RETRIEVAL OF FUEL MOISTURE CONTENT FROM HYPERSPECTRAL DATA VIA PARTIAL LEAST SQUARE

Jie Zhang[1,2,3], Jianjun Wu[1,2,3,], Lei Zhou[1,2,3]*

1. State Key Laboratory of Earth Surface Process and Resource Ecology, Beijing Normal University, Beijing 100875, China; 2. Key Laboratory of Environment Change and Natural Disaster, MOE, Beijing Normal University, Beijing 100875, China; 3. Academy of Disaster Reduction and Emergency Management, MOCA/MOE, Beijing Normal University, Beijing 100875, China

ABSTRACT

As an important indicator of vegetation moisture status, Fuel Moisture Content (FMC) is commonly used for predicting vulnerability to wild fire. Currently, the FMC estimation using spectral data is mainly based on spectral indices derived from several bands and these methods do not make full use of the entire spectrum. Partial Least Square (PLS) is a new multivariate statistical method which can effectively reduce collinearity. In this paper, using LOPEX dataset, we mainly explored the performance of PLS coupled with different feature selection methods for FMC retrieval. According to the results, PLS shows great potential to extract FMC from spectral data; when coupled with different band selection approaches, the models also generate high estimation precision; with band selection, the PLS coupled models involved fewer bands, lowering the model complexity. Thus, the high estimation precision and much simpler modeling make band selection-PLS coupled methods superior to original PLS for FMC retrieval.

Index Terms— Retrieval, hyperspectral, PLS

1. INTRODUCTION

Dynamic monitoring of vegetation moisture status plays an important role in agriculture and forestry. The changes of vegetation water content can result in different spectral characteristics, which provides more potential to explore vegetation moisture status using spectral data. Vegetation water content expressed as Fuel Moisture Content (FMC) is commonly used in forestry for predicting vulnerability to wild fire [1-4]. Currently, making use of spectral data, FMC prediction is mainly based on the relationships between FMC and calculated spectral indices. However, these methods only use limited number of moisture feature bands and thus tend to neglect the hidden moisture information lying in the entire spectrum. Partial Least Square (PLS) is a newly proposed multivariate statistical method which can effectively deal with multi-fold linearity [5]. Currently, PLS is mainly applied in the field of chemistry. However, little work has been done for FMC estimation via PLS. Using LOPEX dataset, this paper mainly investigated the potential of PLS and band selection-PLS coupled methods for FMC retrieval. First, the FMC estimation model was established via PLS using the entire spectra and the strength of PLS models to estimate FMC over various vegetation types was evaluated. Next, two band selection approaches (band selection by correlation coefficient and genetic algorithm) were coupled with PLS (they were respectively called CC-PLS and GA-PLS in this paper). Besides, the performance of CC-PLS and GA-PLS to extract FMC from LOPEX dataset was examined.

2. DATA

For a wide range of water content, we used the Leaf Optical Properties EXperiment (LOPEX) dataset established by the Joint Research Center (JRC). Collected over various vegetation types in 1993, LOPEX dataset included leaf spectral data and related leaf biophysical/biochemical parameters. In this study, vegetation samples of single leaf with reflectance spectra were selected for analysis. The data involved in this study mainly included leaf reflectance data and leaf weight data (both fresh and dry weight). The reflectance was acquired by Perkin Elmer Lambda 19 double-beam spectrophotometer and the vegetation water content was determined through calculating the difference between fresh and dry weight. In all, 335 fresh leaf measurements were used in this study. Detailed description of LOPEX dataset can be referenced in [6].

*Corresponding Author, E-mail: wjj@ires.cn.

This research has received financial support from the National High-tech R & D Program (863 Program) of China (Grant Number: 2006AA120108).

3. METHODOLOGY

3.1. Fuel Moisture Content

Fuel Moisture Content (FMC) is defined as the ratio of leaf water mass versus fresh weight [7].

$$FMC = \frac{FW - DW}{FW} \quad (1)$$

Where FM and DW respectively represent the fresh and dry weight of leaves.

3.2. Partial Least Square Method

Partial Least Square (PLS), brought out by Wold and Albano in 1983, was firstly applied in chemistry. PLS regression can simultaneously aggregate multivariable linear regression (MLR), principal component analysis (PCA) and canonical correlation analysis (CCA). PLS is superior to MLR since the regression models can still be developed in spite of the severe correlation between independent variables. In the analysis of hyperspectral data, PLS can extract the most useful information from hundreds of bands and combine that into several latent variables. Therefore, PLS provides a new perspective to extract information from hyperspectral data. Detailed description about PLS can be found in [5].

3.3. Band Selection-Partial Least Square Coupled Method

Traditional PLS commonly integrates band information from the entire region, which makes the established models very complicated. Thus, the predictive band selection makes up an important step before applying PLS regression. In this study, two band selection approaches (band selection by correlation coefficient and genetic algorithm) were used. When combined with PLS method, they were respectively mentioned as CC-PLS and GA-PLS in this paper.

3.3.1. CC-PLS

The correlation between leaf spectra and water content varies at different bands and those bands with high correlation can be called moisture predictive bands. Thus, the correlation coefficients can be used for band selection. After feature selection, the selected bands can enter into the following PLS analysis for FMC estimation.

3.3.2. GA-PLS

Genetic algorithm (GA) is a computer model to simulate natural selection. During this process, duplication, transformation and mutation are used to select the good, get rid of the bad, and eventually discover the best solution. Given its capability to simulate an individual's natural evolution, GA is well suitable for solving the problems of variable subset selection. In this paper, GA was used for band selection and then combined with PLS for FMC retrieval. More detailed description of GA-PLS method can be referenced in [8].

3.4. Evaluation of Model Performance

During PLS modeling, the determination of optimal number for latent variables is important. In this paper, this is implemented via cross-validation. Cross-validation could also be used to evaluate the performance of PLS models. In this study, we adopted venetian blind cross-validation for PLS/CC-PLS/GA-PLS. More detailed description about venetian blind can be found in [9].The strength of PLS/CC-PLS/GA-PLS models was evaluated by the coefficient of determination (R^2) and the root mean square error (RMSE) for both calibration and validation models.

In this study, all calculations were done in Matlab PLS toolbox and the reflectance/FMC data in use were normalized with auto-scale method.

4. RESULTS

4.1. Statistics of Fuel Moisture Content

In this paper, FMC was used to describe vegetation water content. And the statistical parameters of FMC for all samples are described in Table 1.

Table 1. Statistical parameters of sample FMC

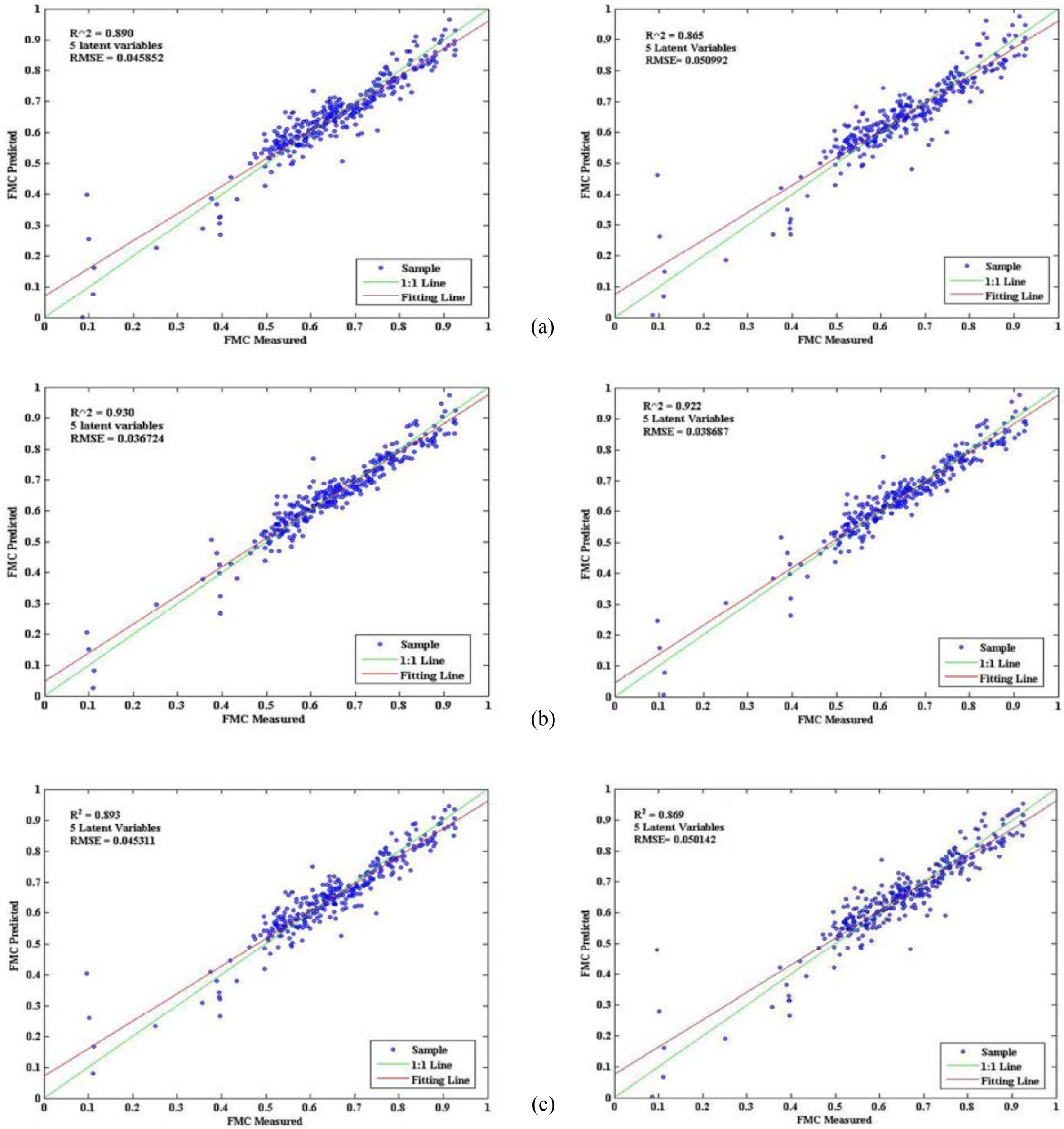
	Min	*Max*	μ	σ	*CV*
FMC	*0.0862*	*0.9264*	*0.6525*	*0.1385*	*21.23%*

From Table 1, the samples have a large variation of FMC, ranging from 0.0862 to 0.9264, which corresponds with the samples collected from different vegetation types.

4.2. FMC Retrieval with PLS

Based on the entire spectrum, the PLS model for FMC prediction was first established and then evaluated with cross-validation. The statistical parameters for PLS models were calculated for both calibration and cross-validation models. The results are presented in Figure 1 and the statistical parameters are listed in Table 2.

As observed from Fig.1 (a) and Table 2, PLS generates good estimation for FMC using the entire spectrum. Via cross-validation, 5 latent variables were determined. For calibration, the precision (R^2) is 0.890 and the estimation error (RMSE) is 0.045852. Compared to calibration model, the validation model results in slightly poorer performance with estimation precision (R^2) of 0.865 and estimation error (RMSE) of 0.050992.



(1) Calibration (2) Cross-Validation

Fig.1. Performance of FMC estimation models(1) Calibration (2) Cross-validation

(a) PLS (b) CC-PLS (c) GA-PLS

4.3. FMC Retrieval with CC-PLS

4.3.1. Correlation between FMC and Reflectance

The correlation between FMC and reflectance at different wavelengths indicates the relative importance of each band for FMC prediction. A larger value demonstrates this band is more important for estimating FMC. The relationships between FMC and reflectance are displayed in Figure 2.

4.3.2. FMC Estimation with CC-PLS

As shown in Figure 2, the correlation between FMC and reflectance varies at different bands. In this paper, 0.4 is defined as the threshold for band selection. Those bands with the absolute value of correlation above 0.4 can be used as predictive bands. After feature selection, the selected bands were combined with PLS for FMC retrieval. The relationships between predicted and measured FMC are presented in Fig.1. And Table 2 also shows the statistical parameters of CC-PLS models.

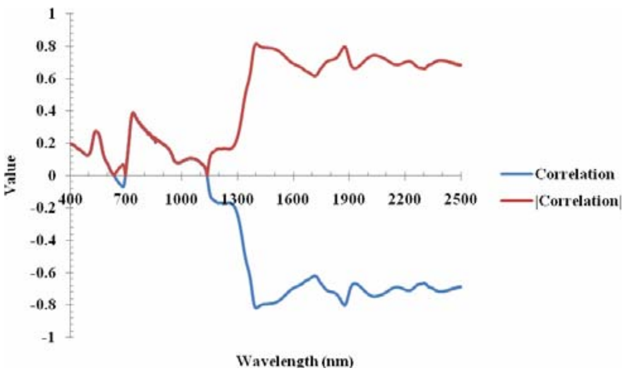


Fig.2. Correlation between FMC and reflectance

From Fig.1 (b) and Table 2, CC-PLS models also yield good results for both calibration and validation analysis. Similar with PLS modeling, the validation model shows slightly poorer performance. Using 5 latent variables, the model precision (R^2) is 0.930 and the estimation error is 0.036724 for calibration; the cross-validation model results in estimation precision of 0.922 and estimation error of 0.038687. Compared to PLS models, CC-PLS models produce better results despite fewer bands are used.

4.4. FMC Retrieval with GA-PLS

Before PLS modeling, GA was first applied for band selection. Then, the quantitative models of FMC estimation were established via PLS. The results are presented in Fig.1 and the statistical parameters are summarized in Table 2.

According to Fig.1 (c) and Table 2, GA-PLS shows good performance for FMC prediction in both calibration and validation models. Similarly, the validation models produce slightly lower precision and accuracy than calibration. Via cross-validation, 5 latent variables were also used in GA-PLS modeling. The estimation precision is 0.893 and the estimation error is 0.045311 for calibration. The validation model has a precision of 0.869 and an estimation error of 0.050142. As compared to PLS, GA-PLS yields slightly better estimation with a much reduced model complexity.

Table 2. Performance evaluation of FMC estimation models

	Calibration		*Validation*	
	R^2	*RMSE*	R^2	*RMSE*
PLS	*0.890*	*0.045852*	*0.865*	*0.050992*
CC-PLS	*0.930*	*0.036724*	*0.922*	*0.038687*
GA-PLS	*0.893*	*0. 045311*	*0.869*	*0.050142*

5. CONCLUSIONS AND DISCUSSIONS

In this paper, we mainly explored the strength of PLS and band selection-PLS coupled method to extract FMC from hyperspectral data. The main conclusions are:

Using LOPEX dataset, PLS shows good performance to estimate FMC from spectral data. For PLS, the R^2 is 0.890 for calibration and 0.865 for validation. When coupled with band selection, the established models also yield good estimation for both CC-PLS and GA-PLS. The CC-PLS model results in an estimation precision (R^2) of 0.930 for calibration and 0.922 for validation. Also, the GA-PLS models have a determination coefficient (R^2) of 0.893 for calibration and 0.869 for validation. Besides, after band selection, the established models incorporate fewer bands, thus lowering the model complexity. Therefore, the high estimation precision and much simpler modeling make the band selection-PLS coupled methods superior to original PLS. Compared to CC-PLS, GA can be automatically implemented in software package, making it more convenient for analysis. However, influenced by internal parameter settings of GA, there may be negligible variations between GA-PLS estimation precision corresponding to different GA runs. However, this does not influence the potential of GA-PLS for estimating FMC. The results demonstrate that PLS and PLS coupled methods provide a new perspective to extract water content from vegetation reflectance and the hyperspectral imagery.

6. REFERENCES

[1] S.L. Ustin, D.A. Roberts, J. Pinzón*, et al.*, "Estimating Canopy Water Content of Chaparral Shrubs Using Optical Methods," *Remote Sensing of Environment,* vol. 65, pp. 280-291, 1998.

[2] S.J. Pyne, P.A. Andrews and R.D. Laven, "Introduction to wildland fire," ed. New York: Wiley, 1996.

[3] E. Chuvieco, D. Cocero, D. Riaño*, et al.*, "Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating," *Remote Sensing of Environment,* vol. 92, pp. 322-331, 2004.

[4] D. Riaño, P. Vaughan, E. Chuvieco*, et al.*, "Estimation of fuel moisture content by inversion of radiative transfer models to simulate equivalent water thickness and dry matter content: Analysis at leaf and canopy level," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 43, pp. 819-826, 2005.

[5] H.W. Wang, *Partial least-squares regression method and its application.* Beijing: National Defense Industry Press, 1999.

[6] B. Hosgood, S. Jacquemoud, G. Andreoli*, et al.*, "Leaf Optical Properties EXperiment 93 (LOPEX93)," *Report EUR 16095 EN,* 1994.

[7] M. Maki, M. Ishiahra, and M. Tamura, "Estimation of leaf water status to monitor the risk of forest fires by using remotely sensed data," *Remote Sensing of Environment,* vol. 90, pp. 441-450, 2004.

[8] L. Li, Y.B. Cheng, S.L. Ustin*, et al.*, "Retrieval of vegetation equivalent water thickness from reflectance using genetic algorithm (GA)-partial least squares (PLS) regression," *Advances in Space Research,* vol. 41, pp. 1755-1763, 2008.

[9] B.M. Wise, N.B. Gallagher, R. Bro*, et al.*, "Chemometrics Tutorial for PLS_Toolbox and Solo," 2006.