Abstract

Evaluation of partial least squares, least squares support vector machines and locally weighted regression for near infrared and mid-infrared quantitative analysis of dried and field moist soils

J.B. Reeves III,[a,c,*] B. Igne,[b] G.W. McCarty,[c] W.D. Hively,[c] E. Lund[d] and C.R. Hurburgh Jr[b]

[a]*EMBUL, BARC East, Beltsville, MD, USA. E-mail: james.reeves@ars.usda.gov*
[b]*Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa, USA*
[c]*HRSL, BARC West, Beltsville, MD, USA*
[d]*Veris Technologies, 601 Broadway, Salina, Kansas, USA*

Introduction

Soil testing requires analysing large numbers of samples that is time consuming and expensive. Mid-infrared (mid-IR) spectroscopy and near-infrared (NIR) spectroscopy are fast, non-destructive, and inexpensive methods that have been used for soil analysis. A comparison of the use of spectral pretreatment as well as the implementation of linear and non-linear regression methods was performed. This study presents an overview of the use of infrared spectroscopy for the prediction of five physical (sand, silt, and clay) and chemical (total carbon and total nitrogen) parameters with NIR and mid-IR units in bench top, portable and field setups. In addition, samples were also scanned in both the NIR and mid-IR wavelength regions on selected instruments, using field moist samples.

Materials and methods

A set of 314 samples collected from 5 bare-soil fields (0–20 cm, recently plow-tilled and vegetation-free) were used. Sampling occurred in transects (20 m apart) corresponding to transects of tillage-based

NIR spectrometer (Veris On-The-Go, 350 nm to 2225 nm, average resolution of 8 nm) measurements. Soil samples were also dried (50°C, two days), crushed (hammer mill) and ground (roller mill), and spectra were collected in the NIR and mid-IR regions on a Digilab FT-IR (FTS7000, 64 co-added scans, 4 cm^{-1} resolution, designated as bench-FT-NIR and bench-FT-mid-IR). A portable FT-IR instrument (SOC-400) was similarly used to collect spectra (8 cm^{-1} resolution). Samples were assayed for sand, silt, and clay by hydrometer, and total C and N by combustion. For the dried samples, every third sample based on field spatial distribution was used for validation. Two types of spectral pretreatment were used: (1) derivative (Savitzky-Golay), normalisation (unit area under curve), and scatter corrections methods (standard normal variate (SNV), multiplicative scatter correction (MSC), extended MSC, and loopy MSC and EMSC; and (2) Fourier and wavelet decompositions (tuned by iterative processes) of spatially pretreated spectra. The best performing preprocessing methods (or combination of preprocessing methods) found when developing PLS models were used to complement frequency based pretreatment methods. Partial least squares, least squares support vector machines (LS-SVM) a non-linear regression technique, and locally weighted regression (LWR), a local linear method, were used to develop calibration models. Spatial preprocessing methods used for LS-SVM and LWR were the best performers found with PLS. Auto scaling (mean zero and unit variance) was used to scale spectra, after pretreatment methods, before developing all regression models. Standard error of prediction (*SEP*) was used to evaluate the precision of each model and r^2 was used to evaluate model fit.

Results and Discussion

For dried soil samples, even though no significant differences existed among pretreatment methods, models using second derivatives performed better. Calibration models (Figure 1) showed that the LS-SVM did not outperform linear methods for most components while LWR that creates simpler models performed well.

The present results tend to show that soil models are quite sensitive to the complexity of the model. The ability of LWR to select only the appropriate samples did help in the development of robust models. Results also proved that field units performed as well as bench top instruments. This was true for both NIR and mid-IR. However, efforts with field moist soils showed that no method produced robust calibrations when using mid-IR spectra in contrast to the NIR where robust calibrations were found (Data not presented).

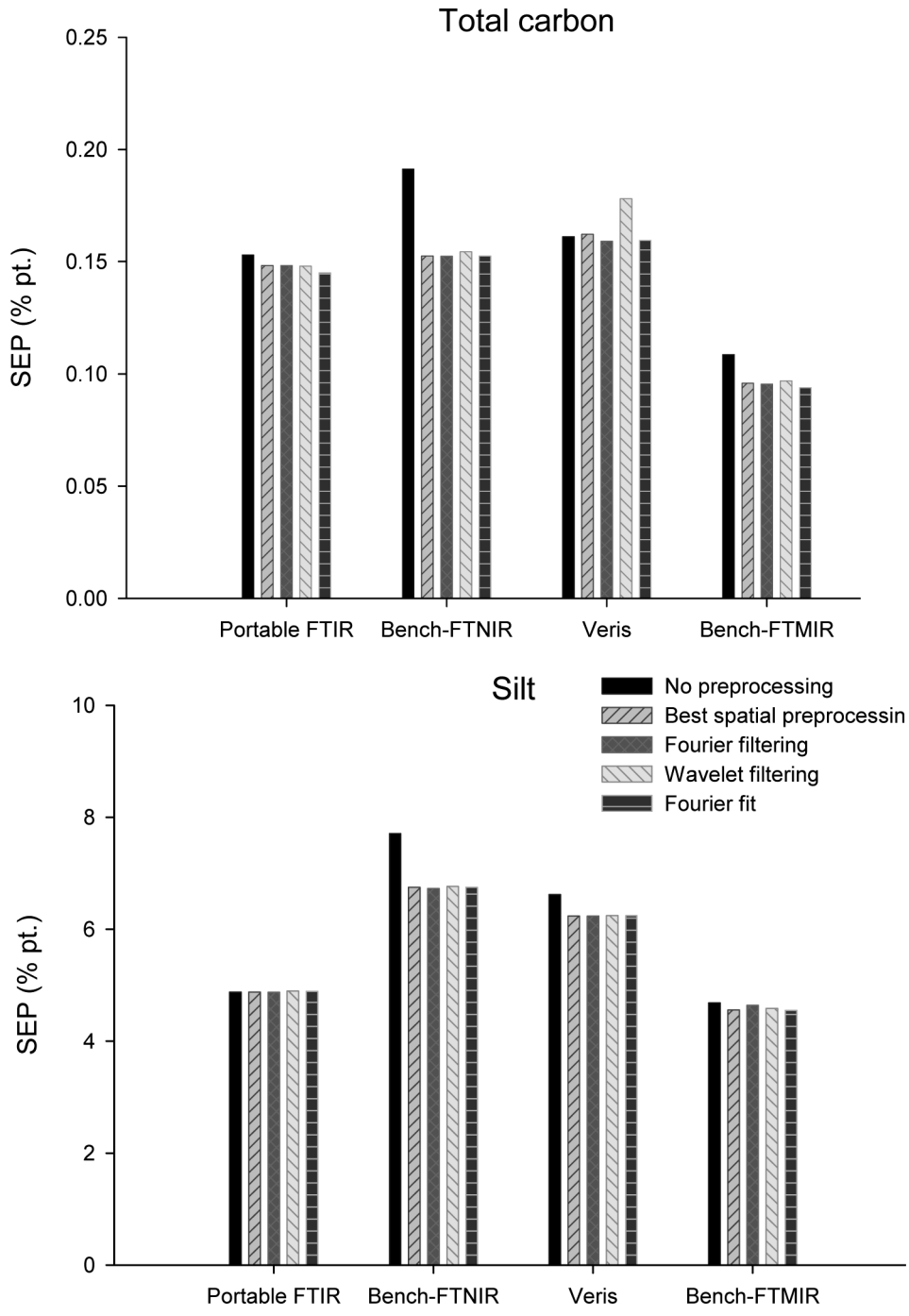

Figure 1. Comparison on spatial and frequency based preprocessing methods.