Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1

Article

Moisture Content Measurement of Broadleaf Litters using Near-Infrared Spectroscopy Technique

Ghiseok Kim [1,3], Suk-Ju Hong [1], Ah-Yeong Lee [1], Ye-Eun Lee [2] and Sangjun Im [2,3,*]

[1] Department of Biosystems and Biomaterials Science and Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea; ghiseok@snu.ac.kr (G.K.); hsj5596@snu.ac.kr (S.H.); ayoung3308@snu.ac.kr (A.L.);

[2] Department of Forest Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea; dldpdms109@naver.com (Y.L.); junie@snu.ac.kr (S.I.)

[3] Research Institute of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea; ghiseok@snu.ac.kr (G.K.); junie@snu.ac.kr (S.I.)

* Correspondence: junie@snu.ac.kr; Tel.: +82-2-880-4759

Abstract: Near-infrared spectroscopy (NIRS) was implemented to monitor the moisture content of broadleaf litters. Partial least-squares regression (PLSR) models, incorporating optimal wavelength selection techniques, have been proposed to better predict the litter moisture of forest floor. Three broadleaf litters were used to sample the reflection spectra corresponding the different degrees of litter moisture. Maximum normalization preprocessing technique was successfully applied to remove unwanted noise from the reflectance spectra of litters. Four variable selection methods were also employed to extract the optimal subset of measured spectra for establishing the best prediction model. The results showed that the PLSR model with the peak of beta coefficients method was the best predictor among all candidate models. The proposed NIRS procedure is thought to be a suitable technique for on-the-spot evaluation of litter moisture.

Keywords: near-infrared spectroscopy; multivariate analysis; partial least-squares regression; floor litter; optimal wavelength selection

1. Introduction

Floor litter refers to the relatively fresh organic residue on the uppermost layer of soil; it plays an important role in the water dynamics of the forest floor [1]. It can retain water within the layer during storm periods and deplete stored moisture through evaporation [2]. The moisture variation of litter can influence the hydrologic, carbon, and nutrient cycles of forests by altering the wetting and drying phases [3,4]. Moreover, litter moisture content is one of the critical determinants for fire ignition and spread in forests [5].

Several attempts have been made to quantitatively measure the moisture variation of floor litter over the past couple of decades [6]. Among them, gravimetric method is the most common technique that determines the moisture amount from the difference in the litter weight of wet and dry conditions [7]. Yet it is a highly cumbersome and labor-intensive measurement.

Some researchers have attested the continuous non-destructive techniques for litter moisture measurement. Gillespie and Kidd proposed the electrical impedance grids with mock leaf sensors to monitor leaf wetness of crops [8]. This technique was further improved by Hanson et al. [9]. On the other hand, time-domain reflectometry (TDR) probes have been employed in litter moisture measurement [10,11]. Ataka et al. developed a rather simple technique to continuously detect moisture content in litters using the modified capacitance sensors [1]. Robichaud and Bilskie developed a commercial device for measuring dead fuel moisture with a frequency domain (FD) sensor [12]. However, these methods exhibit inherently electrical bias and limitations for *in situ* measurement of litter moisture, because floor litter is highly heterogeneous and porous material with lower density and compactness.

 doi:10.20944/preprints201709.0099.v1

Spectrometry has been adopted to analyze the quality of crops, chemicals and biomaterials. Near-infrared spectroscopy (NIRS) is a promising spectroscopic method that treats the near-infrared (NIR) region of an electromagnetic spectrum, which corresponds to the wavelength range of 750 to 2,500 nm [13]. As a molecule has a different absorption frequency in the NIR region due to the vibrational patterns of chemical bonds (O-H, C-H, and N-H), NIRS can detect the special spectrum emitted from the intrinsic components of a test object. The frequency at which a certain vibration occurs is determined by the magnitude of bonds and mass of component atoms. A NIR device structurally consists of a light splitter, such as fiber optic cable, a monochromator, and a detector. Since this simple assembly is relatively easy to implement, NIRS is being widely used in nondestructive quality analysis for food, feed, pharmaceutical, and agricultural industries [14–16].

Sometimes NIRS technique is facing a serious difficulty in use. Spectral reflectance could be adulterated with the undesired environment and instrument causes. This undesired influence can reduce the detection accuracy of a prediction model. However, by applying suitable preprocessing techniques, the inappropriate information can largely be eliminated [17]. Preprocessing is a particular data analysis technique that eliminates unwanted noise for enhancing spectral features to clearly represent object characteristics. Rinnan et al. provided an overview of preprocessing techniques widely used in NIRS applications [18].

NIR spectra possess complex and overlapping absorption bands, so that mathematical procedures are needed for turning the spectra into meaningful information. Multivariate analysis refers to statistical procedure for analysis of data sets with more than one variable. It is capable of describing how the measured spectral features are related to the property of interest [19,20]. Until now, a number of multivariate techniques, such as partial least-squares regression (PLSR), principal component regression (PCR), and multiple linear regression (MLR), are commonly used to extract quantitative and qualitative information from NIR spectra [21,22].

Recently, considerable effort has been made in variable selection on multivariate data analysis. A highly correlated large spectrum can rather reduce the predictive ability of a model. Variable selection is a critical step to establish a reliable and robust model so it extracts a subset of spectral frequencies to produce better prediction results [23,24]. Many mathematical approaches for optimal wavelength selection have been derived from the scientific knowledge on the spectroscopic properties of samples [25].

In this study, we developed a non-contact technique to quantitatively monitor the moisture content of broadleaf litters using the NIRS. Floor litter samples were taken from deciduous forests, and gravimetric method was introduced to obtain the reference moisture content of litter samples. PLSR multivariate analysis has been employed to quantitatively predict the moisture content of floor litter. Four different methods were used to extract the optimal wavelength range through model development, and the model performance was also evaluated in conformity with variable selection methods.

2. Materials and Methods

2.1. Litter Moisture Measurement

The NIR spectra were acquired from the current year's litters, collected in the Seoul National University Arboretum in Seoul, Korea. The broadleaf litters were manually gathered under deciduous trees such as Chinese cork oak (*Quercus variabilis*), Sawtooth oak (*Quercus acutissima*), and Mongolian oak (*Quercus mongolica*). Sampled litters were first sorted to three types according to tree species, and each type of litter was separated into its 11 subsamples to represent different moisture conditions. All samples were immersed in water for 12 h in order to fully saturate, and then placed in a screened plate for a certain times to naturally drip water from samples. The reflectance spectra and litter moisture have been measured from three leaves of each sample at every one hour. The weight of each litter was immediately measured to obtain the wet weight of litter when NIR measurement was done. This process was repeated for all subsamples. After the experiment was completed, all litter samples were placed in an oven at a temperature of 70 °C for 48 h, and measured

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1

the weight of oven-dried litter to obtain the dry weight. The moisture content of litter was calculated based on the gravimetric method as follow:

$$\text{Moisture content}(\%) = \frac{w_w - w_d}{w_d} \times 100 \quad (1)$$

where w_w and w_d are the litter weight of wet and oven-dried conditions, respectively.

2.2. NIR Spectra Measurement

NIR reflectance spectra of litter samples were sampled in spectral range from 904 to 1,707 nm with the NIR spectrometer (CDI-NIR128, Control Development Inc., USA). The tungsten-halogen lamp (LS-1, Ocean Optics, USA) was installed as a light source as shown in Figure 1. The reflectance spectrum was taken at three different positions on the litter surface, and then the averaged values were recorded according to the moisture content of litter samples. The integration time was set to 0.25 s, and the distance between the sample and a probe was maintained at 1 cm. The reflectance of the measured spectra was adjusted with white-referenced and dark-referenced spectra (Equation (2)). The white-referenced spectrum was acquired from a white Teflon board, while the dark-referenced was measured with a completely blocked fiber optic cable in a dark room.

$$\text{Reflectance}(\%) = \frac{S_i - D}{R - D} \times 100 \quad (2)$$

where S_i is the raw reflectance of the i-sample, and D and R represent the raw reflectance of dark- and white- referenced spectra, respectively.

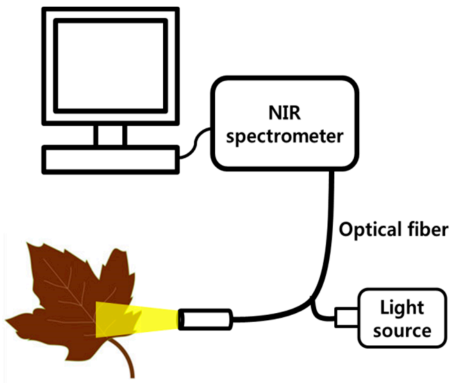


Figure 1. Schematic diagram of NIR spectra measurement.

2.3. Multivariate Model Development

The pretreatment of raw NIR spectra is the first step for model development and optimization to achieve the better NIR veracity. The instrument and environment causes may lead to sample-to-sample variations such as noise, light scattering, and optical path changes. Generally, spectral preprocessing is strongly demanded to exclude noise components in reflectance spectra. In this study, two scatter correction techniques (multiplicative scatter correction(MSC), standard normal variate(SNV)), three normalization techniques (maximum normalization, mean normalization, range normalization), and three Savitzky-Golay(SG) filters (smoothing, the first- and second-derivative techniques) were used for spectra pretreatment.

The PLSR model was also employed to establish the relationship between litter moisture content and reflectance characteristics. PLSR is a classical and widely used statistical method that bears a relation between independent and dependent variables in large data sets. The general structure of PLSR model is written as follow:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1

$$
\begin{aligned}
X &= TP^T + E \\
Y &= TQ^T + F
\end{aligned}
\tag{3}
$$

where X is the n by m matrix of predictors, Y is the n by p matrix of responses, T is the n by 1 matrix of the score matrix, E and F are error terms, and P and Q are the m by 1 and p by 1 loading matrices.

The performance of PLSR model is quantitatively assessed with the root mean square error (RMSE) and the coefficient of determination (R^2). For calibration purpose, the regression model with a small RMSE value and a large value of R^2 could be selected as an appropriate model. In this study, latent variables (LVs) set with the smallest RMSE value was determined through the calibration and validation procedures, and then the prediction model performance was evaluated through the further test process.

A calibration model has to be evaluated with a validation set of samples to get an impression of its predictive ability. Model test has sometimes conducted in the NIRS research to further ensure model performance when the calibration and validation have completed. Among 99 spectra dataset, model calibration, validation, and test have conducted from 60, 15, and 24 dataset, respectively. Both preprocessing and regression analysis were done by using MATLAB commercial software (ver. R2016, MathWorks, Natick, MA, USA).

2.5. Optimal Wavelength Selection

The reflectance spectra, ranging 904 to 1,707 nm, were used in this study for establishing a regression model to predict the moisture content of litter. PLSR technique generally uses the entire range of wavelength's spectra, but recent studies have employed the most appropriate wavelength. It proved that data analysis involving only a representative part of the spectra can lead to a better prediction performance. Selection of optimal wavelength is to pick carefully the subset of spectral data, which closely relate to the property of the interest. If the number of wavelength bands are greater than that of spectral samples, the predictive model is likely to enhance its capability with the optimal wavelength. In this study, various techniques such as the peak of beta coefficients, variable importance in projection (VIP), bootstrap of beta coefficients, and interval PLS (iPLS) were applied to determine the optimal wavelength from preprocessed spectra information.

2.5.1. Peak of Beta Coefficient

The peak of beta coefficients (beta-peak method) extracts the optimal wavelength corresponding to the peak of beta coefficients. The beta coefficient, the regression coefficient of PLSR model, is a linear vector between the measured spectra and the predicted values. It measures the contribution of each wavelength set on the model's predictive ability so as to select the optimum wavelength.

2.5.2 Variable Importance in Projection

VIP is a method that evaluates the contribution of a variable on the weight matrix of a multivariate analysis such as a PLSR model. It has been widely applied to determine major wavelength bands when the variable matrix X defined in Equation (3) was considered [26]. Lohumi et al. estimated how much the VIP value (Equation 4) could be influenced by the relationship between the dependent variable matrix Y and the independent variable matrix X [27]:

$$
v_j = \sqrt{\frac{p\sum_{a=1}^{A}[SS_a(w_{ak}/\|w_a\|^2)]}{\sum_{a=1}^{A}(SS_a)}},
\tag{4}
$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1

where v_j represents the VIP value, p is the number of variables, SS_a is the sum of squares explained by the a^{th} component, w_a is the weight of a^{th} component, and w_{ak} is the weight of the a^{th} component at the k^{th} variable.

The variable selection process by a VIP technique is terminated when the calculated VIP value approaches to the threshold. In general, a threshold value of 1 is set in many studies[28]. In this study, we compared the accuracies of developed models for selected wavelength sets by adjusting a threshold value.

2.5.3 Bootstrap of Beta coefficients

The bootstrap of beta coefficients (bootstrap method) is a useful technique to set the confidence interval, and estimate the significance level. It is mainly used when it is difficult to estimate the distribution of samples. In the bootstrap method, a new dataset is first obtained by randomly re-sampling the sample population. Thereafter, statistical values are obtained from the selected data sets, and then the confidence interval is set [29]. The PLSR-bootstrap method can re-sample the sample data n times and perform a PLSR for each re-sampled data set to obtain the beta coefficient. From the distribution of beta coefficients, the confidence interval is obtained according to the significance level of the beta coefficient of the specific variable. Finally, optimal wavelength is determined by removing the variables with zero value in its confidence interval. The confidence interval is computed by

$$I_k = \overline{b_k} \pm c s_k , \tag{5}$$

where I_k represents the confidence interval, $\overline{b_k}$ is the mean of beta coefficients at the kth variable, c is a constant that determines the confidence interval, and S_k is the standard deviation of beta coefficients at the kth variable.

In equation (5), the constant c is explicitly determined regarding the level of significance. A higher significance level could select fewer wavelength bands. In this study, the number of re-sampling is set to 1,000, and the model accuracy with the selected wavelength is evaluated by changing the c values [30].

2.5.4 Interval PLS

iPLS executes a PLSR model with the wavelength of a specific interval rather than the entire wavelength band. It removes unnecessary wavelength bands and uses only the certain bands of a given interval. The interval is repeatedly sampled to achieve the lowest RMSE value. In this study, the entire wavelength band divided into 20 intervals and PLSR analysis was conducted to acquire the optimal wavelength by shifting the wavelength range to the right or the left until the smallest RMSE value was pursued [31].

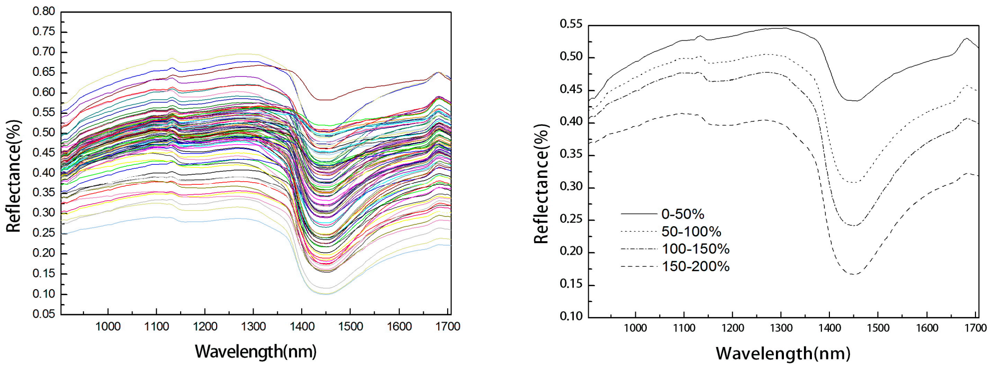
3. Results

3.1. Reflectance Spectra of Litters

Figure 2 depicts the raw and average NIR reflectance spectra of litter samples in the wavelength range of 900-1,700 nm prior to the pretreatment. There seems to be a slight variation in all reflectance spectra through the liter moisture (Figure 2a). Figure 2b illustrates the averaged spectra with different degrees of litter moisture (0–50%, 50–100%, 100–150%, and 150–200%). The spectral difference appeared more clearly in Figure 2b. The absorption rate of NIR spectrum increased in the wavelength range of 1,400–1,500 nm regardless of litter moisture. It is commonly known that NIR spectrum is mostly absorbed at 980, 1,222, and 1,450 nm wavelength by the overtone and combination of the vibrational transitions of water molecules [32]. On the NIR reflectance of litters, the absorption coefficients at 980 and 1,222 nm wavelength were estimated to be less than 1.3 cm^{-1}, while the value was estimated to be 29.8 cm^{-1} at a wavelength of 1,450 nm [32]. Figure 2b exhibits that the reflectivity

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1

of NIR spectrum decreased sharply beyond the wavelength of 1,400 nm and had the lowest value at the wavelength of 1,450 nm for all litter samples.



(a) (b)

Figure 2. Near-infrared reflectance spectrum of litter samples prior to preprocessing: (a) raw spectra; (b) averaged spectra.

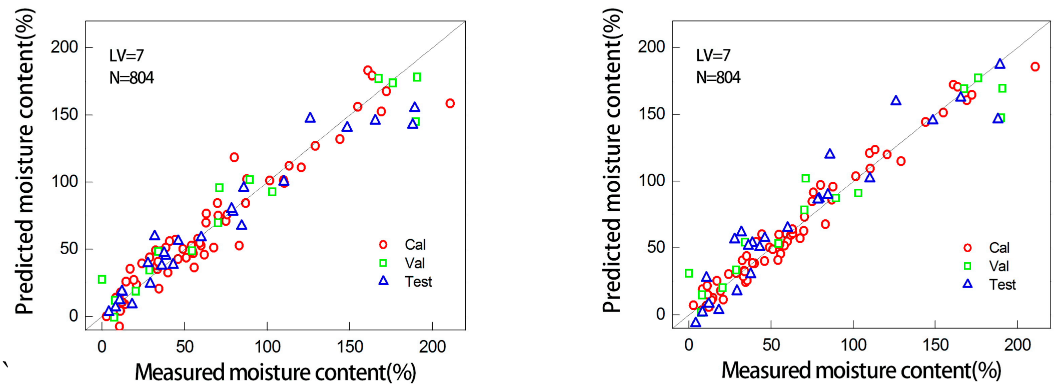
3.2. PLSR model for different preprocessing methods

A number of pretreatment techniques for NIR spectra have been tested to achieve better performance in litter moisture determination. Table 1 compares the results of PLSR model with different spectral preprocessing techniques. The entire range of NIR dataset was treated and used to develop PLSR models with the different preprocessing techniques. The performance of the model was evaluated with two statistical criteria, coefficient of determination (R^2) and root mean square error (RMSE), between the reference and prediction of litter moisture content. The PLSR models treated by the preprocessing methods of MSC, SNV, maximum normalization, and range normalization have good prediction performance for validation. The test results of the models showed that maximum normalization achieved the highest performance (R_t^2: 0.922, $RMSE_t$: 15.711), and followed by the model with range normalization preprocessing (R_t^2: 0.920, $RMSE_t$: 15.970). Figure 3 shows the prediction result of two most highly correlated PLSR models for litter moisture estimation.

Table 1. PLSR results with different preprocessing techniques.

Preprocessing Method	LVs	Calibration		Validation		Test	
		R_c^2	$RMSE_c$	R_v^2	$RMSE_v$	R_t^2	$RMSE_t$
Raw data	9	0.930	12.757	0.920	19.041	0.884	19.170
SG smoothing	8	0.918	13.848	0.915	19.571	0.897	18.103
SG-1st derivative	6	0.918	13.882	0.925	18.376	0.883	19.259
SG-2nd derivative	6	0.937	12.114	0.913	19.792	0.840	22.506
MSC	8	0.926	13.132	0.933	17.432	0.914	16.482
SNV	8	0.927	13.069	0.943	16.106	0.915	16.426
Max. normalization	7	0.920	13.699	0.938	16.797	0.922	15.711
Mean normalization	9	0.926	13.149	0.926	18.308	0.892	18.479
Range normalization	7	0.922	13.535	0.931	17.722	0.920	15.970

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1



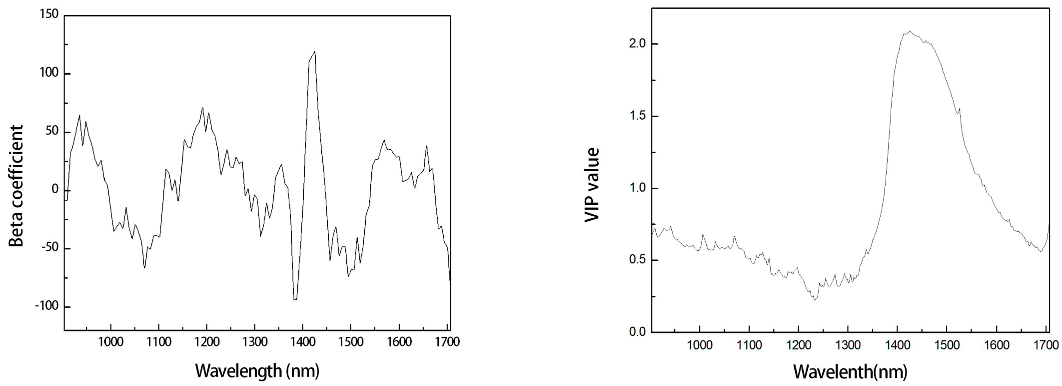
(**a**) (**b**)

Figure 3. Comparison of predicted and measured moisture content. The prediction values were taken from the PLSR model with **(a)** maximum normalization, and **(b)** range normalization preprocessing.

3.3. PLSR models for optimal wavelength selection methods

Four variable selection methods were used to set the optimal wavelength bands from around 800 data set of wavelength (900–1,700 nm). Maximum normalization was used as a pretreatment technique through this step because the PLSR model incorporating maximum normalization produced the highest performance in litter moisture estimation. The selected wavelength set was then input the PLSR model for predicting the moisture content of floor litter. Figure 4 illustrates the beta coefficients and VIP values of PLSR model derived from the entire wavelength's spectra. As shown in Figure 4a, the peak points of beta coefficients were evenly distributed over the entire wavelength band, and the largest peak point was discovered around 1,450 nm. But, large VIP values are observed only in the range of 1,400–1,500 nm.

Figure 5 shows the first four weight vectors with respect to equivalent latent variables of the PLSR model obtained from the entire wavelength's spectra. The weight vector can imply the correlation between the measured spectrum and calculated latent variables through PLSR analysis. It revealed how much the measured spectrum contributed to the formation of corresponding latent variables. As shown in Figure 5a–d, the highest peak is observed around the wavelength of 1,400–1,500 nm, which is highly correlated with the largest absorption coefficient of a water molecule.



(**a**) (**b**)

Figure 4. Beta coefficients (a) and VIP values (b) of the PLSR model with the entire wavelength's spectra.

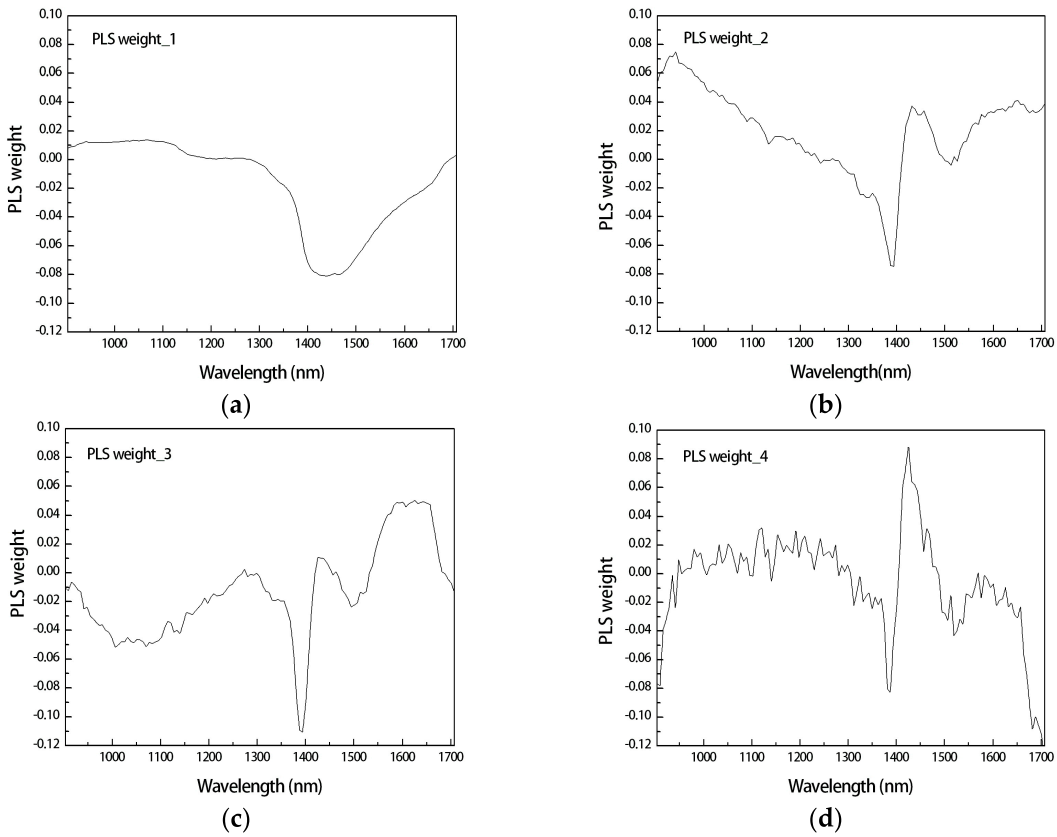


Figure 5. First four weight vectors with respect to equivalent latent variables of the PLSR model with the entire wavelength's spectra: **(a)** First weight vector; **(b)** Second weight vector; **(c)** Third weight vector; **(d)** Fourth weight vector.

Figure 6 shows the optimal wavelength bands determined by four variable selection methods. By applying the beta-peak method, 63 optimal wavelength bands corresponding to all peak points of the beta coefficient were selected, and they were uniformly distributed over the entire wavelength, as shown in Figure 6a. VIP method was implemented with different threshold values ($v = 0.7, 1.0, 1.5, 2.0$), and the wavelength ranged from 1,375 to 1,577 nm was finally selected for $v = 1.0$ (Figure 6b). In the case of the bootstrap method, the significant level constant c was set to four cases ($c = 1.0, 1.3, 1.6, 1.9$). For $c = 1.3$, the extracted wavelength bands were relatively evenly distributed over the entire range, as shown in Figure 6c. After applying the iPLS technique with several intervals, 150 variables were selected in the wavelength range of 1,394–1,543 nm (Figure 6d).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1

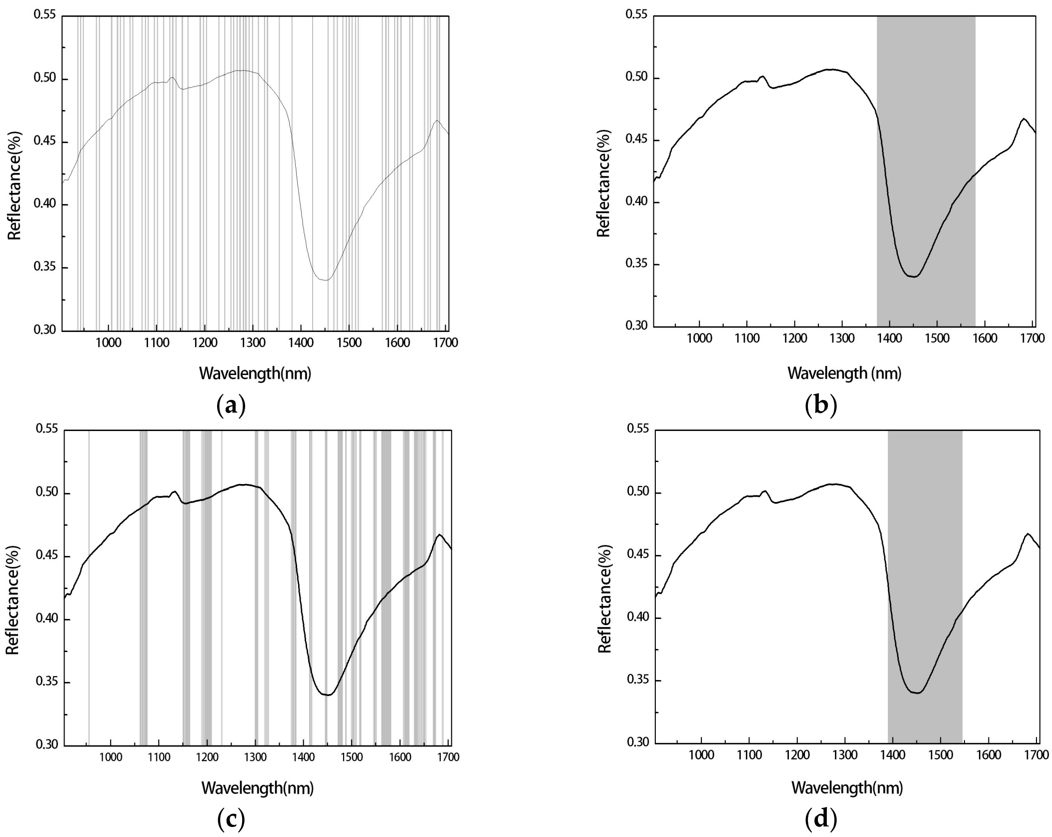


Figure 6. Optimal wavelength derived by variable selection methods: **(a)** beta-peak (N = 63); **(b)** VIP (v = 1.0, N = 203); **(c)** bootstrap (c = 1.3, N = 192); **(d)** iPLS (N = 150).

Table 2 presents the calibration, validation, and test performance of the PLSR models developed from the optimally selected wavelength's spectra. Predictive abilities of different models have been compared against the PLSR model with the entire wavelength's spectra (hereafter the full PLSR model). As 63 wavelength bands, selected by the beta-peak method, were input in the PLSR model, the model could better predict the litter moisture than the full PLSR model. The VIP method also confirmed that the performance of the PLSR model taken from the 316 spectra showed better prediction performance (R_t^2: 0.923, $RMSE_t$: 15.597) than that of the full PLSR_model. VIP value of 1.5 marked 150 wavelength set among the entire wavelength range, and had a similar ability for litter moisture estimation (R_t^2: 0.920, $RMSE_t$: 15.916) with the entire wavelength PLSR_model.

The bootstrap method was applied to select the optimal wavelength. The PLSR model with 305 spectra data (c = 1.0) showed better capability in data analysis (R_t^2: 0.927, $RMSE_t$: 15.164), while the PLSR model with 106 spectra (c = 1.6) also showed similar prediction performance (R_t^2: 0.918, $RMSE_t$: 16.128) when compared to the full PLSR model. A total of 150 wavelength bands was optimally selected in the iPLS technique application, and the prediction performance of PLSR model achieved the R_t^2 value of 0.918, $RMSE_t$ value of 16.115.

As presented in Table 2, the predictive abilities of the PLSR models with reduced wavelength bands by employing the variable selection methods showed similar or better performance with the full PLSR_model, excepting only two cases of VIP (v = 2.0) and bootstrap method (c = 1.9). But, optimal wavelength selection can improve the prediction accuracy by effectively identifying the best subset of candidate spectra. It also eliminates unnecessary information so as to enhance the efficiency and effectiveness of model run. Figure 7 shows the prediction results of the best PLSR models, which were derived from the optimally selected wavelength's spectra.

Table 2. PLSR performance comparison with variable selection methods.

Method		LVs	No. of Variables	Calibration		Validation		Test	
				R_c^2	$RMSE_c$	R_v^2	$RMSE_v$	R_t^2	$RMSE_t$
Full PLSR		7	804	0.920	13.699	0.938	16.797	0.922	15.711
iPLS		7	150	0.900	17.825	0.930	17.825	0.918	16.115
VIP	v = 0.7	7	316	0.924	13.391	0.934	17.276	0.923	15.597
	v = 1.0	7	203	0.905	20.131	0.910	20.131	0.918	16.130
	v = 1.5	7	150	0.909	14.620	0.926	18.314	0.920	15.916
	v = 2.0	7	40	0.884	16.522	0.853	25.818	0.859	21.167
Boots-trap	c = 1.0	7	305	0.919	13.795	0.941	16.392	0.927	15.164
	c = 1.3	7	192	0.916	14.012	0.940	16.472	0.924	15.526
	c = 1.6	7	106	0.909	14.586	0.929	17.954	0.918	16.128
	c = 1.9	7	35	0.875	17.1670	0.884	22.913	0.878	19.655
Beta-peak		6	63	0.918	13.834	0.932	17.494	0.930	14.905

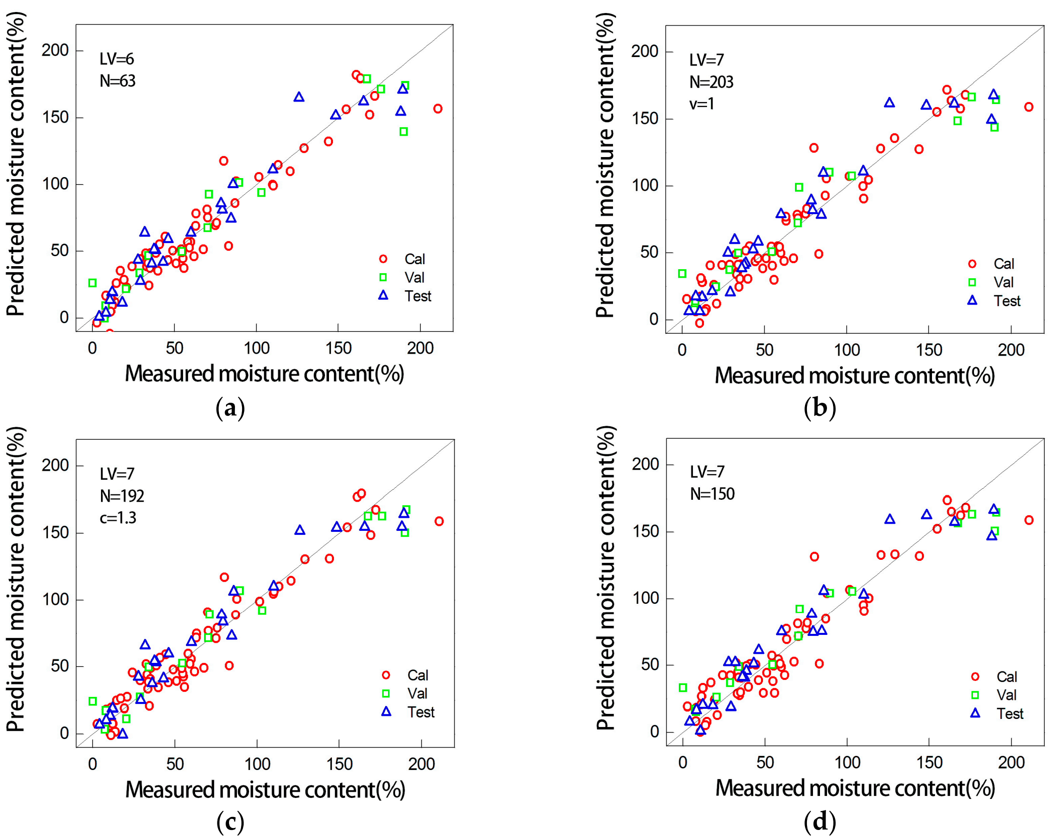


Figure 7. Comparison of the PLSR results with variable selection methods: **(a)** beta-peak; **(b)** VIP; **(c)** bootstrap; **(d)** iPLS.

4. Conclusions

Litter refers to the organic residue on the forest floor, and plays an important role in forest water cycle. Litter moisture is one of the critical determinants for evaporation, infiltration, and fire ignition in a forest environment. In this study, the commonly used technique, NIRS, has been implemented to fast, non-destructively monitor the moisture content of broadleaf litters. Several techniques for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1

data pretreatment and optimal wavelength selection have been tested to establish a reliable and robust multivariate model.

Prior to establishing a PLSR model, the pretreatment was conducted to eliminate the unwanted noise in raw NIR spectra. Maximum normalization preprocessing, one of well-known normalization techniques, achieved the best prediction ability in litter moisture estimation among 8 pretreatment methods. The entire wavelength may uphold the whole information of target samples, but the redundant data causes the decline in model accuracy. The peak of beta coefficients seems to be the best variable selector to extract the optimal wavelength bands as to yield the better predictive ability in litter moisture estimation.

The NIRS has been commonly used in both qualitative and quantitative analysis of target samples with respect to the vibrational energy of molecules. The low optical absorbance characteristics of NIR rays facilitates deeper penetration than mid-infrared rays. This optical method has some merits in reducing time-consuming work and supporting reliable results by minimizing the errors that may originate from the quick changes of litter moisture on the spot.

The NIRS has also limited ability to resolve noise that is caused by the instrument and environment causes. NIR spectra tends to be highly complex and over-parameterized, which sometime yields a poor prediction. In the face of its drawback, multivariate analysis through variable reduction and proper pretreatment have been successfully introduced in practical uses. It is believed that this research provides with great opportunities for multicomponent analysis of agricultural products and foods, as well as many other scientific sectors, such as life sciences and biomaterial research.

Acknowledgments: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A02061769).

Author Contributions: G.K. designed the experimental concept and data analysis methods; S.H and A.L. performed experiments and analyzed the data; Y.L supported experiments and data analysis; S.I. wrote the article and have been supervising, discussing the experiments and edited the presented work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ataka, M.; Kominami, Y.; Miyama, T.; Yoshimura, K.; Jomura, M.; Tani, M. Using capacitance sensors for the continuous measurement of the water content in the litter layer of forest soil. *Appl. Environ. Soil Sci.* **2014**, ID627129. Doi:dx.doi.org/10.1155/2014/627129
2. Stocks, J.B. Moisture in the forest floor - its distribution and movement. Dept. of Fisheries and Forestry, Canadian Forestry Service, **1970**, No. 1271. p. 20.
3. Xiao, Q.; McPherson, E.G. Surface water storage of twenty tree species in Davis, California. *J. Environ. Qual.*, **2016**, 45, 188–198. Doi:10.2134/jeq2015.02.0092
4. Borken, W.; Matzner, E. Reappraisal of drying and wetting effects on C and N mineralization and fluxes in soils. *Glob. Change Biol.*, 2009, 15, 808–824.
5. Johnson, E.A.; Miyanishi, K. *Forest Fires: Behavior and Ecological Effects*; Academic Press: Washington, DC, USA, 2001, p. 594.
6. Schunk, C.; Ruth B.; Leuchner, M.; Wastl, C.; Menzel, A. Comparison of different methods for the in situ measurement of forest litter moisture content. *Nat. Hazard Earth Sys.*, **2016**, 16, 403–415. doi:10.5194/nhess-16-403-2016.
7. Wotton, B.M.; Stocks, B.; Martell, D.L. An index for tracking sheltered forest floor moisture with the Canadian Forest Fire Weather Index System. *Int. J. Wildland Fire*, **2005**, 14, 169–182. Doi:10.1071/WF04038
8. Gillespie, T.J; Kidd, G. Sensing duration of leaf moisture retention using electrical impedance grids. *Can. J. Plant Sci.*, **1978**, 58, 179–187
9. Hanson, P.J.; O'Neill, E.G.; Chambers, M.L.S.; Riggs, J.S.; Joslin, J.D.; Wolfe, M.H. Soil respiration and litter decomposition. In *North America Temperature Deciduous Forests Response to Changing Precipitation Regimes*; Hanson, P.J.; Wullschleger, S.D., Eds.; Springer: New York, USA, 2003, pp 163–189.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 September 2017 doi:10.20944/preprints201709.0099.v1

10. Borner, T.; Johnson, M.G.; Ryglewlcz, P.T.; Tingey, D.T.; Jarrell, G.D. A two-probe method for measuring water content of thin forest floor litter layers using time domain reflectometry. *Soil Technol*., **1996**, 9, 199–207
11. Pumpanen, J.; Ilvesniemi, H. Calibration of time domain reflectometry for forest soil humus layers. *Boreal Environ. Res*., **2005**, 10, 589–595
12. Robichaud, P.R.; Bilskie, J. A new tool for fire managers-an electronic duff moisture meter. *Fire Management Today*, **2004**, 64, 15–18.
13. American Society for Testing and Materials, E1866-97: Standard Guide for Establishing Spectrophotometer Performance Tests, and E1655: Standard practices for infrared, multivariate, quantitative analysis, Official ASTM Publications, Philadelphia, PA, **2000**.
14. Choe, E.; van der Meer, F.; van Ruitenbeek, F.; van der Werff, H.; de Smeth, B.; Kim, K.W. Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: a case study of the Rodalquilar mining area, SE Spain. *Remote Sens. Environ*., **2008**, 112, 3222–3233.
15. Choi, C.H.; Lee, K.J.; Park, B.S. Prediction of soluble solid and firmness in apple by visible/near-infrared spectroscopy. *J. Korean Society for Agricultural Machinery*, **1997**, 22, 256–265. (In Korean)
16. Lu, R.F.; Guyer, D.E.; Beaudry, R.M. Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *J. Texture Stud*., **2000**, 31, 615–630.
17. Porep, J.U; Kammerer, D.R; Carle R. On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology*, **2015**, 46, 211–230.
18. Rinnan, A°; Berg, F. van den; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, **2009**, 28, 1201–1222.
19. Kim, Y.T; Suh, S.R. Comparison of performance of models to predict hardness of tomato using spectroscopic data of reflectance and transmittance. *J. Biosystems Eng*., **2008**, 33, 63–68. (In Korean)
20. Kawano, S.; Watanabe, H.; Iwamoto, M. Determination of sugar content in intact peaches by near infrared spectroscopy. *J. Japan. Soc. Hort. Sci*., **1992**, 61, 445–451.
21. Dull, G.G.; Leffler, R.G.; Birth, G.S.; Smittle, D.A. Instrument for non-destructive measurement of soluble solids in honeydew melons. *Trans. Am. Soc. Agric. Eng*., **1992**, 35, 735–737.
22. Chung, H.-I.; Kim, H.-J. Near-infrared spectroscopy: Principles. *Anal. Sci. Technol*., **2000**, 13, 138–151.
23. Sarathjith ,M.C; Das, B. S; Wani, S.P; Sahrawat, K.L. Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils, *Geoderma*, **2016**, 267, 1–9.
24. Zhang, C; Jiang, H; Liu, F; He, Y. Application of Near-Infrared Hyperspectral Imaging with Variable Selection Methods to Determine and Visualize Caffeine Content of Coffee Beans. *Food and Bioprocess Technology*, **2017**, 10, 213–221.
25. Fernández Pierna, J.A; Abbas, O; Baeten, V; Dardenne, P. A backward variable selection method for PLS regression (BVSPLS). *Analytica Chimica Acta*, **2009**, 642, 89–93.
26. Anderson, C.M.; Bro R. Variable selection in regression-a tutorial. *J. Chemometric*., **2010**, 24, 728–737.
27. Lohumi, S.; Lee, S.; Cho, B.-K. Optimal variable selection for Fourier transform infrared spectroscopy analysis of starch-adulterated garlic powder. *Sensor. Actuator. B*, **2015**, 216, 622–628.
28. Mehmood, T. et al. A review of variable selection methods in partial least squares regression. *Chemometric. Intell. Lab*., **2012**, 118 62–69.
29. Bickel, P.J.; Freedman, D.A. Some asymptotic theory for the bootstrap. *Ann. Stat*., **1981**, 9, 1196–1217.
30. Lazraq, A. et al. Selecting both latent and explanatory variables in the PLS1 regression model. *Chemometric. Intell. Lab*., **2003**, 66, 117–126.
31. Norgaard, L. et al. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Appl. Spectrosc*., **2000**, 54, 413–419.
32. Collins, J.R. Change in the infra-red absorption spectrum of water with temperature. *Phys. Rev*., **1925**, 26, 771–779.