Soil & Tillage Research 175 (2018) 267–275

Contents lists available at ScienceDirect

Soil & Tillage Research

journal homepage: www.elsevier.com/locate/still

Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data

Haijun Qi[a,c], Tarin Paz-Kagan[b], Arnon Karnieli[c], Xiu Jin[a], Shaowen Li[a,*]

[a] School of Information and Computer Science, Anhui Agricultural University, Hefei 230036, China
[b] Department of Sensing, Information and Mechanization Engineering, Institute of Agricultural Engineering, Agricultural Research Organization (ARO), Volcani Center, Israel
[c] The Remote Sensing Laboratory, Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boqer Campus 84990, Israel

ARTICLE INFO

Keywords:
Spectral analysis
Soil measurements
Prediction methods
Calibration
Regression algorithms

ABSTRACT

Soil nutrients, including available nitrogen (N), phosphorous (P), and potassium (K), are critical properties for monitoring soil fertility and function. Spectroscopy analysis has proven to be a rapid and effective means for predicting soil properties, in general, and NPK, in particular. However, different calibration methods, including preprocessing transformations (PPTs) and regression algorithms (RAs), considerably affect the performance of prediction models. In this study, raw spectrum and 21 PPTs, combined with three RAs, for a total of 66 calibration methods, were investigated for modeling and predicting soil NPK using hyperspectral VNIR data (400–1000 nm). The ratio of performance to deviation (RPD) of validation set was selected to evaluate the prediction accuracy and the ratio between the interpretable sum squared deviation and the real sum squared deviation (SSR/SST) of the validation set was also used to evaluate the explanatory power of the models. It was found that there is a tradeoff between RPD and SSR/SST values; under this tradeoff, the multiplicative scatter correction, combined with the back-propagation neural network, was preferred for predicting P (RPD = 2.23, SSR/SST = 0.81). The Savitzky-Golay filtering + logarithmic transformation, combined with the partial least squares – regression, was preferred for predicting K (RPD = 1.47, SSR/SST = 0.95). However, with extremely low RPD and SSR/SST values, the prediction of N was unreliable in this study. The evaluation approach presented in this paper suggests a framework for choosing a calibration method for spectroscopy analysis for predicting soil NPK and perhaps some other properties.

1. Introduction

Soil available nutrients, including available nitrogen, phosphorus, and potassium (NPK), playing an important role in enhancing soil fertility and plant productivity for the growth and development of agricultural systems. Wheat and corn are important grain crops in Anhui Province, China, and in many other regions across the globe with high NPK demand. However, excessive fertilization not only affects soil fertility and increases the economic investment, but it also leads to environmental pollution (e.g., Chen et al., 2009; Sadowski et al., 1987; Savci, 2012). It is essential to improve the efficiency and accuracy of soil available nutrient detection for the reasonable fertilization and the sustainable development of agricultural systems. Traditional laboratory methods for quantifying NPK are expensive and time-consuming, and thus cannot meet the requirements of modern soil quality assessment and management, particularly with respect to precision agriculture. Alternatively, previous studies have suggested the reflectance spectroscopy analysis approach as a rapid, non-destructive, reproducible, and cost-effective analytical method for assessing soil properties (Ben-Dor and Banin, 1995).

Reflectance spectroscopy is used in chemometrics to construct soil spectral classification and regression models to predict many soil attributes. Several soil properties with high concentration have a specific spectral absorption signal that can be well predicted with the reflectance spectroscopy analysis approach. For example, soil water content has significant absorption bands around 1400 and 1900 nm (Stoner and Baumgardner, 1980). Organic matter has broad sensitive bands from the visible to the shortwave infrared range (350–2500 nm) due to the overtones and combination absorptions of O—H, C—H, and N—H bonds (Clark et al., 1990). The most common sensitive bands associated with clay minerals are the 1400–1410 and 2160–2200 nm due to the metal—OH band plus the O—H stretch combination and C—O (Galvão et al., 1997; Hunt and Salisbury, 1970). Unfortunately, soil NPK do not have any obvious spectral feature and usually exist in low

---

* Corresponding author.
E-mail addresses: shwli@ahau.edu.cn, shwli2017@163.com (S. Li).

http://dx.doi.org/10.1016/j.still.2017.09.006
Received 7 August 2017; Received in revised form 10 September 2017; Accepted 14 September 2017

Table 1
Summary of previous research results for predicting NPK by various calibration methods.

Literature	Spectral range	Calibration method		Nutrient	Results
		PPT	RA		
Confalonieri et al. (2001)	1100–2498 nm	LG	Modified PLS-R	P	$R^2 = 0.23–0.57$
				K	$R^2 = 0.48–0.82$
Viscarra Rossel et al. (2006)	400–14286 nm	LG	PLS-R	NO_3^--N	$R^2_{adj} = -0.02–0.00$
				P	$R^2_{adj} = 0.01–0.20$
				K	$R^2_{adj} = 0.29–0.47$
Mouazen et al. (2010)	350–2500 nm	5 points averaged + normalization + SG + FD	BPNN-LVs	P/K	$R^2 = 0.68–0.74$, RPD = 1.77–1.94
Shao and He (2011)	800–2500 nm	SG + MSC	LS-SVM	N	r = 0.90
				P	r = 0.83
				K	r = 0.83
	4000–400 cm^{-1}	SG + MSC	LS-SVM	N	r = 0.87
				P	r = 0.88
				K	r = 0.89
Gholizade et al. (2013)	700–2500 nm	SG + SD	SMLR	P	$R^2 = 0.52$
				K	$R^2 = 0.48$
Wu et al. (2014)	350–2500 nm	MSC + FD	Global BPNN	N	r = 0.78
				P	r = 0.65
				K	r = 0.86
			Local BPNN	N	r = 0.90
				P	r = 0.82
				K	r = 0.94
Ji et al. (2014)	350–2500 nm	LG + SG	PLS-R	N	$R^2 = 0.86$, RPD = 2.49
				P	$R^2 = 0.29$, RPD = 1.17
				K	$R^2 = 0.07$, RPD = 0.77
			LS-SVM (in situ spectrum)	N	$R^2 = 0.76$, RPD = 1.91
				P	$R^2 = 0.36$, RPD = 1.27
				K	$R^2 = 0.14$, RPD = 0.91
Paz-Kagan et al. (2014), Paz-Kagan et al. (2015)	350–2500 nm	SG + auto scale + generalized least squares weighting	PLS-R	NH_4^+-N	$R^2 = 0.27–0.83$, RPD = 1.69–2.43
				NO_3^--N	$R^2 = 0.74–0.82$, RPD = 1.76–2.68
				P	$R^2 = 0.21–0.74$, RPD = 0.53–1.98
				K	$R^2 = 0.61–0.76$, RPD = 2.00–2.25
Hu et al. (2016)	350–2500 nm	LG + normalization + 5 points averaged	PLS-R	P	$R^2 = 0.35$, RPD = 1.20
				K	$R^2 = 0.42$, RPD = 1.24
		LG + normalization + 5 points averaged + DOSC	PLS-R	P	$R^2 = 0.63$, RPD = 1.64
				K	$R^2 = 0.66$, RPD = 1.72
Sarathjith et al. (2016)	350–2500 nm	FD + WT	SVM	P	$R^2 = 0.80$, RPD = 2.27
				K	$R^2 = 0.71$, RPD = 1.89
Yu et al. (2016)	350–2500 nm	SG + LG + FD	SMLR + spectral index	NH_4^+-N	$R^2 = 0.74–0.92$, RPD = 1.89–3.49
				NO_3^--N	$R^2 = 0.33–0.53$, RPD = 1.02–1.37
				P	$R^2 = 0.44–0.51$, RPD = 1.29–1.37
				K	$R^2 = 0.51–0.95$, RPD = 1.32–4.26
			PLS-R + spectral index	NH_4^+-N	$R^2 = 0.78–0.92$, RPD = 2.06-3.40
				NO_3^--N	$R^2 = 0.30–0.73$, RPD = 0.95–1.89
				P	$R^2 = 0.39–0.51$, RPD = 1.17–1.38
				K	$R^2 = 0.37–0.71$, RPD = 1.15–1.59
Shaddad et al. (2016)	350–2500 nm	3 points averaged + normalization + SG + FD	PLS-R	P	$R^2 = 0.77$, RPD = 2.08
				K	$R^2 = 0.55$, RPD = 1.48

Abbreviations used: preprocessing transformation (PPT); regression algorithm (RA); logarithmic transformation (LG); first derivative (FD); Savitzky-Golay filtering (SG); multiplicative scatter correction (MSC); wavelet transformation (WT); direct orthogonal signal correction (DOSC); partial least squares − regression (PLS-R); back-propagation neural network (BPNN); least squares − support vector machine (LS-SVM); stepwise multiple linear regression (SMLR); support vector machine (SVM); available nitrogen (N); available phosphorous (P); available potassium (K); ammonium nitrogen (NH_4^+-N); nitrate nitrogen (NO_3^-—N); coefficient of determination (R^2); correlation coefficient (r); the ratio of performance to deviation (RPD).

concentrations in the soil (Ji et al., 2014). Consequently, their identification using spectral approach (calibration and prediction) is difficult to achieve. In addition, the existence of unexpected irrelevant information in spectra also greatly affects the performance of calibration models for quantifying soil NPK.

Calibration methods, including different combinations of preprocessing transformations (PPTs) and regression algorithms (RAs) have been widely applied to improve the prediction accuracy of soil NPK. PPTs, based on various mathematical functions, can be used to correct for non-linearity, measurement and sample variations, and