Checks and Mass Balances for In Situ Quantification of Mineral Composition using Proximal Soil Sensors

Soil mineral composition affects soil behavior, but field estimation of mineral composition has proved difficult. To investigate the potential of predicting soil mineral composition in situ, 15 soils representing diverse mineral composition from New South Wales, Australia, were scanned with visible near-infrared (VisNIR) and portable X-ray fluorescence (pXRF) spectrometers to a depth of 1 m at 2.5-cm scanning increments. The presence of phyllosilicate and Fe oxides was assessed using a pattern-matching algorithm with VisNIR spectra from mineral end-member libraries. Rule-based iterative partitioning was then applied on pXRF elemental compositions based on mineral stoichiometry to determine the abundance of kaolinite, smectite, illite, hematite, goethite, calcite, gypsum, and quartz. This fusion model correctly identified the most abundant mineral in a sample with 72% accuracy. Mineral predictions were stable under variable moisture and surface conditions, as experienced when scanning samples in situ and under air-dry and ground conditions in the laboratory. Relative changes in mineral composition within a profile and across horizon boundaries were accurately expressed. The fusion model accurately quantified the abundance of quartz (Cohen's kappa coefficient, 0.67), and CaCO$_3$ (Cohen's kappa coefficient, 0.76; Lin's concordance correlation coefficient, 0.96; RMSE, 20.9 g kg$^{-1}$). The dominant phyllosilicate mineral was identified correctly with 86% accuracy, although accurate quantification of phyllosilicates and Fe oxides was not achieved. Conjoint use of VisNIR and pXRF spectroscopy as part of a fusion model approach showed great potential to provide comprehensive estimation of soil mineral composition in situ.

Abbreviations: LCCC, Lin's concordance correlation coefficient; pXRF, portable X-ray fluorescence; RPD, ratio of performance to deviation; VisNIR, visible near-infrared; XRD, X-ray diffraction.

Core Ideas
- VisNIR and pXRF were used to quantify soil mineral composition in situ.
- The fusion model approach leveraged the complementary strengths of VisNIR and pXRF.
- The most abundant mineral of a sample was predicted with 78% accuracy.
- Quantitative predictions were achieved for quartz, CaCO$_3$, and gypsum.
- This is the most comprehensive investigation of its kind to date.
and are limited to angles greater than 20° 29 (Gianoncelli et al., 2008). This is problematic because most phyllosilicates have primary peaks at smaller angles (<10° 29); thus, these devices are less equipped to estimate these important soil constituents. Subsequently, portable XRD devices do not offer a solution for rapid soil assessment and have been mostly limited to archeological and lithological investigations (e.g., Cannon et al., 2015; Nakai and Abe, 2012; Uda, 2004).

There is a growing interest in using proximal soil sensors to obtain information on soil systems. Two devices—visible near-infrared diffuse reflectance (VisNIR) and portable X-ray fluorescence (pXRF) spectrometer—have shown potential as field diagnostic devices because they can provide a wealth of information in a timely manner (e.g., Viscarra Rossel et al., 2011; Weindorf et al., 2014).

Many soil minerals have characteristic absorption features in the VisNIR range (Clark et al., 1990). Studies have successfully predicted a range of minerals in air-dry and ground samples. Brown et al. (2006) used boosted regression trees to predict the presence of kaolinite and montmorillonite in the clay fraction on a 0 to 5 ordinal scale, with 96 and 88% of samples falling within one ordinal unit, respectively. Mulder et al. (2013) parameterized absorption features in the 2.1- to 2.4-µm band of the VisNIR spectrum and predicted the abundance of kaolinite, dioctahedral mica, smectite, and calcite using regression tree analysis. The model worked well in the presence of ancillary minerals not accounted for in the training phase, with kaolinite, dioctahedral mica, and calcite having RMSE values of <8%. Meanwhile, Malone et al. (2014) used a pattern-matching algorithm based on diagnostic absorbance of mineral end members first demonstrated by Clark et al. (2003). This approach predicted the presence of kaolinite, smectite, illite, hematite, and goethite, which were used to map terrons in the viticultural study area.

Studies predicting soil mineral content in situ are limited. Viscarra Rossel et al. (2009) investigated absorbance values at diagnostic wavelengths from continuum removed spectra, collected in situ, to gain estimates of kaolinite, smectite, illite, hematite, goethite, gibbsite, calcite, and attapulgite. The authors describe qualitative agreement with observations from XRD diffractograms but noted that quartz could not be estimated from this approach. One limitation of VisNIR for comprehensive mineral characterization is a lack of absorbance features for quartz and for low-Fe feldspars in the 350- to 2500-nm wavelength range (Clark et al., 1990). To date, no attempt has been made to quantify quartz and feldspars from VisNIR spectra, and subsequently no attempt has been made to offer a full description of sample mineralogy. To provide a full description of sample mineralogy, other solutions must be explored.

Applications of pXRF to soil science are increasing (Weindorf et al., 2014). Zhu et al. (2011) demonstrated that pXRF data could be used to predict soil textural attributes, which is of interest because texture is related to mineral composition (McKenzie et al., 2004). Weindorf et al. (2009) found that, after correcting for Ca associated with calcite, pXRF data could predict soil gypsum content within 6% of laboratory values. Improvements in pXRF technology, including lower detection limits for lighter elements, have since allowed the direct prediction of gypsum content using pXRF reported S and Ca contents (Weindorf et al., 2013). Chakraborty et al. (2017) also found that pXRF reported Ca content of intact aggregates and that ground samples were useful in establishing the developmental stage of secondary carbonates in soils.

There are clear benefits from the combined applications of VisNIR and pXRF. Visible near-infrared diffuse reflectance spectroscopy gives information on sample color as well as molecular overtones and combination vibrations (Burns and Ciurczak, 2007). Portable X-ray fluorescence spectroscopy can give an estimate of the concentration of elements with atomic number ≥12 and is relatively stable under varying moisture conditions (Stockmann et al., 2016). Using the data from both devices, information on the elemental composition of a sample is provided via VisNIR, and some molecular information is provided via VisNIR. Weindorf et al. (2016) showed that combining information from the two devices produced the best prediction of calcium carbonate equivalent in a diverse set of arid soils from Spain (ratio of performance to deviation [RPD], 1.74). Conjoint use of VisNIR and pXRF information has also been found to produce more accurate predictions of a large range of agronomically relevant soil properties through model averaging of predictions given from individual devices (O’Rourke et al., 2016) or directly constructing regression models using information from both devices as model inputs (Duda et al., 2017).

Although previous studies have shown the potential of VisNIR and pXRF for the estimation of soil mineral composition, many of these studies estimated only a small number of soil minerals or provided only qualitative interpretation of results. Additionally, the vast majority of studies were conducted on air-dry and ground samples under laboratory conditions. To achieve the greatest potential of VisNIR and pXRF, efforts must be made to obtain results in situ, or at least on site, to allow investigators to do something useful with the results provided. A focus should also be given to providing a comprehensive soil mineral composition rather than estimating a few select minerals. The objective of this study was to determine if proximal soil sensors could be used to provide in situ quantification of mineral composition. It was hypothesized that conjoint use of VisNIR and pXRF spectroscopy as part of a fusion model approach would provide comprehensive estimation of soil mineral composition in situ.

**MATERIALS AND METHODS**

**Soil Sampling Locations**

Fifteen sites were chosen that exhibited a diverse range of mineralogy from across the state of New South Wales, Australia (Fig. 1). The 15 selected sites encompassed Alfisol, Aridisol, Inceptisol, Spodosol, Ultisol, Vertisol, and Oxisol soil orders according to US Soil Taxonomy (Soil Survey Staff, 2014). The sites were characterized with udic, ustic, and aridic soil moisture regimes and thermic soil temperature regimes.
In Situ Scanning Procedures

At each site a soil pit was excavated, and a 1 m × 1 m surface of the pit wall was prepared to a smooth surface by shearing excess soil with a shovel. A final shearing was conducted with an asparagus knife parallel to the soil surface and progressing from the top of the soil pit to the bottom to minimize contamination from falling debris. The soils were scanned in situ using VisNIR and pXRF at 2.5-cm increments in three 1-m vertical transects spaced 50 cm apart. This methodology resulted in 41 scan locations per 1 m vertical transect and 123 scan locations per soil pit. The VisNIR spectra were obtained with an AgriSpec device (Analytical Spectral Devices) connected via a fiberoptic bundle to a contact probe attachment with illumination provided by an inbuilt halogen lamp. Baseline readings were taken with a Spectralon tile (Labsphere Inc.) every 15 to 20 measurements, and pXRF readings were taken with an Olympus Delta Premium. Readings were taken in the proprietary Geochem mode. Samples were irradiated with X-rays of energy 50 and 10 keV successively for 30 s each. An inbuilt calibration reported the elemental composition for elements Al and heavier (i.e., Z > 12). A 316 stainless steel alloy clip was used to perform a calibration check daily immediately prior to operation. A SiO₂ blank and NIST standards 2709a, 2710a, and 2711a were scanned hourly during operation to track the performance of the device over time. The recovery of elements used in the fusion model (Al, Si, S, K, Ca, Ti, Fe) for all NIST standards ranged between 0.95 and 1.17. Results for one standard (NIST 2710a, Montana I soil) were (pXRF reported/NIST certified [recovery]): Al, 63.0/59.5 g kg⁻¹ [1.06]; Si, 310/311 g kg⁻¹ [1.00]; K, 23.1/21.7 g kg⁻¹ [1.06]; Ca, 9.8/9.6 g kg⁻¹ [1.02]; Fe, 41.0/43.2 g kg⁻¹ [0.95]. The recovery of S could not be assessed because a certified element content of S is not provided for these standards and because reported Ti was often below the detection limit of the pXRF spectrometer when operated in Geochem mode. After in situ scanning, horizons-based samples were taken for laboratory analysis and for scanning with VisNIR and pXRF again under air-dried, ground (<2 mm) conditions.

Data Fusion

A fusion model was used for predicting the mineral composition of the soil profiles. The VisNIR spectra were used to predict clay content and to identify phyllosilicates and Fe oxides. The results were moderated using an elemental mass balance from pXRF data (Fig. 2). This approach differs from that presented by Jones and McBratney (2016) in that clay content is predicted using VisNIR alone and not in conjunction pXRF data. Gypsum has also been added to the model. The fusion model approach and validation statistics were programmed in R (R Core Team, 2018).

Prediction of phyllosilicates (kaolinite, smectite, and illite) and Fe oxides (hematite and goethite) was performed using a shape-fitting algorithm across diagnostic wavelength ranges to quantify the deviation from convex hull corrected reference mineral spectra. This method is defined in more detail in Malone et al. (2014) and was previously used in the TetraCorder decision-making framework by the US Geological Survey (Clark et al., 2003). The VisNIR spectra were also used to estimate total clay concentration with spectral models constructed using the Cubist package in R (Kuhn and Quinlan, 2017). The spectral models were calibrated using 2594 samples from five spectral libraries across the state of New South Wales, Australia. Information on the five spectral libraries may be found in Filippi et al. (2018), Minasny et al. (2011), Malone et al. (2014), Fajardo et al. (2016), and Pino (2016). Estimated total clay content was then used to scale phyllosilicate speciation ratios to give quantitative predictions of each phyllosilicate species.
The average stoichiometry of minerals was used to calculate the proportional element mass contribution of Al, Si, S, K, Ca, Ti, and Fe as well as light elements, which represent constituent elements unquantifiable using pXRF (i.e., Z ≤ 12) (Table 1). Stoichiometric values were derived from the US Geological Survey digital spectral library (Clark et al., 2007) and other sources (Brigatti et al., 2006). The seven elements comprised a mean of 99.45% (SD, 0.65%) of the mass of total identifiable elements (Z > 12) for all in situ scans. The required elements for each phyllosilicate species were then tested against the pXRF elemental concentrations by iteratively subtracting 1% of the predicted elemental requirement from the available elements. In the event that individual quotas were filled or if a constituent element was exhausted, the partitioning of further elements to that mineral species was halted, and the remaining species continued to draw elements to fill their quota. For example, illite required contributions from Si, Al, and K. If K ran out before the illite quota is filled, then the partitioning of elements to illite was halted, but kaolinite continued to fill its quota if Si and Al were still available.

Once all quotas were filled or constituent elements exhausted, the moderated values were recorded, and the remaining elements were used for Fe oxides in a similar fashion. Residual elements were then used to predict gypsum (CaSO₄·H₂O), calcite, and quartz using known stoichiometry of these minerals. Gypsum content was dictated by the least limiting availability of S and Ca. The assumption was made that all remaining Ca was in the form of CaCO₃. Because the samples contained negligible quantities of other silicate minerals, such as olivine, pyroxene, or amphibole, the assumption was made that all remaining Si was in the form of tectosilicates (i.e., quartz and feldspars). This was further simplified in the fusion model to quartz only because few samples contained significant quantities of feldspar. Reported values for individual minerals were divided by the sum of observed mineral mass of a sample to give the proportional mass contribution.

Horizon-Based X-Ray Diffraction

Horizon-based samples were ground to a fine powder, and randomly oriented samples were analyzed using monochromatic CuKα radiation at 30 kV and 28.5 mA with a mini-materials analyzer X-ray diffractometer (GBC Scientific Equipment Pty Ltd). The samples were scanned from 4° to 65° 2θ at a speed of 1° 20 min⁻¹ and using a step size of 0.01° 2θ. The clay fraction of samples was isolated using a sedimentation process based on Stoke's law. Oriented samples from the clay fraction were analyzed following four pretreatments for the identification of phyllosilicate species (i.e., Mg-saturated, Mg-saturated and ethylene glycol solvated, K-saturated, and K-saturated) and heated to 550°C (Brown and Brindley, 1980). A randomly oriented sample of the clay fraction was scanned to aid in the identification of minerals other than phyllosilicates. The presence of kaolinite, smectite, illite, hematite, goethite, calcite, gypsum, and quartz was quantified based on expert interpretation of powdered and oriented XRD diffractograms into six ordinal classes representing 0, 0 to 5, 5 to 20, 20 to 40, 40 to 60, 60 to 80, and 80 to 100% of the total mineral composition. The four most abundant minerals in each sample were identified and used for comparison with fusion model reported values.

Correlation between Minerals and Portable X-Ray Fluorescence Values

Spearman’s rank order correlation coefficient (ρ) was used to explore relationships between XRD-reported mineral abundance and pXRF-reported elemental composition under air-dry, ground condition. Spearman’s rank correlation coefficient was chosen over Pearson’s correlation coefficient because the XRD ranking scale was not based on a uniform interval and observations were not normally distributed. The Spearman’s rank order correlation coefficient describes how well the relationship between two variables can be described using a monotonic function, which is not necessarily linear, as is stipulated with Pearson’s correlation coefficient. Spearman’s rank correlation coefficient is defined as the Pearson correlation coefficient between ranked variables and was calculated using the following formula:

$$\rho = \frac{\sum_{i=1}^{n} (r_x_i - \bar{r}_x)(r_y_i - \bar{r}_y)}{\sqrt{\sum_{i=1}^{n} (r_x_i - \bar{r}_x)^2 \sum_{i=1}^{n} (r_y_i - \bar{r}_y)^2}}$$

where $r_x_i$ and $r_y_i$ are the i-th ranked variables of $r_x$ and $r_y$, respectively, and $\bar{r}_x$ and $\bar{r}_y$ are the mean values of the ranked variables $r_x$ and $r_y$, respectively.

### Table 1. Average mass-based mineral stoichiometry used in the elemental mass balance.

<table>
<thead>
<tr>
<th>Mineral</th>
<th>LE†</th>
<th>Al</th>
<th>Si</th>
<th>S</th>
<th>K</th>
<th>Ca</th>
<th>Ti</th>
<th>Fe</th>
<th>Element contribution to total mineral mass (g kg⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaolinite</td>
<td>576</td>
<td>201</td>
<td>212</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Smectite</td>
<td>588</td>
<td>98</td>
<td>279</td>
<td>0</td>
<td>3</td>
<td>11</td>
<td>1</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Illite</td>
<td>545</td>
<td>118</td>
<td>243</td>
<td>0</td>
<td>63</td>
<td>3</td>
<td>4</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Hematite</td>
<td>301</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>699</td>
<td></td>
</tr>
<tr>
<td>Goethite</td>
<td>371</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>629</td>
<td></td>
</tr>
<tr>
<td>CaCO₃</td>
<td>600</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>400</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Gypsum</td>
<td>581</td>
<td>0</td>
<td>186</td>
<td>0</td>
<td>233</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Quartz</td>
<td>533</td>
<td>0</td>
<td>467</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

† Light elements (LE) are those with atomic number ≤12, which are unidentifiable by portable X-ray fluorescence.
Evaluation of Mineral Estimations

In situ mineral estimations were aggregated by horizon, and the mean value was compared with XRD estimations of ordinal abundance by calculating the linear-weighted Cohen’s kappa coefficient, $\kappa_{w}$ (Cohen, 1960). This statistic measures the inter-rater agreement between the fusion model and XRD analysis using the following formula:

$$\kappa_{w} = 1 - \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{i,j} (x_{i,j} - m_{i,j})^2}{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{i,j} - \sum_{i=1}^{k} \sum_{j=1}^{k} w_{i,j}}$$

where $k$ is the number of ordinal classes, $w$ is the weight matrix, $x$ is the observed matrix obtained from the fusion model, and $m$ is the expected matrix based on chance agreement. The weight matrix consisted of zeros on the diagonal, cells one off the diagonal are weighted 1, cells two off the diagonal are weighted 2, and so on.

Evaluation of Carbonate Prediction

The $\text{CaCO}_3$ estimations were further validated using carbonate equivalent values obtained by the rapid titration method (Rayment and Lyons, 2011). Because these results are continuous in nature, validation metrics such as the RMSE, Lin’s concordance correlation coefficient (LCCC), and the RPD were calculated. Prior to validation, in situ estimates of $\text{CaCO}_3$ contribution to the mineral fraction from the fusion model was converted to the $\text{CaCO}_3$ contribution to the whole soil mass, $[\text{CaCO}_3]_{\text{soil}}$, by accounting for the mass contribution of C from organic matter. To achieve this, fusion model estimations of $\text{CaCO}_3$ were aggregated by horizon, and the horizon mean was then calculated and adjusted using the following formula:

$$[\text{CaCO}_3]_{\text{soil}} = \frac{[\text{CaCO}_3]_{\text{mineral}} \times (1000 - \text{OC} \times 1.72) / 1000}{1 - 1.72 \times \text{OC}}$$

where $[\text{CaCO}_3]_{\text{mineral}}$ is the $\text{CaCO}_3$ contribution to the mass of the mineral fraction, OC is the organic carbon content expressed in g kg$^{-1}$, and 1.72 is a factor to convert organic carbon to organic matter.

RESULTS AND DISCUSSION

Mineral Composition from X-ray Diffraction Analysis

The 15 soils showed diverse mineral composition between sites and frequently between horizons within a profile. Five separate minerals were identified as the most dominant in at least one of the profiles, demonstrating the diversity of soils sampled (Table 2). Quartz and kaolinite were the most ubiquitous minerals; both were identified in all samples, although at times only in trace amounts. Quartz was the most dominant mineral in half of the samples, reflecting the parent material and the highly weathered nature of many of the profiles. Relationships between mineralogy and soil type were observed. Smectite was the dominant mineral in 17 samples, corresponding mostly with the three Vertisol profiles. Iron oxides were dominant in the Oxisol. Calcite was the dominant mineral in two of the subsoil horizons of one profile. Illite and gypsum were not the most dominant mineral in any soil. Illite was found to be the second most dominant mineral in three horizons. Gypsum was only identified in a single horizon, and 13 horizons had only three identifiable minerals. A number of minerals not directly accounted for in the fusion model were also observed in the XRD analysis. These were vermiculite, hydroxy-interlayered vermiculite, rutile, anatase, and gibbsite. K-feldspar, anorthite, and albite were also identified in a number of horizons. All of these minerals were present in low amounts, and future efforts should attempt to quantify all auxiliary minerals and individual feldspar species for comprehensive mineral characterization.

Correlation between Mineral Composition and pXRF Reported Elements

Spearman’s rank correlation coefficients revealed that mineral abundance was significantly correlated with pXRF-reported elemental composition. Each mineral was positively correlated with at least one element, and significant negative correlations were observed (Table 3).

Kaolinite was found to be positively correlated with Al (0.56). This may be due to two factors: (i) kaolinite is a 1:1 layer silicate, meaning it has a greater proportion of octahedral alumina sheets compared with 2:1 layer silicates, such as illite and smectite, and (ii) kaolinite is the most abundant phyllosilicate in highly weathered soils, such as Oxisols, which are also more likely to contain Al in the form of gibbsite or associated with Fe oxides (Isbell, 1994). This is also supported by a positive correlation of kaolinite with Fe (0.33).

Ililite had a significant positive correlation with K (0.52). In contrast to other phyllosilicates, illite has a large amount of K ions in the interlayer space to offset the negative charge resulting from the isomorphic substitution in the mineral. The strong correlation here may also be attributable to the lack of significant quantities of other K-enriched minerals, such as K-feldspars, which were only observed in trace amounts in two horizons.

Smectite had a significant positive correlation with Ca (0.78). Although Ca is not a structural element of smectite, soils that are dominated by smectite have a large cation exchange capacity, and Ca is the predominant cation on the exchange sites. An even stronger correlation was observed between calcite and Ca (0.82).

Hematite had a significant correlation with Fe (0.81), Al (0.62), and Ti (0.62). This reflects the strong association with Al and Ti oxides in weathered soils (Isbell, 1994). Goethite showed

Table 2. Count of dominant mineral allocation observed by X-ray diffraction (XRD) analysis of 65 soil horizon samples sourced from 15 diverse soil profiles within the state of New South Wales, Australia.

<table>
<thead>
<tr>
<th>Mineral</th>
<th>XRD dominant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>Kaolinite</td>
<td>9</td>
</tr>
<tr>
<td>Smectite</td>
<td>17</td>
</tr>
<tr>
<td>Illite</td>
<td>0</td>
</tr>
<tr>
<td>Fe oxides</td>
<td>5</td>
</tr>
<tr>
<td>Calcite</td>
<td>2</td>
</tr>
<tr>
<td>Gypsum</td>
<td>0</td>
</tr>
<tr>
<td>Quartz</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>65</td>
</tr>
</tbody>
</table>

www.soils.org/publications/sssaj
similar but less significant relationships. Goethite was often found in small quantities in sandstone-derived alluvial soils, which may have influenced the correlation values and explain why it does not have a significant negative correlation with Si, as observed with hematite.

Gypsum showed a significant relationship with S (0.46), although only one sample contained identifiable levels of gypsum. Quartz showed a highly significant positive relationship with Si (0.77) and a negative or no correlation to all other elements. These significant relationships between pXRF-reported elemental composition and ordinal XRD mineral abundance measures underlie the potential of the conjoint use of VisNIR and pXRF for estimation of mineral abundance, compared with using VisNIR alone.

High-Resolution Mineral Composition Estimations

Fusion model predictions gave high-resolution estimates of mineral composition within a profile. Mineral composition estimates often matched with XRD-derived values. Comparing estimated values between two distinct profiles demonstrates the success of the fusion model used in this study (Fig. 3). Dominant mineral abundance between profiles and relative changes within profiles were well represented with changes in mineral composition often occurring across horizon boundaries.

X-ray diffraction analysis of an Ultisol profile indicated that the A and E horizons of this profile were dominated by quartz, whereas kaolinite was the dominant phyllosilicate in the clay fraction and Fe oxides were present in trace amounts. In the Bt1 and Bt2 horizons, kaolinite and Fe oxides became more dominant, whereas traces of illite appeared and the overall abundance of quartz decreased. These observations closely matched the fusion model predictions (Fig. 3a). The change in phyllosilicate minerals in transition between the E and Bt1 horizon appears gradual; however, this is because the horizon was wavy, and the average of the three vertical transects is shown.

The A horizon of an Aridisol was also observed to be quartz dominated by XRD analysis. Kaolinite, smectite, and illite were all observed in the clay fraction, and no Fe oxides were identified. A texture contrast boundary was observed between the A and Bt horizons. The Bt horizon experienced a decrease in quartz and an increase in all phyllosilicates, as expected. Phyllosilicate content remained constant throughout the B horizons. Calcite was observed in Btk1 and Btk2 horizons only. Again, these observations closely matched the fusion model predictions (Fig. 3b).

Dominant Minerals

The fusion model was effective in identifying the dominant mineral in a sample. After aggregating in situ scans by horizon, it was observed that the dominant mineral was correctly assigned with 72% accuracy, whereas the dominant mineral was incorrectly assigned in 28% of cases to the second most dominant mineral (Table 4).

Four of the incorrect assignments occurred when quartz was assigned as the dominant mineral in place of kaolinite; however, in each case XRD analysis identified quartz as constituting 20 to 40% of the mineral phase. One incorrect assignment occurred in a sesquioxides accumulated horizon, where kaolinite was assigned in place of quartz. This may be attributed to the increased Fe and Al in this horizon, incorrectly assigned to kaolinite. Five incorrect

### Table 3. Spearman's rank-order correlation coefficient between X-ray diffraction observed ordinal mineral abundance and portable X-ray fluorescence (pXRF) reported element composition of 65 soil horizon samples sourced from 15 diverse soil profiles within the state of New South Wales, Australia.

<table>
<thead>
<tr>
<th>Mineral</th>
<th>LE†</th>
<th>Al</th>
<th>Si</th>
<th>S</th>
<th>K</th>
<th>Ca</th>
<th>Ti</th>
<th>Fe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaolinite</td>
<td>−0.16</td>
<td>0.56***</td>
<td>−0.12</td>
<td>0.05</td>
<td>0.01</td>
<td>−0.50***</td>
<td>−0.09</td>
<td>0.33**</td>
</tr>
<tr>
<td>Smectite</td>
<td>0.45***</td>
<td>−0.05</td>
<td>−0.29*</td>
<td>−0.10</td>
<td>0.38**</td>
<td>0.78***</td>
<td>−0.13</td>
<td>0.14</td>
</tr>
<tr>
<td>Illite</td>
<td>0.05</td>
<td>0.03</td>
<td>0.29*</td>
<td>−0.36***</td>
<td>0.52***</td>
<td>0.09</td>
<td>−0.45***</td>
<td>−0.33**</td>
</tr>
<tr>
<td>Hematite</td>
<td>0.00</td>
<td>0.62***</td>
<td>−0.57***</td>
<td>0.36**</td>
<td>−0.21</td>
<td>−0.12</td>
<td>0.62***</td>
<td>0.81***</td>
</tr>
<tr>
<td>Goethite</td>
<td>−0.36**</td>
<td>0.28*</td>
<td>−0.08</td>
<td>0.35**</td>
<td>−0.47***</td>
<td>−0.60***</td>
<td>0.32**</td>
<td>0.25*</td>
</tr>
<tr>
<td>Calcite</td>
<td>0.38**</td>
<td>−0.07</td>
<td>−0.40***</td>
<td>−0.06</td>
<td>0.23</td>
<td>0.82***</td>
<td>−0.07</td>
<td>0.15</td>
</tr>
<tr>
<td>Gypsum</td>
<td>−0.17</td>
<td>−0.11</td>
<td>0.03</td>
<td>0.46***</td>
<td>−0.14</td>
<td>0.19</td>
<td>0.00</td>
<td>−0.04</td>
</tr>
<tr>
<td>Quartz</td>
<td>−0.23</td>
<td>−0.54***</td>
<td>0.77***</td>
<td>−0.43***</td>
<td>0.01</td>
<td>−0.44***</td>
<td>−0.19</td>
<td>−0.71***</td>
</tr>
</tbody>
</table>

* Signifcant at the 0.05 probability level.
** Signifcant at the 0.01 probability level.
*** Signifcant at the 0.001 probability level
† Light elements (LE) are those with atomic number ≤12, which are unidentifiable by pXRF.
assignments occurred in the Oxisol, where kaolinite, and not Fe oxides, was identified as the dominant mineral. The prediction of Fe oxides may be improved by accounting for the Al substitution. Singh and Gilkes (1992) investigated 39 soils in southwestern Australia and identified that goethite contained 13 to 35 mol% Al substitution and hematite contained 4 to 23 mol% Al substitution. This is reflected with the significant correlation of the Fe oxides with Al and with Ti found in this study (Table 3). Eight incorrect assignments of dominant mineral occurred in Vertisols, where smectite was correctly estimated as the dominant phyllosilicate in each instance. This suggests the model could be improved for Vertisols and Oxisols, which are dominated by smectite and Fe oxides, respectively. The second, third, and fourth most dominant mineral was estimated with 51, 49, and 48% accuracy, respectively (Table 4). For comparison, the fusion model was also run with VisNIR spectra and pXRF data obtained from air-dry, ground horizon samples, and laboratory-based clay content. Under this laboratory scenario, the correct allocation was improved to 77, 60, 60, and 48% accuracy, respectively (data not shown).

Two of the profiles that gave poor predictions of the dominant mineral were in smectite-rich Vertisols. The underprediction of smectite in these soils may be due to differences in the elements involved in isomorphic substitution and charge balancing (Moore and Reynolds, 1997), which may also affect absorbance features in the VisNIR spectra (Clark et al., 1990). Analysis of the seven minerals in the USGS library used to construct the smectite mass balance revealed the following ranges in constituent oxides: 52.0 to 69.6% SiO$_2$, 15.0 to 20.9% Al$_2$O$_3$, and 1.17 to 5.18% Fe$_2$O$_3$. In contrast, the six minerals used to construct the kaolinite mass balances had much smaller ranges: 42.0 to 47.1% SiO$_2$, 36.4 to 38.0% Al$_2$O$_3$, and 0.26 to 1.61% Fe$_2$O$_3$. Smectite was also overpredicted in agriculture topsoil samples. This may be due to enrichment of Ca from plant material, organic matter, fertilizers, or lime application. Both of these areas are highlighted for further improvement for the model.

### Phyllosilicate Prediction by VisNIR and Fusion Model

The VisNIR predictions of phyllosilicate speciation were stable under variable moisture content and surface condition, as observed by scanning in situ and under air-dry, ground condition in the laboratory. The LCCC values were >0.90 for each phyllosilicate species (Fig. 4). This demonstrates that the metrics chosen to estimate phyllosilicate speciation were robust under variable moisture conditions.

The chosen metrics were also effective in identifying the dominant phyllosilicate. The dominant phyllosilicate was identified correctly using only VisNIR in 85 and 88% of cases in situ and in air-dry, ground conditions, respectively (Table 5). However, VisNIR alone was less effective at predicting the ordinal dominance of all three phyllosilicate species in a sample. This was only achieved in 34% of cases in both in situ and under air-dry and ground conditions in the laboratory. The fusion model outperformed VisNIR alone when estimating the dominance of all three phyllosilicate species. The correct order of phyllosilicates was identified in 63% of cases under both in situ and air-dry, ground condition for the fusion models. The estimation of the dominant phyllosilicates in a sample was very effective, but the absolute abundance of individual phyllosilicates was poorly estimated.

### CaCO$_3$ and Gypsum Predictions from the Fusion Model

The fusion model was effective in predicting CaCO$_3$ content (LCCC, 0.96; RMSE, 20.9 g kg$^{-1}$; RPD, 4.47) when compared with carbonate equivalent values obtained by the rapid ti-

<table>
<thead>
<tr>
<th>Model dominant</th>
<th>XRD dominant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>I</td>
<td>47</td>
</tr>
<tr>
<td>II</td>
<td>18</td>
</tr>
<tr>
<td>III</td>
<td>0</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
</tr>
<tr>
<td>&gt;IV</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>65</td>
</tr>
</tbody>
</table>

Fig. 4. Comparison of the proportion of phyllosilicate speciation predicted using visible near-infrared (VisNIR) spectra obtained in situ and aggregated to horizon versus horizon-based samples scanned after air-drying and grinding (<2 mm). Samples were obtained from 65 soil horizons sourced from 15 diverse soil profiles within the state of New South Wales, Australia. Il, illite; Ka, kaolinite; LCCC, Lin's concordance correlation coefficient; Sm, smectite.
Calcium carbonate equivalent (g kg⁻¹) from horizon-based carbonate equivalent data acquired by the rapid titration method and horizon aggregated in situ CaCO₃ predictions from the fusion model (Lin's concordance correlation coefficient, 0.96; rMSE, 20.9 g kg⁻¹). Samples were obtained from 65 soil horizons sourced from 15 diverse soil profiles within the state of New South Wales, Australia.

Quartz Prediction from the Fusion Model

Quartz was also estimated with high accuracy, given its $k_w$ of 0.67 (Table 6). Because quartz has no spectral response in the VisNIR region, the accuracy of this prediction is attributable to the fusion model approach and the strong relationship between quartz and pXRF recorded Si ($p = 0.77$) (Table 3). This highlights the strength of the fusion approach because quartz has no direct absorption bands in VisNIR spectra.

Model Improvements

A number of improvements could be made to the model. The inclusion of gibbsite, Ti oxides, feldspars, and vermiculite must be addressed, as well as any other relevant minerals. Gibbsite and vermiculite have absorbance features in the VisNIR region, which could be incorporated into the fusion model. The lack of absorption features of low-Fe feldspars in the VisNIR range will continue to limit their prediction. Distinguishing between hematite and goethite also presented difficulty, as identified previously (e.g., Malone et al., 2014; Vendrame et al., 2012). Raman spectroscopy may offer a solution because it has been demonstrated to provide, under laboratory conditions, a number of absorption peaks related to feldspars and Fe oxides as well as quartz and common igneous minerals such as olivine and pyroxene (Freeman et al., 2008). Mid-infrared spectroscopy would also be useful to incorporate because it contains fundamental absorbances of phyllosilicate minerals and calcite (Farmer and Russell, 1964; Janik et al., 1995). The mid-infrared spectrum is more susceptible to variable moisture and scanning surface condition compared with VisNIR, which may limit field applications, although field-portable devices are available (Sorak et al., 2012).

The mass balance is limited in that the pXRF spectrometer used in this study was only able to detect elements heavier than Mg. Other pXRF sensors are available with the capacity to measure Mg in the field, which is an integral component in many of the predicted minerals, as well as vermiculite. An alternative to pXRF in the model would be laser-induced breakdown spectroscopy. Laser-induced breakdown spectroscopy provides quantitative estimates of light elements that cannot be quantified using

Table 5. Comparison of the accuracy of prediction of the dominant phyllosilicate and the ordinal abundance of all phyllosilicates by visible near-infrared (VisNIR) spectra only and using the fusion model approach. Samples were obtained from 65 soil horizons sourced from 15 diverse soil profiles within the state of New South Wales, Australia. Results are presented for samples scanned in situ and under air-dried and ground (<2 mm) conditions.

<table>
<thead>
<tr>
<th>Model estimated†</th>
<th>VisNIR only</th>
<th>Fusion model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In situ</td>
<td>Ground</td>
</tr>
<tr>
<td>Dominant phyllosilicate</td>
<td>85</td>
<td>88</td>
</tr>
<tr>
<td>All phyllosilicates</td>
<td>34</td>
<td>34</td>
</tr>
</tbody>
</table>

†, not present; tr, trace (<5%); x, 5–20%; xx, 20–40%; xxx, 40–60%; xxxx, 60–80%; xxxxx, 80–100%.

Table 6. Contingency table of semi-quantitative quartz abundance from X-ray diffraction (XRD) analysis and the fusion model approach. Samples were obtained from 65 soil horizons sourced from 15 diverse soil profiles within the state of New South Wales, Australia. Bold numerals on the diagonal represent a perfect match between fusion model estimates and XRD reported quartz abundance.

<table>
<thead>
<tr>
<th>XRD abundance</th>
<th>Model estimated†</th>
<th>VisNIR only</th>
<th>Fusion model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>–</td>
<td>tr</td>
<td>x</td>
</tr>
<tr>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>tr</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>x</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>xx</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>xxx</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>xxxx</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>xxxxx</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

†, not present; tr, trace (<5%); x, 5–20%; xx, 20–40%; xxx, 40–60%; xxxx, 60–80%; xxxxx, 80–100%.
pXRF, including Li (Jantzi and Almirall, 2011). However, laser-induced breakdown spectroscopy is also affected by moisture content, and the amount of sample analyzed is in the range of nanograms to picograms, which suggests that small heterogeneities in a sample will greatly influence results.

This investigation presented the most comprehensive attempt at estimating the mineral composition of soil in situ. The limitation of a comprehensive prediction of mineral abundance is that if one prediction is incorrect, then the accuracy of all other predictions is reduced as well. It is difficult to assess the performance of the fusion model in context with previous studies. Many existing studies give only qualitative description of the strength of the mineral signal and relative changes in a profile without formal validation (e.g., Malone et al., 2014; Viscarra Rossel et al., 2009). In other instances, validation is only recorded for the clay fraction and not the entire soil (e.g., Brown et al., 2006). Future studies should aim at providing comprehensive mineral estimations and appropriate validation statistics.

CONCLUSIONS

A new method for in situ quantification of soil mineral composition based on VisNIR and pXRF spectroscopy was investigated. This fusion model was centered on pattern-matching of VisNIR spectra to identify the relative abundance of phyllosilicates and Fe oxides, combined with an elemental mass balance based on pXRF-reported elemental composition. Kaolinite, smectite, illite, hematite, goethite, CaCO$_3$, gypsum, and quartz were estimated, making this model the most comprehensive attempt to estimate soil mineral composition using proximal soil sensors. The fusion model correctly identified the most abundant mineral in a sample with 72% accuracy, with the remainder assigned to the second most abundant mineral of the sample. The second, third, and fourth most abundant minerals were correctly assigned with 51, 49, and 48% accuracy, respectively. Prediction of CaCO$_3$, quartz, and gypsum was quantitative. The abundance of phyllosilicates was qualitative only because the model was able to predict the relative dominance of phyllosilicate species but not their absolute abundance. This may be due to variation in the chemical composition from various isomorphic substitutions, different cations balancing the negative charge of these minerals, and noncrystalline phases in the sample that were not identified from the XRD analysis. The model was stable under in situ as well as air-dry and ground conditions, demonstrating that it is a viable field technique.

Timely access to soil mineral composition acquired using the fusion model approach has the potential to enhance soil morphological description by facilitating the direct investigation of lithologic discontinuities and other diagnostic pedogenetic features. The fusion model results may also provide valuable information for the identification of terrons in the Lower Hunter Valley, Australia. Geoderma 132:621–636. doi:10.1016/j.geoderma.2005.04.025


Clark, R.N., G.A. Swayze, R. Wise, K.E. Livo, T.M. Hoefer, et al. 2007. USGS digital spectral library sphib06s. USGS Data Ser. 231, USGS, Reston, VA.


Soil Science Society of America Journal


