Comparison of two NIR systems for quantifying kaolinite in Weipa bauxites

Luke McArthur[1] and Colin Greensill[2]

[1] Process Engineering and Light Metals Centre, Central Queensland University, Australia
[2] Faculty of Science, Engineering and Health, Rockhampton, 4701, Australia

E-mail: luke.mcarthur@riotinto.com

Abstract
Two laboratory-based near infrared systems were compared for the purpose of quantifying kaolinite within Weipa bauxites. The two systems were based on different instrument technologies, a Bruker MPA (multi-purpose analyser) FTNIR (Fourier transform near infrared) instrument, and a Foss Model 6500 pre-dispersive grating instrument. Systems were compared in terms of accuracy, robustness, sample throughput and ease of use. The MPA-based system was superior in terms of accuracy (root mean squared error of prediction (RMSEP) = 0.88%m/m), throughput and ease of use. The Model 6500 had an advantage of offering the most robust measurement with greater time stability and lower errors in repeated and repacked measurements. However, addition of a small number of temporally displaced and repacked samples in the calibration set improved robustness of the MPA system, matching the Model 6500.

Keywords: kaolinite, bauxite, near infrared, mineral, signal to noise, 6500, MPA

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Bauxite quality is assessed on its suitability for alumina refining, namely, the potential alumina yield and running costs for refinement. In the Bayer process of alumina extraction, silica dissolved in the digestion phase must be removed as sodalite, consuming valuable caustic soda. The mineral kaolinite, also termed reactive silica, is the primary source of soluble silicates and hence a significant expense in the refinement process. Estimated soda losses can account for as much as 20% of alumina production costs in high silica bauxites [1]. Precise determination of kaolinite content is therefore important to bauxite mining and refining industries but accurate analyses are complicated by the complex bauxite mineralogy. Techniques such as atomic absorption spectroscopy and x-ray fluorescence cannot directly discriminate minerals of similar atomic composition such as gibbsite, boehmite and kaolinite. On the other hand, wet chemical methods and x-ray diffraction are prohibitively slow for large-scale analysis.

Near infrared spectroscopy (NIRS) offers a rapid and economical alternative for quantitative mineralogical determinations. NIRS utilizes the 12 500–4000 cm^{-1} region of the electromagnetic spectrum, residence to the numerous bands arising from vibrational overtones and combinational transitions of OH, CH, NH functional groups [2]. The strong dependence of the band position to the molecular structure and atomic composition suggests that bauxite minerals containing hydroxyl (OH) groups such as kaolinite, gibbsite, boehmite, hydrohematite and goethite should possess a characteristic presence in the near infrared (NIR) spectrum. The main advantage of NIRS is low absorptivity of NIR radiation. This characteristic permits analysis of solid samples with large path length without requiring a costly and sometimes hazardous preparation. NIRS is however a secondary method and due to the inherently poor resolution of NIR spectra, requires careful calibration for quantitative analysis.

Quantitative mineralogical analysis of bauxite using NIRS remains relatively unexplored, or at least undocumented in the public domain. A number of studies have been conducted

analysing the spectra of bauxite minerals including kaolinite [3–7], gibbsite and boehmite [6] demonstrating that the NIR spectra of these bauxite minerals were characteristic of the location of OH groups within the lattice structure. Walker appears to be the first and only researcher to analyse bauxite [8, 9] and published preliminary work developing quantitative calibration models using the PIMA II portable instrument [10]. Walker developed partial least-squares-based (PLS) calibration models for kaolinite, gibbsite, boehmite and hematite, with corresponding R^2 values of 0.73, 0.87, 0.89 and 0.85. It is expected that the calibration performance can be improved by utilizing systems based on modern laboratory grade instruments as well as application of new chemometric data treatment prior to PLS calibration.

In this study, two NIR systems (one Fourier transform, the other pre-dispersive) were trialled and assessed for suitability in determination of kaolinite content in Weipa bauxites. The spectral data were collected and PLS calibration models were optimized utilizing a number of proven data treatments.

Systems were compared on four criteria: (1) accuracy, (2) robustness, (3) sample throughput and (4) ease of use. Accuracy was assessed by the root mean squared error of prediction (RMSEP) when the model was applied to an independent test set (see section 2). The throughput was assessed on time taken for sample presentation and measurement. Ease of use was assessed on the number of steps and degree of complexity for preparing the sample and measuring its NIR spectrum.

In this context, robustness is the stability of measurement against natural experimental variations within same-operator, same-laboratory conditions. Robustness was evaluated by assessing the error introduced through four aspects of routine measurement. The first aspect was time stability of the system, which is subject to instrumental drift and seasonal effects [11]. The contribution from the latter is only expected to be minor given that the laboratory was air-conditioned. The second aspect was concerned with instrumental noise inherent in the repeated measurement of one sample. The third and fourth aspects were concerned with the operator and sample variability due to handling and repacking of the sample, respectively. These two aspects are most likely to impact the scattering profile and homogeneity of the sample, having obvious impact on the NIR spectrum.

2. Theory

2.1. Assessment of accuracy/error

Accuracy of calibration models and error from experimental variations were assessed by RMSEP on the appropriate validation set:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{N} \sum_{g=1}^{G} (\hat{y}_{i,g} - y_i)^2}{N \cdot G}}$$

where $\hat{y}_{i,g}$ is the prediction made on the gth of G repeated spectral measurements for the ith of N samples. The significant difference between the accuracy of two models was tested using the method reported by Fearn [12], using a 5% level of significance.

2.2. Spectral noise

Spectral noise was quantified using the root mean square (RMS) of residuals. The signal-to-noise ratio (SNR) provides a measure of the noise relative to the average signal strength. The SNR was determined as the averaged RMS noise to mean signal ratio at each wavenumber:

$$\mathbf{rms} = \sqrt{\frac{\sum_{i=1}^{N} \sum_{g=1}^{G} (\mathbf{x}_{i,g} - \bar{\mathbf{x}}_i)^2}{N \cdot G}}$$

$$\mathbf{snr} = \mathbf{rms}./\bar{\mathbf{x}}$$

$$\text{SNR} = \frac{1}{p} \sum_{j=1}^{p} \text{snr}_j$$

$\mathbf{x}_{i,g}$ is the spectrum of the gth repeated measurement of the ith sample
$\bar{\mathbf{x}}_i$ is the mean spectrum for this ith sample
$\mathbf{snr}$ is the signal-to-noise ratio spectrum
$\bar{\mathbf{x}}$ is the mean spectrum of the set of N samples
$./$ is elementwise division
snr_j is the signal-to-noise ratio for the jth of p wavenumbers/wavelengths.

3. Experimental details

3.1. Samples

Rio Tinto Aluminium Mining and Refining Research and Technical Support (RTA M&R R&TS) provided 263 bauxite samples with reference analysis from two different mine sites in the Weipa region. To develop calibrations for kaolinite, samples were divided into a calibration set consisting of 204 samples and a validation set consisting of 59 samples. The calibration and validation sets had similar and heavily skewed distributions of the mineral contents. A distribution of the sample sets is shown in figure 1.

Weipa bauxites have only two silica bearing minerals: kaolinite and quartz. Kaolinite content was therefore determined through the difference between total silica obtained from XRF and quartz content measured using a gravimetric digestion in hydrofluoric acid.

Prior to reference analyses and spectral measurement the bauxite samples were sieved, ground and dried. The samples were dry sieved to with an upper sieve size of 16 mm and a lower sieve size of 1.3 and 0.3 mm for the first and second mine site respectively. The sieved faction was ground using a Rocklabs ring grinder for 3 min producing a particle size distribution with 95% material less than 75 μm. The samples were then dried for 2 h at 105 °C to remove free moisture.

3.2. Instruments

The two new candidate systems were based on different laboratory grade instruments. The first system was based on the multi-purpose analyser (MPA) (Bruker Optik GmbH, Ettlingen, Germany) and the second system was based on the Model 6500 (Foss NIRSystems Inc., Silver Springs, USA).

The MPA is a Fourier transform (FT) NIR instrument with a single lead-sulfide (PbS) detector offering a full range
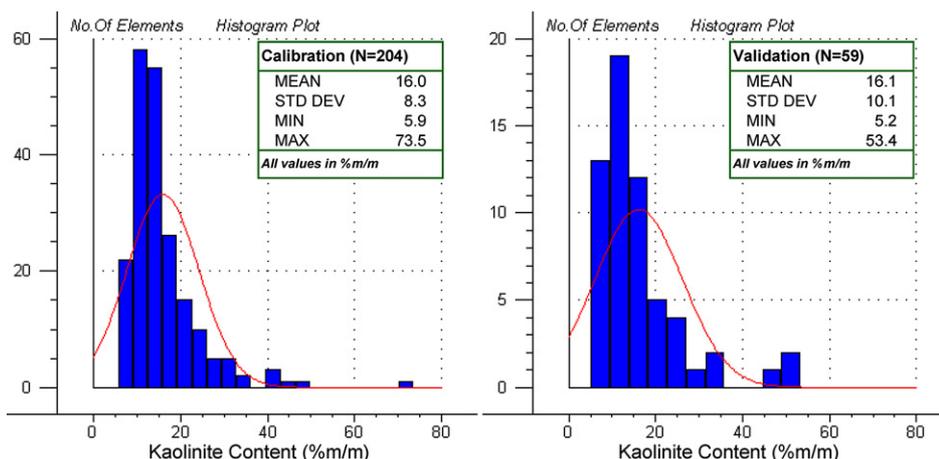
Figure 1. Distribution of kaolinite content in the calibration (left) and validation (right) sets.

measurement between 12 000 and 3600 cm^{-1} at variable resolution (2 cm^{-1} maximum).

The Model 6500 is a pre-dispersive grating NIR instrument with a dual silicon and PbS detector system offering full range measurement of visible and near infrared, 25 000–4000 cm^{-1} (400–2500 nm). The Model 6500 has a fixed resolution of 10 nm corresponding to a resolution of 37 cm^{-1} at 6060 cm^{-1} and makes spectral measurements in 2 nm steps (7 cm^{-1} at 6060 cm^{-1}).

3.3. Spectral acquisition

Sample presentation was optimized for each instrument. The MPA instrument was capable of 30 sample batch measurements using a circular sample carousel. Samples were presented in disposable glass vials (MGlas AG IN 440) approximately 20 mm in diameter. Approximately 5 g of powdered bauxite sample was dispensed into the vial to provide a sample thickness of 15–20 mm to ensure a condition of zero transmittance. Spectral measurement was performed over the instrument's full spectral range (12 000–3600 cm^{-1}) at a 16 cm^{-1} resolution. For each sample in the carousel, a reference was measured followed by ten spectra measured before progressing to the next position. Each spectrum (average of 32 forward and backward scans) took approximately 9 s to acquire.

The Model 6500 instrument offered presentation of the samples in a spinning cup module to provide a measurement stable against sample orientation. The sample cups were filled with 9 g of bauxite powder which was levelled and compacted at 3.2 kPa (382 g compression mass). The sample cup was placed into the spinning cup module and rotated during the measurement. For each sample a reference was collected, followed by ten spectra measured over the instrument's full range, 25 000–4000 cm^{-1} (400–2500 nm). The Model 6500 had a constant 10 nm resolution (37 cm^{-1} resolution at 6060 cm^{-1}) and performed measurement at 2 nm intervals (7 cm^{-1} resolution at 6060 cm^{-1}). An individual spectrum was an average of 32 forward and backward scans and had a 35 s acquisition time. Sample cups were cleaned using ethanol and a low-lint soft tissue Kimwipe (Kimberly-Clark Australia Ltd, 1098).

Table 1. Calibration models and predictive results for both systems.

System	Treatment	RMSEP	R^2	PLS factors
MPA[a]	AS	1.3	0.989	4
MPA[a]	SNV, SG2, AS	0.94	0.993	5
MPA[a]	*AS, OSC2*	*0.89*	*0.994*	*2*
6500	AS	1.4	0.986	4
6500	AS, OSC2	1.1	0.989	1
6500	**SG1, AS, OSC1**	**0.99**	**0.992**	**6**

[a] Optimal MPA models developed at a 16 cm^{-1} resolution.

The order of sample presentation was randomized by sample number, ensuring no correlation between the order of measurement and mineral content. The same sample order was used for both systems.

A number of 'runs' (using identical sample presentation methods described above) were conducted to obtain the relevant data sets to assess robustness of the models with respect to the four experimental aspects: drift, repeated measurement, operator handling and replicate measurement (repacking). These runs are described below.

Due to the limited availability of the MPA instrument, the measurements to assess the different aspects of the experimental error were conducted at 8 cm^{-1} prior to the investigation on the impact of spectral resolution [13] which identified a 16 cm^{-1} resolution for model development. Consequently models developed at 8 cm^{-1} for the experimental error analysis did not perform as well as the optimized models at a 16 cm^{-1} resolution reported in table 1.

Run 1 (full set measurement)—the spectral data for all 263 samples. Ten consecutive measurements were made for each sample before progressing to the next. This set served as a basis for model development, optimization and comparison of the two systems.

A subset of run 1 consisting of spectral data of ten randomly selected samples was used to assess the instrumental error. The calibration models were developed on the remaining 253 samples and then used to predict on the ten-sample set.

Run 2 (drift)—spectral data of all 263 samples were measured a second time using the same procedure as for run 1. The measurement of the samples for run 2 was conducted at least one month after the first. This data set was used to test the models for stability against drift.

Run 3 (operator)—each of the ten samples used to assess the instrumental error was measured ten consecutive times. Between each successive measurement the sample was removed from its holder (carousel or spinning cup module) and then replaced. This data set was used to assess the operator experimental error and also encompasses elements of the instrumental error.

Run 4 (replicate)—using the same ten samples as in run 3, ten consecutive spectral measurements were made for each sample, repacking the sample between each successive measurement. This data set was used to assess the replicate measurement error, which encompasses the operator and instrumental errors.

3.4. Chemometric model development

Quantitative calibration models were developed using PLS regression. Models were optimized in a three step process: (1) find optimal spectral window, (2) optimize data treatment parameters and (3) find the best permutation of data treatments.

The best contiguous spectral window was found by iteratively developing PLS models based on the autoscaled data for each possible contiguous window. The best spectral window was identified by the model producing the lowest RMSEP.

Once the spectral window had been determined, the data treatment needed to be optimized. The data treatments trialled included autoscaling (AS), mean centring (MC), first- and second-order Savitzky Golay derivatives (SGx where x is the order of differentiation), multiplicative scatter correction (MSC), standard normal variate (SNV) and orthogonal signal correction (OSCx where x is the number of components to use in the correction). The first step was to optimize the 'polynomial order' and 'window width' parameters for Savitzky Golay derivative and the 'tolerance' parameter for orthogonal signal correction. This was again achieved through a linear search of possible parameter values selecting those that produced a PLS module with the lowest RMSEP. Once parameters were set, all possible permutations of data treatments were trialled and PLS models developed on the treated data. The final calibration was selected as the model with (significantly) the lowest RMSEP and simplest data treatment.

3.5. Software

The manufacturers' recommended software OPUS 5.5 (Bruker Optik GmbH, 2005) and Vision 2.51 (Foss NIRSystems, Inc. 2001) were used to control and interface the MPA and Model 6500 instruments respectively. Data treatment, PLS modelling and noise calculations were performed using Matlab® 7.1 (Mathworks, Inc. 2005) with PLS Toolbox 3.5 (Eigenvector Research, Inc. 2004). The selection of spectral window, data treatment parameters and the search of different data treatment permutations were accomplished using Matlab scripts.

4. Results/discussion

4.1. Accuracy

The spectra of three bauxite samples (high medium and low kaolinite concentration) measured using both the MPA and
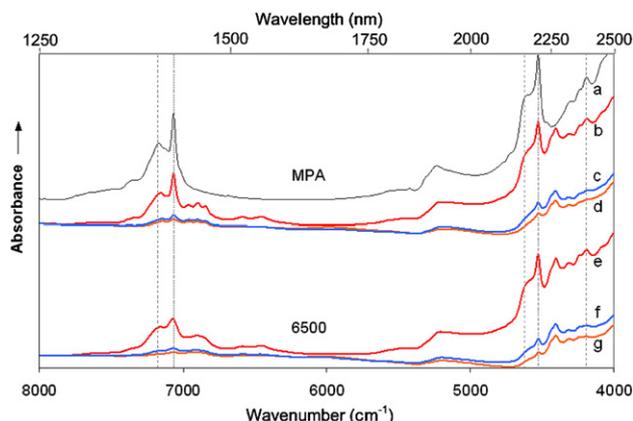

Figure 2. (*a*) Pure kaolinite spectrum acquired on the MPA. (*b*)–(*d*) Bauxite spectra acquired using the MPA system with high, medium and low amounts of kaolinite respectively. (*e*)–(*g*) Bauxite spectra acquired using the Model 6500 system with high, medium, low kaolinite content respectively. Vertical grey lines mark dominant absorption bands. Spectra offset for clarity.

Model 6500 systems are shown in figure 2 below. A pure kaolinite spectrum measured using the MPA system is also shown. The Weipa bauxite samples were rich in kaolinite and gibbsite and consequently these two minerals dominated the NIR spectrum. The bauxite samples in figure 2 were selected to have relatively low gibbsite content to best illustrate the impact of the kaolinite mineral on the bauxite spectrum. Five dominant kaolinite bands were most prevalent in the bauxite spectra: four peaks at 7176, 7067, 4527, 4193 cm^{-1} and a shoulder at 4622 cm^{-1}. The peaks at 7176 and 7067 cm^{-1} are attributed to the first overtones of the fundamental OH vibrations at 3669 and 3619 cm^{-1} [3]. The remaining bands 4622, 4527 and 4193 cm^{-1} attributed to combinations of the fundamental Al–OH stretching and deformation modes [5]. The broad band located at 5100 cm^{-1} increases with increasing kaolinite content but does not originate from molecular vibrations from the kaolinite structure. Rather, the 5100 cm^{-1} band is due to absorbed or coordinated water resulting from a shearing of the kaolinite layers during grinding [5, 6].

The difference in resolution of the two instruments was obvious in the OH overtone region (6400–7400 cm^{-1} [5]). The MPA system had a constant wavenumber resolution of 16 cm^{-1} whilst the Model 6500 had a constant wavelength resolution of 10 nm. At lower wavenumber, such as the Al–OH combination region (4000–4800 cm^{-1}), the resolution was visibly similar for both instruments. However, at higher wavenumber, in the OH overtone region, the MPA provided much better spectral definition, particularly around the 7065 cm^{-1} peak. Additionally, the MPA system also resolved a fine peak structure: 6958, 6896 and 6850 cm^{-1}.

The SNR was evaluated at each wavelength for each instrument and is presented in figure 3. The obvious observation was that the MPA system demonstrated a superior SNR in all wavelength regions. The superior SNR was purely a result of a stronger absorbance signal and not lower signal noise. The stronger absorbance signal was attributed to the high optical throughput of the Fourier transform instrument in combination with the reduced sample density which would
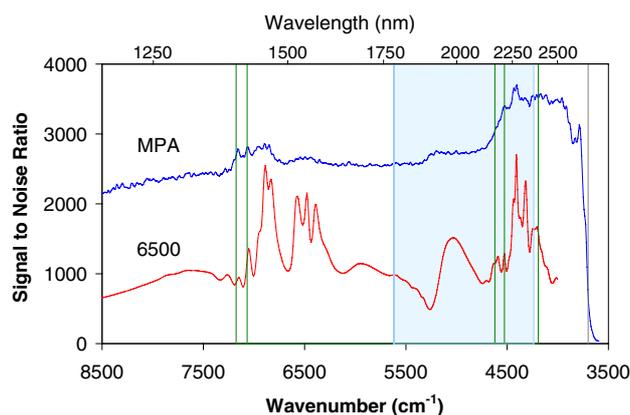
Figure 3. Signal-to-noise ratio for each wavenumber for both MPA and Model 6500 systems. Vertical green lines mark the position of dominant kaolinite bands. The dashed vertical line denotes the end of MPA window whilst the shaded area denotes the spectral window selected for the Model 6500.

increase average sample reflectance. Another observation was that the Model 6500 system had a poor SNR at the kaolinite bands, particularly in the OH overtone region where the kaolinite bands correspond to local minima.

The spectral windows for both systems were very different. The MPA system achieved best calibration of autoscaled data over the 8500–3700 cm^{-1} window whilst the best window for the Model 6500 was determined to be 5620–4240 cm^{-1}. The two spectral windows are also shown in figure 3.

Extension of the spectral window into the noisy lower wavenumbers was highly beneficial to model performance for the MPA system. This result was attributed to the presence of the fundamental OH vibration located at 3694 cm^{-1} [3]. Extension of the window beyond 3700 cm^{-1} proved fruitless and calibrations degraded as a direct result of the low signal to noise at the end of the PbS detectors useful range. Comparatively, the Model 6500 made use of a small spectral window, concentrating on the Al–OH combination bands at 4527 and 4622 cm^{-1} with calibrations improving marginally by omission of the overtone region around 7000 cm^{-1}. The slight improvement observed by utilizing the region containing only the Al–OH combination bands was best explained by the greater SNR observed at the 4527 and 4622 cm^{-1} bands, providing more stable data at the influential wavenumbers for subsequent calibration.

Trialling of the many permutations of MC, AS, SNV, MSC, first and second order SG derivatives and OSC was a computationally and temporally expensive process. The single most beneficial treatment was rudimentary AS which yielded very similar prediction results, RMSEP $\sim$ 1.35%m/m, on both systems requiring four PLS factors in the final model. Applied individually, SNV, MSC and derivation provided no additional benefit for calibration, suggesting that baseline effects and multiplicative scatter effects were not the most dominant sources of spectral variation in either system. However, when applied prior to AS, some combinations of SNV/MSC and derivatives proved beneficial. OSC (preceded by AS) performed well on the data sets for both systems, offering a 30% reduction in the RMSEP over the autoscaled model and requiring only two PLS factors.
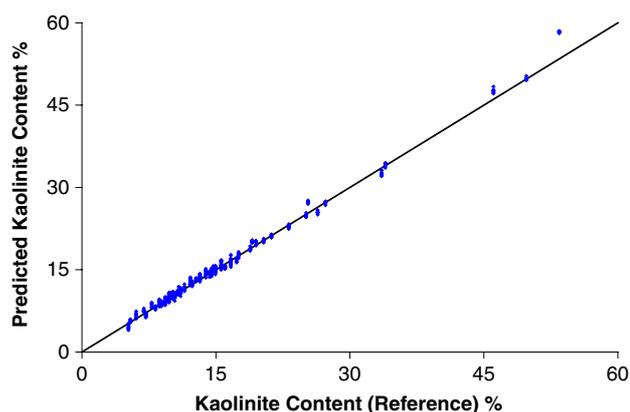

Figure 4. Predicted versus reference values for the AS, OSC2 kaolinite model.
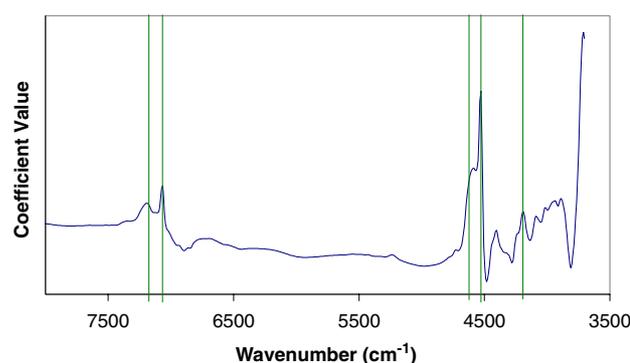
Figure 5. Model coefficients for the MPA AS, OSC2 model. Vertical lines mark the position of reported kaolinite bands.

Approximately 1200 calibration models were developed across both systems; the two best models for each instrument are shown in table 1 with the models developed from autoscaled data for reference. For the MPA system, the best calibration using spectral data collected at a 16 cm^{-1} resolution that was autoscaled with two-component OSC. The Model 6500 achieved best calibration results by first derivative, AS and single-component OSC.

The autoscaled two-component OSC calibration for the MPA system was significantly better than all other models (at the 5% level) achieving an accuracy of 0.89% with the R^2 value of 0.994 requiring only two PLS factors. Predicted kaolinite content against the reference kaolinite content is shown in figure 4. The predicted values were tightly clustered around the ideal predicted-equals-reference line with the greatest residual error arising from the samples of higher kaolinite concentration, a result directly reflective of the skewed kaolinite distribution within the calibration set. The model coefficients (figure 5) for the resulting calibration provide confidence in the robustness of the model, displaying a smooth spectral-like form with the positive and most influential coefficients aligned with the kaolinite bands observed in the bauxite spectra. The importance of extending the spectral window to 3700 cm^{-1} to capture variation of the 3694 cm^{-1} fundamental was justified by the large magnitude of the coefficient in this region. The coefficients also confirmed that the band at 5100 cm^{-1} had little correlation to kaolinite content.
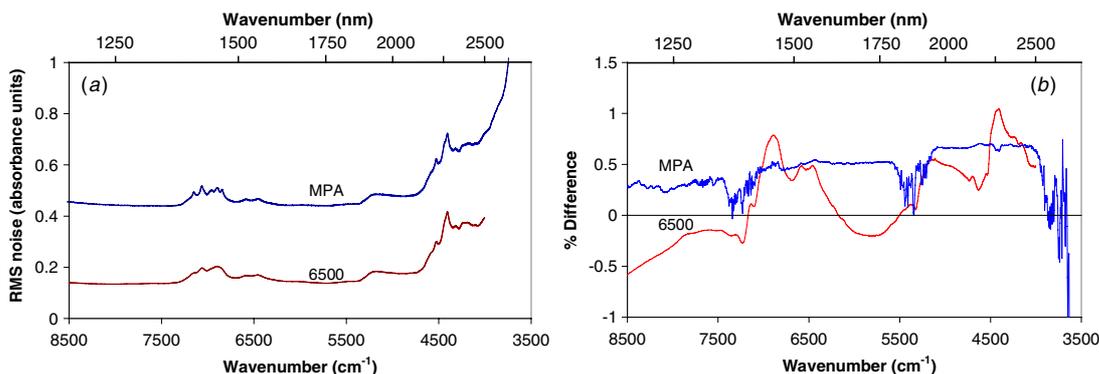
Figure 6. (*a*) Mean spectra on the MPA and Model 6500 systems for run 1 and run 2. Mean spectra for both runs lie on top of each other. (*b*) The percentage difference between run 1 and run 2 for the MPA and Model 6500 systems.

Table 2. RMSEP (%m/m) on the validation sets for the first and second runs (R1 and R2 respectively) for both MPA and Model 6500 systems. Models were developed from the calibration set for the first run and the calibration set for the first run extended with 20 samples of the second run.

Model	System	Model using run 1 calibration set		Model using extended set	
		R1	R2	R1	R2
Autoscaling	MPA	1.34	1.42	1.33	1.40
Autoscaling	Model 6500	1.36	1.18	1.36	1.17
AS, OSC2	MPA	1.09	1.23	1.11	1.14
SG1, AS, OSC2	Model 6500	1.00	0.95	0.99	0.95

MPA models developed from data measured at an 8 cm^{-1} resolution.

4.2. Robustness against aspects of the experimental method

4.2.1. Time stability (drift). The first aspect to be assessed was the time stability of the systems. The mean spectra and difference between the mean spectra of the validation sets of run 1 and run 2 for each system can be seen in figures 6(*a*) and (*b*), respectively. The mean spectra were very similar for both runs on each system. Both systems displayed an average increase across the spectral range in the order of 0.5%. The MPA showed a relatively constant offset of spectra interrupted by minima around 5341 cm^{-1} and 7343 cm^{-1} reflecting changes in the ambient humidity and temperature, coinciding with a failing air-conditioning unit.

The mean difference between runs on the Model 6500 system had an unexpected spectral profile. It was not anticipated that the sample structure changed at all during the month between measurement and it was difficult to explain this profile in terms of instrumental drift. It was suspected that the difference in the mean spectra was due to a change in the experimental method as the operator became more adept at preparing the samples.

Table 2 shows the RMSEP for models developed using the calibration set from the first run (column 3), and the calibration set of the first run extended with 20 samples from the second run (column 4), when used to predict on the validation sets of the first and second runs. The autoscaled models for both the MPA and Model 6500 are included to give a reference and demonstrate the impact of the data treatments.

Table 3. Averaged signal-to-noise ratio for the instrument, handling and repack sets on the MPA and Model 6500 instruments.

	Window (cm^{-1})	Instrument	Handling	Repack
MPA	8500–3700	3000	160	130
6500	5640–4240	1600	700	84

The autoscaled models show a distinct difference in the accuracy of predictions between the two runs. The MPA predicts worse on the second run whilst the Model 6500 predicts far better on the second run, in agreement with the prior suggestion that the operator is learning the presentation method.

Treating the MPA data with AS and OSC improved the predictions on both sets. However, the optimized MPA model is still less accurate on the second run, indicative of a unique spectral variation introduced in the second run. The model was stabilized by inclusion of spectral data of 20 samples from the second run in the calibration set without sacrificing accuracy suggesting this method as an acceptable method for model maintenance [14].

Application of the SG1, OSC1 data treatment to the Model 6500 spectral data yielded approximately equal results when predicting on the validation sets of either run 1 or run 2 demonstrating time stability.

4.2.2. Repeat, operator and repack error. The RMS residual spectra (refer to figures 7(*a*) to (*c*)) and averaged SNR (refer table 3) were calculated for the instrument, handling and repack data sets.

The MPA has a clearly superior instrument noise (refer figure 7(*a*)) with an averaged signal to noise almost double that of the Model 6500 due to both slightly improved RMS noise and superior signal strength.

The operator's handling of the sample produced great spectral variation in the MPA system, comparable to that of repacking the sample. The origin of these variations was attributed to changes in the scattering profile due to either redistribution of the powder sample by mechanical vibration or differences in orientation of the sample to the beam when placed in the sample carousel. This conclusion was confirmed by figure 7(*b*). The MPA operator RMS spectrum displayed a footprint of the bauxite spectrum—a result best explained
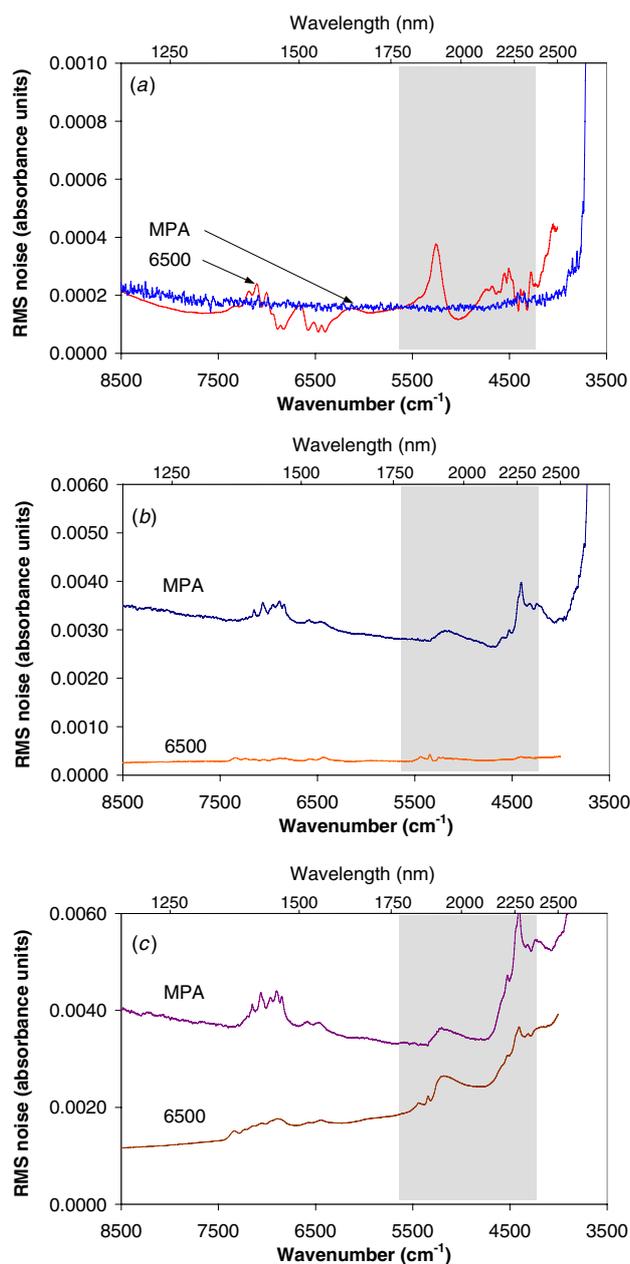
Figure 7. RMS residual spectra for (*a*) instrument sets, (*b*) handling and (*c*) repack sets. The shaded area depicts the Model 6500 window.

a changing degree of multiplicative scatter. The Model 6500 system on the other hand displayed much greater stability against handling of the sample. This result was most likely due to the rotation of the sample during measurement, averaging differences in scatter due to sample orientation. However, it may also be due to stabilization of the powder due to the compression of the sample.

RMS noise for both systems was comparable for replicate measurements. The main source of spectral variation for both systems was believed to be a changing scattering profile. This would appear valid for both systems where, again, the RMS noise spectra (figure 7(*c*)) display a footprint of the bauxite spectrum.

Table 4. Standard deviation (%m/m) of model predictions for each experimental aspect.

Model	System	Repeated	Operator	Replicate
Autoscaling	MPA	0.04	0.21	0.18
Autoscaling	Model 6500	0.21	0.11	0.19
AS, OSC2	MPA	0.10	0.39	0.36
AS, OSC2[a]	MPA	0.19	0.23	0.22
SG1, AS, OSC1	Model 6500	0.17	0.17	0.18

[a] Calibration set of run 1 extended with 20 samples from run 2.

The variation in predictions on the different data sets yielded results that were, for the most part, in agreement with the signal-to-noise values of table 3, that is, smaller spread of predictions for higher SNR. The repeatability values calculated for each data set for both autoscaled and the optimized models are shown in table 4. Autoscaled models were also used as a rudimentary reference to examine the impact of data treatment on the spread of predictions.

For autoscaled data, the MPA model displayed excellent repeatability for consecutive measurements (0.04%m/m) and displayed repeatability in the order of 0.2%m/m for both handling and repacking of the sample.

The Model 6500 system produced interesting results for the autoscaled data, displaying a large error in repeated measurements (0.21%m/m) whilst the error in operator aspect was considerably less (0.11%m/m). This counter-intuitive result was attributed to heating of the sample with sustained exposure to the NIR beam (order of minutes) during the ten repeated measurements. The RMS spectra for the repeated measurement set (figure 7(*a*)) displayed a spurious variation in the region about 5257 cm^{-1} (overtones of water [4, 5]) which consequently aligns with large model coefficients. The 5257 cm^{-1} peak was not present in the RMS spectrum of the operator set since there is a delay between the repeated measurements during the removal/replacement of the sample from/to the holder. Indicative of the SNR, the repack error was identical to that of the MPA system (∼0.2%m/m).

The optimized model for the Model 6500 system was robust to experimental variations demonstrating good stability with a constant experimental error of approximately 0.2%m/m across each aspect. The optimized model for the MPA system was much less stable than the autoscaled model, with the spread of the predictions almost doubling in each aspect. For the MPA system, removal of major spectral variation uncorrelated to kaolinite concentration through OSC emphasized minor, and weakly-correlated, spectral variations introduced though sample-to-sample variations in the experimental method. Extending the calibration set with 20 replicate measurements eliminated the weak correlations, creating a robust model matching the stability of the Model 6500 system.

4.3. Throughput and ease of use

This study was aimed to directly compare the MPA and Model 6500 spectral systems consisting of both the instrument and presentation method. The two important points of difference between the systems were the sample throughput and ease of use.

The MPA system had a 30 s preparation time (dispensing 5 g of sample into a disposable glass vial) and a 9 s spectral acquisition time with batch measurement of up to 30 samples.

Conversely, the Model 6500 required approximately 30 s to dispense, level and compact the sample, 35 s of spectral measurement, and an additional 45 s to clean the sample cup after the measurement. The process was also complicated by scratching of the quartz window of the sample cup with frequent use. The scratching was believed to originate from abrasion from hard minerals when wiping the window clean. This would need to be addressed if the Model 6500 system was to be adopted for large-scale analysis. In terms of both throughput and ease of use the MPA system was clearly superior with a 30 s sample preparation and 9 s spectral measurement, with the ability of 30-sample batch measurement.

5. Conclusion/recommendation

Two NIR spectroscopic systems were developed, trialled and compared for the purpose of quantitative measurement of kaolinite content in Weipa bauxites. For both systems, PLS calibrations were optimized by an iterative selection of spectral window and data treatments. The FTNIR MPA-based system achieved a predictive accuracy of 0.89%m/m with R^2 of 0.994. The pre-dispersive grating based Foss Model 6500 achieved a predictive accuracy of 0.99%m/m and R^2 of 0.992. The MPA system produced the most accurate calibration, significant at the 5% level.

Both systems were robust against the repeated measurement, operator handling and replicate measurement (repacking), with the MPA requiring replicate measurement of a small number of samples. Time stability demonstrated a need for both systems to keep models updated with new samples (frequency yet to be determined) to capture new variations due to changes in instrument behaviour, experimental method, or environmental conditions.

The MPA system displayed a number of practical advantages over the Model 6500 system including preparation time, acquisition time, ease of sample preparation and batch measurement of samples.

In conclusion, the MPA system was recommended for the quantification of kaolinite within Weipa bauxites due to its superior accuracy, throughput and ease of use.

Acknowledgments

References

[1] McCormick P G, Picaro T and Smith P A I 2002 Mechanochemical treatment of high silica bauxite with lime *Miner. Eng.* **15** 211–4

[2] Bakeev K A 2004 Near infrared spectroscopy as a process analytical tool part I: laboratory applications *Spectroscopy* **18** 32–5

[3] Petit S, Decarreau A, Martin F and Buchet R 2004 Refined relationship between the position of the fundamental OH stretching and the first overtones for clays *Phys. Chem. Miner.* **31** 585–92

[4] Carmody O, Kristof J, Frost R L, Mako E, Kloprogge J T and Kokot S 2005 A spectroscopic study of mechanochemically activated kaolinite with the aid of chemometrics *J. Colloid Interface Sci.* **287** 43–56

[5] Frost R L, Mako E, Kristof J and Kloprogge J T 2002 Modification of kaolinite surfaces through mechanochemical treatment - a mid-IR and near-IR spectroscopic study *Spectrochim. Acta* A **58** 2849–59

[6] Ruan H D, Frost R L and Kloprogge J T 2001 Application of near-infrared spectroscopy to the study of alumina phases *Appl. Spectrosc.* **55** 190–6

[7] Crowley J K and Vergo N 1988 Near infrared reflectance spectra of mixtures of kaolin group minerals: use in clay mineral studies *Clays Clay Miner.* **36** 310–6

[8] Walker G S 1994 *Near Infrared Analysis as a Means of Characterising Selected Minerals and Bauxite Pisoliths* ed G D Battern, P C Flinn, L A Welsh and A B Blakeney (Lorne, Australia: Near Infrared Spectroscopy Group) pp 12–8

[9] Walker G S 1993 *VIS-NIR and MIR Analysis of Selected Minerals and Bauxite Pisoliths* ed J Graham, B Griffin and H Linge (Fremantle, Australia: CSIRO W.A. Laboratories) pp 238–41

[10] Walker G S 1996 *Use of Near Infrared for the Quantitative Analysis of Minerals* ed A Niedbalska, A Szymanski and A Wiewiora (Warsaw, Poland: Warsaw University of Technology) pp 85–9

[11] Barboza F D and Poppi R J 2003 Determination of alcohol content in beverages using short-wave near-infrared spectroscopy and temperature correction by transfer calibration procedures *Anal. Bioanalytical Chem.* **377** 695–701

[12] Fearn T 1996 Comparing standard deviations *NIR News* **7** 5–6

[13] Mcarthur L and Greensill C 2007 Impact of resolution on NIR PLS calibration of kaolinite content with Weipa bauxite *Meas. Sci. Technol.* **18** 1343–7

[14] Greensill C V, Wolfs P J, Spiegelman C H and Walsh K B 2001 Calibration transfer between PDA-based NIR spectrometers in the NIR assessment of melon soluble solids content *Appl. Spectrosc.* **55** 647–53