In Situ Characterization of Soil Clay Content with Visible Near-Infrared Diffuse Reflectance Spectroscopy

Travis H. Waiser, Cristine L.S. Morgan*, David J. Brown, and C. Tom Hallmark

T.H. Waiser, C.L.S. Morgan, and C.T. Hallmark Dep. of Soil and Crop Sciences, Texas A&M University, 2474 TAMU, College Station, TX 77843-2474. D. J. Brown, Dep. of Land Resources and Environmental Sciences, Montana State University, PO Box 173120, Bozeman, MT 59717-3120. Received 06 June 2006. *Corresponding author (cmorgan@ag.tamu.edu).

KEYWORDS
Diffuse reflectance spectroscopy; in situ; VNIR; PLS regression; clay content; water potential

ACKNOWLEDGEMENTS
The authors would like to thank the Texas Water Resource Institute and Texas Agricultural Experiment Station for their financial support. The extension agents from Erath and Comanche County, Texas were instrumental in putting us in contact with landowners of the study sites. We appreciate the work the Texas Agricultural Experiment Station Soil Characterization Laboratory did in running some of the analyses especially Dr. Richard Drees and Ms. Donna Prochaska.
**ABSTRACT**

Visible and near-infrared (VNIR, 400-2500 nm) diffuse reflectance spectroscopy (DRS) is a rapid, proximal-sensing method that has proven useful in quantifying constituents of dried and ground soil samples. However, very little is known about how DRS performs in a field setting on soils scanned *in-situ*. The overall goal of this research was to evaluate the feasibility of VNIR-DRS for *in situ* quantification of clay content of soil from a variety of parent materials. Seventy-two soil cores were obtained from six fields in Erath and Comanche Counties, Texas. Each soil core was scanned with a visible near-infrared spectrometer, with a spectral range of 350-2500 nm, at four different combinations of moisture content and pre-treatment: field-moist *in situ*, air-dried *in situ*, field-moist smeared *in situ*, and air-dried ground. The VNIR spectra were used to predict total and fine clay content of the soil using partial least squares (PLS) regression. The PLS model was validated with 30% of the original soil cores that were randomly selected and not used in the calibration model. The validation clay predictions had a root mean squared deviation (RMSD) of 61 g kg\(^{-1}\) and 41 g kg\(^{-1}\) dry soil for the field-moist and air-dried *in situ* cores, respectively. The RMSD of the air-dry ground samples was between the two *in situ* RMSDs and comparable to values in the literature. Smearing the samples increased the field-moist *in situ* RMSD to 74 g kg\(^{-1}\). Whole-field holdout validation results showed that soils from all parent materials need to be represented in the calibration samples for maximum predictability. In summary, DRS is an acceptable
technique for rapidly measuring soil clay content in situ at variable water contents and parent materials.

Abbreviations: DRS, diffuse reflectance spectroscopy; N, the total number of samples in the validation data; PLS, partial least squares; RMSD, root mean squared deviation; RPD, ratio of standard deviation to RMSD; SD, standard deviation; VNIR, visible and near-infrared; $Y_{\text{pred}}$, predicted values of the validation set using the PLS model; and $Y_{\text{meas}}$, laboratory measurements of the validation set.

INTRODUCTION

The resolution of a 1:24,000 scale soil survey map is too coarse to capture soil variability within soil mapping units and transitions between soil mapping units. However, soil maps that capture soil variability at a 10-50-m scale are necessary for resource management such as, precision agriculture, non-point source pollution modeling, and resource use planning (Pachepsky et al., 2001; Ellert et al., 2002). Researchers have used external landscape features such as terrain modeling, soil surface reflectance, vegetative cover, and rapid surface sensing techniques like electromagnetic devices to capture the spatial variability of soil properties across landscapes followed by collecting soil cores and laboratory analyses of those cores for model calibration and validation (Moore et al., 1993; Sudduth et al., 1997; Zhu et al., 1997, McBratney et al., 2003, Zhu et al., 2004). Each of these methods has the advantage of creating a high-resolution map of horizontal soil variability, but is limited in the ability to get high-resolution, vertical soil information. Soil profile (vertical) information is limited by the
capacity to collect and analyze soil cores. Model calibration and validation are currently
the weak points in digital soil mapping because collecting and analyzing soil cores to
capture vertical variability is time consuming and cost prohibitive. For example, current
laboratory soil analyses can take several weeks to months for each soil pedon and cost
upwards of $2,000. The lack of soil sensors that can rapidly quantify soil profile
information demonstrates the need to find new methods for soil mapping.

A proximal soil sensor that might be useful for rapidly quantifying soil profile
information in the field is visible and near-infrared diffuse reflectance spectroscopy
(VNIR-DRS). Recent research has shown the effectiveness of VNIR-DRS in providing a
non-destructive rapid prediction of soil physical, chemical, and biological properties of
air-dried ground soil samples in the laboratory (Shepherd and Walsh, 2002). In the VNIR
spectrum, absorptions by water bonds associated with clay content and other bonding
associated with clay type provide the opportunity to use VNIR for quantifying clay
information in soil. Previous research on air-dried, ground soil samples has shown VNIR
predictions of soil clay content with $r^2$-values ranging from 0.56 to 0.91 and RMSD’s
ranging from 23 to 11 g kg$^{-1}$ (Ben-Dor and Banin, 1995; Janik et al., 1998, Shepherd and
Walsh, 2002; Islam et al., 2003; Sorensen and Dalsgaard, 2005; Brown et al., 2006). Air-
drying and grinding the soil sample prior to scanning is believed to improve prediction
accuracy of DRS-based models. Air-drying the sample reduces the intensity of bands that
are related to water so signals associated with other soil properties are not masked or
hidden. Additionally, particle size affects accuracy of the spectral scan. Smaller particles
increase reflectance scatter, which reduces the absorption peak height (Workman and
Shenk, 2004). Soil homogenization, accomplished while grinding, may also be beneficial when scanning with VNIR-DRS.

The determination of clay content by VNIR measurements is possible in part due to the distinctive spectral signatures of common clay minerals. The spectral signatures include overtones and combination bands that occur from chemical bonds within soil minerals (Hunt and Salisbury, 1970; Clark, 1999). For example, kaolinite, smectite, and muscovite occur in the clay fraction of soils and have distinct spectral absorption features. Kaolinite \([\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4]\) is an aluminum silicate with two very strong hydroxyl bands near 1400 nm, OH stretch, and 2200 nm, OH stretch Al-OH bend combination (Hunt and Salisbury, 1970; Clark et al., 1990). Smectite (montmorillonite) \([\text{M}^{+0.3}\text{Al}_{2.7}\text{Si}_{3.3}\text{O}_{10}(0\text{H})_2; \text{M}=\text{Ca}^{2+}, \text{Mg}^{2+}, \text{K}^+, \text{etc.}]\) has two very strong water bands around 1400 and 1900 nm from molecular water and another at 2200 nm (Hunt and Salisbury, 1970; Goetz et al., 2001). Muscovite \([\text{KAl}_3\text{Si}_3\text{O}_{10}(\text{OH})_2]\) displays hydroxyl bands at 1400 nm and between 2200 to 2600 nm (Hunt and Salisbury, 1970). Brown et al. (2006) built a VNIR calibration for clay minerals and was able to predict kaolinite and montmorillonite within one ordinal unit from X-ray diffraction data 96% and 88% of the time, respectively. Goetz et al. (2001) investigated using NIR spectroscopy to measure smectite content in selected Colorado soils \((r=0.83)\). Because the literature shows VNIR absorbance associated with clay mineralogy, one may expect that the wavebands significant to predicting soil clay content to be similar to mineralogy wavebands.

There have been few reported studies of \textit{in situ} VNIR-DRS soil characterization. Sudduth and Hummel (1993) tested a portable near-infrared (NIR) spectrometer to measure soil properties \textit{in situ} on the fly. Their research concluded that VNIR-DRS
could be used to accurately predict cation exchange capacity, soil organic matter, and moisture content in the laboratory, but the technique was not sufficiently accurate at quantifying soil organic matter in a furrow. The poor prediction with this in situ technique was attributed to movement of the soil past the spectrometer. Hence low prediction accuracy for in situ scanning may be avoided by holding the spectrometer scanning device more stable, and by holding the soil stationary while scanning.

In the field, DRS measurements may not work as well as laboratory DRS measurements because of potential problems that have not been addressed by research on air-dried ground soils. The potential problems that may reduce the prediction accuracy of in situ DRS analysis include, varying amounts of soil moisture, varying particle size and aggregation, smearing of soil surfaces, small scale heterogeneity (mottles, accumulations, and redox features), and geographic regionality of calibration models. Sudduth and Hummel (1993) indicated that local buried residue and roughness of the soil surface caused a reduction in their prediction accuracies. Smearing of the soil causes reflectance properties to change, possibly altering prediction accuracies. Geographic regionality, such as changes in soil parent material, may affect prediction accuracies when using VNIR-DRS (Ben-Dor and Banin, 1995; Dalal and Henry, 1986; Sudduth and Hummel, 1996).

The overall goal of our study is to evaluate the feasibility of VNIR-DRS for in situ quantification of clay content for soil profiles from a variety of parent materials. Specifically, this research addresses the following objectives: 1) evaluate the precision of 350 to 2500 nm (VNIR region) soil reflectance measurements in quantifying soil clay content of in situ soils at field-moist and air-dry water content; 2) quantify any change in
measurement error or prediction accuracy for *in situ*, field-moist soil with a smeared surface; and 3) quantify any change in prediction accuracies with consideration to field-specific calibrations.

This research will help evaluate the feasibility of using a VNIR spectrometer in the field as a proximal sensor to aid soil mapping activities. For VNIR-DRS to be considered useful in the field, it must be an improvement on current field techniques. The VNIR DRS methods should be reliable, meaning that the errors should be comparable to current field techniques with similar data acquisition speeds, and VNIR-DRS should be rapid, meaning that it should quantify clay much faster than any current techniques. Although time and money are required to calibrate the VNIR instrument, once calibrated, VNIR-DRS can scan whole soil pedons more rapidly than hand texturing in the field or laboratory particle-size methods (i.e. pipette and hydrometer), allowing for more soil core information to be quantified in one day. Therefore VNIR-DRS is considered adequate if VNIR-DRS can quantify clay content accurate enough for precision management, watershed modeling needs, or other applications where soil information is need. In general, most precision management strategies and watershed models require a stronger emphasis on the change in soil texture in space (vertical and horizontal) than on lab-quality soil data (Morgan et al., 2003). Quickly collecting higher spatial resolution data is where VNIR-DRS has an advantage over current field techniques.

**MATERIALS AND METHODS**

**Soil Coring**
Seventy-two soil cores were collected from six fields in Erath County (fields 1, 2, 3, and 6) and Comanche County (fields 4 and 5), Texas in May 2004 using a Giddings hydraulic soil sampler (Windsor, CO) attached to a truck. Each soil core was collected to a maximum depth of 105 cm or to the depth of a coring-restrictive horizon. The soil cores were contained in a 6.0-cm diameter plastic sleeve that was capped on both ends. The soil cores that were collected were chosen to represent the large variability in soil properties over the two counties. For example, 21 soil series were mapped by the Natural Resources Conservation Service in these fields (Table 1), and the parent material of these soils included alluvium, sandstone, shale, and limestone. The large variability was selected to ensure that the VNIR prediction models would not be mineralogy specific. Additionally, though many cores represented the same soil series, variability within each soil series because of 1) differences of morphology within a series, 2) inclusions of other soils within each mapping unit, and 3) varying landscape positions all lead to representation of high morphological variability within each mapped series.

**VNIR-DRS Scanning**

The soil cores were scanned using an ASD “FieldSpec® Pro FR” VNIR spectroradiometer (Analytical Spectral Devices, Boulder, CO) with a spectral range of 350-2500 nm, 2-nm sampling resolution and spectral resolution of 3 nm at 700 nm and 10 nm at 1400 and 2100 nm. The spectroradiometer was equipped with a contact probe. The contact probe has a viewing area defined by a 2-cm diameter circle and its own light source. A Spectralon® panel with 99% reflectance was used to optimize the spectrometer each day; the same panel was used as a white reference before scanning each core. The 72 soil cores were prepared for the first, moist *in situ*, scan by slicing the
plastic sleeve lengthwise with a utility knife, and then halving the soil core lengthwise (surface to subsoil) with a piano wire. One-half of the core was smeared using a stainless steel spatula to simulate possible smearing by a soil probe and the other half was left unsmeared. The smearing test was performed to quantify any reduction in accuracy that might occur if a soil probe, pushed into the ground, smeared the sides of the hole. In this laboratory exercise, smearing moist, high clay content soil required multiple passes with the spatula. Therefore, this test was a worst-case scenario, as the soil was smeared to its maximum capacity. A wire grid was used to identify two 3-cm wide columns and 3-cm wide rows within each core half (Fig. 1). Each row within each column was scanned twice with the FieldSpec® Pro FR with a 90° rotation of the contact probe between scans. Both halves of each core, smeared and unsmeared, were scanned at field-moist water content.

Water potential was measured at each soil horizon from each core, with a maximum of six samples per core, using a SC-10 thermocouple psychrometer (Decagon, Pullman, WA) (Rawlins and Campbell, 1986). The water potential samples were collected immediately after scanning each core to minimize drying. Water potential measurements were made to confirm variability of water content and to quantify the range of soil moisture among the soils scanned. The cores were then placed in a drier at 44°C for two days for air-dried in situ scans. Before taking air-dried scans, the cores were removed from the oven and left on the countertop in the laboratory to equilibrate to room temperature. The soil core half that was scanned in unsmeared condition was then rescanned, in situ, at air-dry moisture content.
Each row of the unsmeared soil core was ground and passed through a 2-mm sieve, and re-scanned as air-dried, ground soil. The air-dried, ground soils were scanned from below with an ASD mug lamp connected to the FieldSpec® Pro FR. A Spectralon® 99% reflectance panel was used to optimize and white reference the spectrometer. Approximately 28 g of ground soil was placed into a Duraplan® borosilicate optical-glass Petri dish. Each sample was scanned twice with a 90º rotation between scans.

**Pretreatment of Data**

The reflectance spectra were spliced where the three detectors overlapped using an empirical algorithm. Reflectance for the 0º and 90º scans were averaged (mean). Scans on the same row, column 1 and 2, were also averaged, creating a mean of four scans for each measured value of clay content. Smoothed reflectance and 1st and 2nd derivatives were extracted on 10-nm intervals (360-2490 nm) from a weighted cubic smoothing spline fit to the mean reflectance data using the R “smooth spline” function (R Development Core Team, 2004), following the methods of Brown et al. (2006).

**Laboratory Analysis**

Although every 3 cm of the soil core was scanned, only selected rows were used for laboratory particle size analysis. Only the VNIR scans from those rows were subsequently used in the VNIR models. Whole rows of soil where combined for laboratory analysis; hence, each row of soil had a total of four scans to represent its spectral properties (Fig. 1). Rows were chosen to represent the primary horizons within a soil core. The entire row from each half of the core was used for particle size analysis. Particle size distribution was determined in the laboratory using the pipette method with an error of ±1% clay (Steele and Bradfield, 1934; Kilmer and Alexander, 1949; Gee and
Or, 2002). Fine clays were measured using centrifugation, USDA NRCS method 3A1b (USDA, 1996). Eighteen samples were selected for determination of clay mineralogy. Clay mineralogy by X-ray diffraction was completed following techniques described in Hallmark et al. (1986); no pretreatment was done. The mineralogy technique used was to determine the absence or presence of clay minerals existing in quantities of greater than 5% by volume of clay fraction.

**Model Calibration and Validation**

Four types of clay prediction models were built to help determine how in situ scans and variable water content affect prediction accuracies for clay content. Air-dried ground, air-dried in situ, field-moist in situ, and smeared field-moist in situ scans were the four soil preparation types used in the models. The clay content models were produced using 70% of the soil cores randomly chosen as the calibration samples. Scans from an individual soil core were not split between validation and calibration datasets. Entire cores were selected for validation or calibration to maintain independence between the calibration and validation data (Brown et al., 2005). For each of the four pretreatment scans (air-dried ground, air-dried in situ, field-moist in situ, and smeared field-moist in situ), three prediction models for clay content were built using soil reflectance, 1st derivative of reflectance and 2nd derivative of reflectance. The calibration models were built using segmented cross validation PLS method in Unscrambler 9.0 (CAMO Tech, Woodbridge, NJ). The segments for cross validation are randomly chosen and represent four percent of the calibration dataset. The remaining 30% of the cores were used to validate the models.
Model calibration and validation sets were also completed on whole-field holdouts of moist *in-situ* scans, as a means to compare to the findings of Brown et al. (2005). Whole-field holdouts were achieved by calibrating a model using PLS with five of the six fields. The sixth field was held out as the validation samples. Six models were created so all six fields were represented as a validation set.

The significant wavelengths in each model were plotted to help determine what portions of the spectra were important for clay content predictions. Significant wavelengths were chosen by Unscrambler 9.0 by comparing Student t-values of the PLS regression coefficients to a p-value of 0.05. Negative clay content predictions were changed to zero clay content before comparison of measured to predicted clay.

Measured versus predicted values of the validation samples were compared using simple regression. The coefficient of determination ($r^2$), root mean squared deviation (RMSD), ratio of standard deviation (SD) to RMSD (RPD) and bias were calculated to compare the accuracy of different PLS models. Statistical formulas to calculate RMSD and bias follow Gauch et al. (2003), Brown et al. (2005) and RPD follows Chang et al. (2005):

\[
\text{RMSD} = \sqrt{\frac{\sum_{n} (Y_{\text{pred}} - Y_{\text{meas}})^2}{N}},
\]

\[\text{RPD} = \frac{\text{SD}}{\text{RMSD}},\]

\[
\text{Bias} = \frac{\sum_{n} (Y_{\text{pred}} - Y_{\text{meas}})}{N};
\]

where $Y_{\text{pred}}$ are predicted values of the validation set using the PLS model, and $Y_{\text{meas}}$ are laboratory measurements of the validation set, and $N$ is the total number of samples in the validation data.
RESULTS AND DISCUSSION

Sample Descriptions

From the 72 soil cores, 270 soil samples were analyzed for clay content and water potential. Of these samples, 188 were used in the calibration model. The other 82 samples were used to validate the PLS model. Though selected randomly, the calibration and validation data sets were similar (Table 2). Clay content range of the 270 soil samples was 12 g kg\(^{-1}\) to 578 g kg\(^{-1}\) dry soil. The mean and median for the calibration and validation data were similar, indicating that the samples were evenly split above and below the mean, and that the prediction model produced should not be skewed. The minimum water potential measured, -5.8 MPa, in the calibration samples was from a tilled surface (0-3 cm) of a loamy-textured soil. The range of water potential confirms that in situ moist scans involved a large range of water contents that may be found in most field situations. The maximum water potential, 0 MPa, occurred in deep horizons in several of the irrigated fields. The samples represented a range of clay mineralogy. The minerals found in the clay fractions were: smectite, kaolinite, mica, quartz, calcite, and feldspar (Table 3). The soils in field 1 were notably high in clay content and were dark colored Vertisols; field 3 soils were low in clay content.

Field-moist vs. Air-dried Validation

The 1\(^{st}\) derivative PLS model performed better than the reflectance and 2\(^{nd}\) derivative PLS models in the field-moist in situ scans. The \(r^2\) value for the 1\(^{st}\) derivative model was 0.83 compared to 0.71 and 0.58 for the reflectance and 2\(^{nd}\) derivative models, respectively. Because the 1\(^{st}\) derivative PLS model performed the best for field-moist in situ, and the 1\(^{st}\) derivative worked the best in other VNIR work (Reeves et al., 1999;
Reeves and McCarty, 2001; Brown et al., 2006), the remaining PLS models reported in this paper all use the 1st derivative of the VNIR spectra between 350 and 2500 nm. The prediction accuracy of three of the four models used to predict clay content were very similar, and the fourth, field-moist in situ smeared, had the largest prediction error and scatter about the validation regression. Table 4 summarizes the prediction accuracies between the four VNIR models. The $r^2$ value of the dried-ground samples were similar to the literature, while the RMSD value of this work was smaller than those cited (Ben-Dor and Banin, 1995; Janik et al., 1998, Shepherd and Walsh, 2002; Islam et al., 2003; Sorensen and Dalsgaard, 2005; Brown et al., 2006). The air-dried ground model was expected to produce the most accurate prediction model because the sample was reduced to a uniform size, moisture was uniform, and grinding homogenized the soil, but the air-dried in situ model did slightly better (Fig. 2). There are two possible explanations for why the air-dried ground model did not perform as well as the air-dried in situ model. One explanation could be random error. The second possible explanation is that in situ soil has a higher bulk density than dried ground soil; and therefore, the in situ soil may have a stronger reflectance signal. The air-dried ground and in situ moist models had a similar number of wavelengths with significant (p-value $\leq 0.05$) regression coefficients, 83 and 78 significant wavelengths, respectively (Fig. 3).

The air-dried ground prediction had the most bias (-16.3), compared to the other predictions. The large bias associated with the air-dried ground prediction was attributed to 60% of the samples being under predicted in clay content and large discrepancies between measured and predicted clay content of soils greater than 450 g kg$^{-1}$ clay. Because the air-dried in situ model slightly outperformed the air-dried ground model
using the 1\textsuperscript{st} derivative of VNIR reflectance spectra, we conclude natural soil heterogeneity, as a result of clay films, translocations, redox features etc., and variability associated with aggregation had little effect on model predictions. Gaffey (1986) found that aggregate size had little effect on prediction accuracy, but Gaffey was looking at carbonate sizes. His results showed a difference in reflectance, but the number of bands, band positions and width, and relative band intensities were unchanged.

Another significant finding was that the clay content model using the field-moist \textit{in situ} model had slightly less prediction accuracy than the air-dried \textit{in situ} model. However, the field-moist \textit{in situ} model performed just as well when compared to the air-dried ground model (Fig. 2). These results indicate that the amount of water in the soil sample did not change the prediction accuracy compared to air-dried laboratory situations. These results show that soil water may reduce accuracy somewhat (increased RMSD by 20 g kg\(^{-1}\)), but this reduction in accuracy was no more than air-drying, grinding, and scanning the soil through a borosilicate Petri dish (Fig. 2). Chang et al. (2005) showed that r\(^2\)-values decreased slightly from 0.79 to 0.76 from an air-dried soil to a moist soil, respectively, which agrees with other findings. The field-moist \textit{in situ} models used 102 significant wavelengths in determining clay content (Fig. 3). Figure 3 shows the 32 significant wavelengths that were common in all three models.

The same set of VNIR models was created to predict fine clay. Fine clay models have lower RMSD values than total clay models but the RPD values are similar to the total clay models (Table 4). The similar RPD values indicate that the smaller RMSD values for fine clay models were due to smaller standard deviations (smaller range); so the total clay and fine clay models performed similarly. VNIR-DRS has proven
effective in identifying smectitic clays (Brown et al., 2006) and studies suggest that Vertisol fine clays are comprised primarily of smectites (Anderson et al., 1972; Yerima et al., 1989). Therefore, our ability to predict fine clay in this study might be due to the mineral signature of this size fraction.

**Field-moist Smeared Validation**

Smearing of the field-moist *in situ* cores reduced the accuracy of the prediction model (Fig. 2). The decrease in the prediction accuracy of the smeared cores was probably due to the change in reflectance properties of the soil. Smearing causes the soil surface to become shiny, which changes the reflectance from diffuse to diffuse specular reflectance. Specular reflectance can mask diffuse absorptions and otherwise confuse the analysis of reflectance spectra (Coates, 1998). When comparing the significant wavelengths between the field-moist *in situ* and field-moist *in situ* smeared cores, a number of wavelengths found in the visible portion of the spectra are absent from the smeared core (Fig. 4). The smeared core model had 70 significant wavelengths and the unsmeared core model had 102, with 50 of these wavelengths being common in both models. The absence of these wavelengths may be the source of loss in prediction accuracy. In a field situation, the loss of prediction accuracy due to smearing might be less because it took multiple passes with the knife to smear many of the soil cores.

**Whole-field Holdout Validation**

As expected (Brown et al., 2005), whole-field cross-validation yielded degraded predictions relative to random cross-validation. For all six fields, whole field validation statistics (Table 5, RMSD and RPD) were worse than for the randomly selected cross-validation results (Table 4). The RPD statistics are particularly telling, with fields 1 and
having RPD values < 1 (no predictive value), fields 4-6 having RPD values < 1.5 (little predictive value), and only field 2 yielding acceptable results (RPD = 1.92).

Both field 1 and 3 contained soil at the extremes for clay content. Field 1 consisted of high clay soils (median 46 % clay) that were dark in color through the entire length of the soil core—high clay contents that were not represented in the calibration soils. Not surprisingly, the prediction bias for this field was -128.1 g kg⁻¹ (Table 5) indicating a significant underprediction of these high clay contents. Conversely, field 3 contained relatively low clay soils (median 3 % clay) not found in the other five fields. Regional clay calibrations will require a far larger and more diverse soil-spectral library than that presented in this study.

**DRS as a Proximal Sensor for Clay Content**

Figure 5 illustrates the performance of VNIR-DRS used as a proximal soil sensor for in-situ soils at field soil moisture. Three very different soil profiles are shown. The first is a Vertisol from field 1. This soil is very high in clay and somewhat uniform in clay content with depth. The VNIR-DRS prediction shows the same uniformity, with 10 % clay under prediction (Fig. 5a.). Nonetheless, the prediction does put the profile in the correct clay texture category. The second soil, Fig. 5b., is an Alfisol, with a sandy surface and an E horizon. In the case of the Alfisol, VNIR-DRS predictions show the continuous change of clay content, the A, E, and Bt horizons are clearly seen. The third soil is a Mollisol with secondary carbonate concentrations in the B horizon (Fig. 5c). The VNIR-DRS prediction of this core was encouraging, because the secondary carbonates did not appear add random scatter to the prediction. The concentrations may have biased the clay content prediction toward the bottom of the profile. In summary, though the
VNIR-DRS predictions do no perfectly agree with the particle size analysis in these three cores, the predictions do provide a rapid, continuous measurement of at least relative soil clay content fluctuations with depth.

**CONCLUSIONS**

In this study of 72 soil cores from Central Texas, we found a strong relationship between VNIR reflectance and clay content for air-dried *in situ* (RMSD = 41 g kg\(^{-1}\)), field-moist *in situ* (RMSD = 61 g kg\(^{-1}\)), and air-dried ground (RMSD = 62 g kg\(^{-1}\)) scans. Visible near-infrared DRS was capable of predicting soil clay content *in-situ* at varying water contents. There was a decrease in predictive accuracy for smeared field-moist *in situ* (RMSD = 74 g kg\(^{-1}\)) scans. However, we obtained worse results for ground and sieved samples than for the non-smeared *in situ* scans—suggesting that soil aggregates and natural heterogeneity did not degrade *in situ* predictions. Variable water contents did reduce predictive accuracy as evidenced by a comparison of air-dried *in situ* and field-moist *in situ* results, but the field-moist *in situ* were still slightly better than those obtained from scans of air-dried, ground-and-sieved samples. Validation RPD statistics were similar for fine clay and total clay predictions.

These results indicate that VNIR-DRS may be useful as a proximal soil sensor in field conditions. In particular, we envision a VNIR-DRS sensor placed in a soil probe for *in situ* soil profile characterization. Such a probe would give soil scientists the ability to acquire soil profile data (e.g. clay content) at greater spatial and vertical sampling densities than is practical with conventional soil excavation, extraction and characterization techniques. To make this vision a reality, continued research is needed
on *in situ* VNIR-DRS applications with greater soil diversity and a wider range of soil properties.

REFERENCES


U.S. Department of Agriculture. 1977. Soil survey of Comanche County, Texas. Soil


Fig. 1. Schematic of a vertically sliced soil core. Columns and rows indicate locations scanned using the contact probe for *in situ* scans. Dried ground soil samples represent the soil from one row, both columns.

Fig. 2. Predicted vs. measured clay content of the validation data set for (a) air-dried ground (b) air-dried *in situ* (c) field-moist *in situ* (d) field-moist *in situ* smeared. Clay content predictions were made from models built with partial least square regression using the 1st derivative of visible near-infrared reflectance (VNIR) spectra (350-2500 nm). RMSD is root mean squared deviation.

Fig. 3. The wavelengths that contributed to significant regression coefficients (p-value ≤ 0.05) in the prediction of clay content are shown for field-moist and air-dry *in situ* and air-dry ground models. The relative magnitude of each regression coefficient indicates the strength of the correlation. The top plot shows the mean of the regression coefficients common in all three models. All plots are on the same x-axis. Values of the y-axis are not shown, but all y-axes are on the same scale.

Fig. 4. The wavelengths that contributed to significant regression coefficients (p-value ≤ 0.05) in the prediction of clay content are shown for field-moist, smeared and unsmeared, *in situ* models. The relative magnitude of each regression coefficient indicates the strength of the correlation. The top plot shows the mean of the regression coefficients common in both models. All plots are on the same x-axis. Values of the y-axis are not shown, but all y-axes are on the same scale.
Fig. 5. Laboratory measured (+) and visible near-infrared (VNIR) predicted (◊) clay content for three soil cores, a) a dark-colored Vertisol, b) an Alfisol with distinct A, E, Bt horizons, and c) a Mollisol with secondary calcium carbonate concentrations in the B horizons. Laboratory measurements were taken to represent a row within a horizon, and VNIR predictions were made every 3-cm.
Table 1. Series and family classification of the soils mapped by the Natural Resources Conservation Service within the fields which were used for VNIR-DRS predictions (USDA, 1973; USDA, 1977; USDA, 2005).

<table>
<thead>
<tr>
<th>Soil series</th>
<th>Family classification</th>
<th>Field number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abilene</td>
<td>Fine, mixed, superactive, thermic Pachic Argiustolls</td>
<td>4</td>
</tr>
<tr>
<td>Altoga</td>
<td>Fine-silty, carbonatic, thermic Udic Haplustepts</td>
<td>6</td>
</tr>
<tr>
<td>Blanket</td>
<td>Fine, mixed, superactive, thermic Pachic Argiustolls</td>
<td>6</td>
</tr>
<tr>
<td>Bolar</td>
<td>Fine-loamy, carbonatic, thermic Udic Calciustolls</td>
<td>1,5,6</td>
</tr>
<tr>
<td>Bosque</td>
<td>Fine-loamy, mixed, superactive, thermic Cumulic Haplustolls</td>
<td>5</td>
</tr>
<tr>
<td>Brackett</td>
<td>Loamy, carbonatic, thermic, shallow Typic Haplustepts</td>
<td>5</td>
</tr>
<tr>
<td>Bunyan</td>
<td>Fine-loamy, mixed, active, nonacid, thermic Typic Ustifluvents</td>
<td>2</td>
</tr>
<tr>
<td>Chaney</td>
<td>Fine, mixed, active, thermic Oxyaquic Paleustalfs</td>
<td>4</td>
</tr>
<tr>
<td>Cisco</td>
<td>Fine-loamy, siliceous, superactive, thermic Typic Haplustalfs</td>
<td>4</td>
</tr>
<tr>
<td>Denton</td>
<td>Fine-silty, carbonatic, thermic Udic Calciustolls</td>
<td>1,6</td>
</tr>
<tr>
<td>Frio</td>
<td>Fine, smectitic, thermic Cumulic Haplustolls</td>
<td>1,4,5</td>
</tr>
<tr>
<td>Houston Black</td>
<td>Fine, smectitic, thermic Udic Haplusterts</td>
<td>1</td>
</tr>
<tr>
<td>Karnes</td>
<td>Coarse-loamy, carbonatic, thermic Typic Calciustepts</td>
<td>5</td>
</tr>
<tr>
<td>Lewisville</td>
<td>Fine-silty, mixed, active, thermic Udic Calciustolls</td>
<td>5,6</td>
</tr>
<tr>
<td>Maloterre</td>
<td>Loamy, carbonatic, thermic Lithic Ustorthents</td>
<td>1</td>
</tr>
<tr>
<td>Nimrod</td>
<td>Loamy, siliceous, active, thermic Aquic Arenic Paleustalfs</td>
<td>3,4</td>
</tr>
<tr>
<td>Pedernales</td>
<td>Fine, mixed, superactive, thermic Typic Paleustalfs</td>
<td>4</td>
</tr>
<tr>
<td>Purves</td>
<td>Clayey, smectitic, thermic Lithic Calciustolls</td>
<td>1,5,6</td>
</tr>
<tr>
<td>Selden</td>
<td>Fine-loamy, siliceous, active, thermic Aquic Paleustalfs</td>
<td>3</td>
</tr>
<tr>
<td>Selden</td>
<td>Fine-loamy, mixed, superactive, thermic Udic Calciustolls</td>
<td>5</td>
</tr>
<tr>
<td>Venus</td>
<td>Fine-loamy, mixed, active, thermic Udic Paleustalfs</td>
<td>2,3,6</td>
</tr>
<tr>
<td>Windthorst</td>
<td>Fine, mixed, active, thermic Udic Paleustalfs</td>
<td>2,3,6</td>
</tr>
</tbody>
</table>
Table 2. Clay content and water potential summary statistics for soil samples used in calibration and validation datasets.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>SD†</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Calibration samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total clay (g kg⁻¹)</td>
<td>188</td>
<td>12</td>
<td>525</td>
<td>255</td>
<td>139</td>
<td>262</td>
</tr>
<tr>
<td>Fine clay (g kg⁻¹)</td>
<td>188</td>
<td>3</td>
<td>380</td>
<td>138</td>
<td>85</td>
<td>140</td>
</tr>
<tr>
<td>Water potential (MPa)</td>
<td>188</td>
<td>-5.8</td>
<td>0.0</td>
<td>-0.49</td>
<td>0.60</td>
<td>-0.35</td>
</tr>
<tr>
<td><strong>Validation samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total clay (g kg⁻¹)</td>
<td>82</td>
<td>28</td>
<td>578</td>
<td>271</td>
<td>144</td>
<td>247</td>
</tr>
<tr>
<td>Fine clay (g kg⁻¹)</td>
<td>82</td>
<td>18</td>
<td>362</td>
<td>152</td>
<td>79</td>
<td>147</td>
</tr>
<tr>
<td>Water potential (MPa)</td>
<td>82</td>
<td>-2.2</td>
<td>0.0</td>
<td>-0.48</td>
<td>0.47</td>
<td>-0.38</td>
</tr>
</tbody>
</table>

† SD, standard deviation.

Table 3. Clay minerals present in soil cores from each field. Mineralogy is presented as a summary by field and not all cores are represented. Cores selected for mineralogy were chosen to represent clay minerals within a field.

<table>
<thead>
<tr>
<th>Clay Minerals</th>
<th>Vermiculite</th>
<th>Smectite</th>
<th>Kaolinite</th>
<th>Mica</th>
<th>Calcite</th>
<th>Quartz</th>
<th>Feldspar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field no.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>XXX†</td>
<td>X</td>
<td>XX</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>XX</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>t</td>
<td>X</td>
<td>XX</td>
<td>X</td>
<td></td>
<td></td>
<td>t</td>
</tr>
<tr>
<td>4</td>
<td>XX</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>XX</td>
<td>X</td>
<td>t</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† XXX, greater than 50% of mineral present; XX, 20-50% of mineral present; X, 5-20% of mineral present; t, trace amounts of mineral present.
Table 4. Prediction accuracies used for total clay and fine clay content models using the 1st derivative of visible near-infrared reflectance spectra (350-2500 nm) and partial least squares regression. Prediction accuracies were calculated by using entire soil cores (30% of the total data set) randomly held out of the calibration model. Also included are the number of principal components (PCs) used in each partial least squares regression model.

<table>
<thead>
<tr>
<th></th>
<th>r²</th>
<th>RMSD</th>
<th>RPD</th>
<th>Bias</th>
<th>PCs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>g kg⁻¹</td>
<td>g kg⁻¹</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total clay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air-dried ground</td>
<td>0.84</td>
<td>62</td>
<td>2.32</td>
<td>-16.3</td>
<td>9</td>
</tr>
<tr>
<td>Air-dried in situ</td>
<td>0.92</td>
<td>41</td>
<td>3.51</td>
<td>-2.0</td>
<td>10</td>
</tr>
<tr>
<td>Field-moist in situ</td>
<td>0.83</td>
<td>61</td>
<td>2.36</td>
<td>3.0</td>
<td>9</td>
</tr>
<tr>
<td>Field-moist smeared</td>
<td>0.75</td>
<td>74</td>
<td>1.95</td>
<td>7.4</td>
<td>6</td>
</tr>
<tr>
<td>Fine clay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air-dried ground</td>
<td>0.81</td>
<td>34</td>
<td>2.32</td>
<td>0.65</td>
<td>4</td>
</tr>
<tr>
<td>Air-dried in situ</td>
<td>0.85</td>
<td>31</td>
<td>2.55</td>
<td>4.2</td>
<td>4</td>
</tr>
<tr>
<td>Field-moist in situ</td>
<td>0.84</td>
<td>32</td>
<td>2.47</td>
<td>4.4</td>
<td>8</td>
</tr>
<tr>
<td>Field-moist smeared</td>
<td>0.75</td>
<td>41</td>
<td>1.93</td>
<td>11</td>
<td>4</td>
</tr>
</tbody>
</table>

† RMSD, root mean squared deviation; RPD, ratio of standard deviation of clay content to RMSD.

Table 5. Prediction accuracies of clay content using whole-field holdouts of field-moist in-situ scans. Models were created using partial least squares regression with five fields and then validated using soil cores from the sixth field. Six separate models were calibrated and validated.

<table>
<thead>
<tr>
<th>Validation field</th>
<th>RMSD</th>
<th>RPD</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>g kg⁻¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>143</td>
<td>0.72</td>
<td>-128.1</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>1.92</td>
<td>4.6</td>
</tr>
<tr>
<td>3</td>
<td>97</td>
<td>0.98</td>
<td>9.2</td>
</tr>
<tr>
<td>4</td>
<td>89</td>
<td>1.18</td>
<td>44.0</td>
</tr>
<tr>
<td>5</td>
<td>72</td>
<td>1.21</td>
<td>31.7</td>
</tr>
<tr>
<td>6</td>
<td>68</td>
<td>1.09</td>
<td>30.0</td>
</tr>
</tbody>
</table>
† RMSD, root mean squared deviation; RPD, ratio of standard deviation of clay content to RMSD.
Fig. 2

(a) \( y = 0.72x + 60 \), \( r^2 = 0.84 \)  
RMSD = 62 g kg\(^{-1}\)

(b) \( y = 0.88x + 31 \), \( r^2 = 0.92 \)  
RMSD = 41 g kg\(^{-1}\)

(c) \( y = 0.76x + 69 \), \( r^2 = 0.83 \)  
RMSD = 61 g kg\(^{-1}\)

(d) \( y = 0.65x + 103 \), \( r^2 = 0.75 \)  
RMSD = 74 g kg\(^{-1}\)
Fig. 3

Significant regression coefficient (b)

Similar wavelengths
Air-dried ground
Field-moist in situ
Air-dried in situ

Wavelength (nm)

500 750 1000 1250 1500 1750 2000 2250 2500
Fig. 4

Similar wavelengths

Smeared in situ

Field-moist in situ

Wavelength (nm)

Significant regression coefficient (b)

Fig. 5

Measured

VNIR prediction

clay content (g kg\(^{-1}\))

depth (cm)