Fuel 89 (2010) 760–767

Contents lists available at ScienceDirect

Fuel

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

Remote bitumen content estimation of Athabasca oil sand from hyperspectral infrared reflectance spectra using Gaussian singlets and derivative of Gaussian wavelets

D. Lyder [a], J. Feng [a], B. Rivard [b,*], A. Gallie [c], E. Cloutis [d]

[a] Earth Observation Systems Laboratory, Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alta., Canada T6G 2E3
[b] Department of Geography, University of Victoria, 3800 Finnerty Rd, P.O. Box 3060, STN CSC, Victoria, BC, Canada V8W 3R4
[c] Department of Earth Sciences, Laurentian University, Sudbury, Ont., Canada P3E 2C6
[d] Department of Geography, University of Winnipeg, Winnipeg, Canada

ARTICLE INFO

Article history:
Received 30 October 2008
Received in revised form 20 February 2009
Accepted 19 March 2009
Available online 9 April 2009

Keywords:
Sand
Bitumen
Infrared

ABSTRACT

Modeling of the total bitumen content, TBC, in Athabasca oil sands was undertaken on the basis of its hyperspectral reflectance spectra. Spectra (8 cm^{-1} resolution) were obtained that covered both the short-wave infrared and thermal infrared (TIR: 3.00–30.00 μm). Two methods, Gaussian fitting and wavelet analyses, were investigated to identify useful bitumen features as well as the best removal of the baseline. We aim to obtain the best determination of the TBC for a suitable suite of test and validation oil sands samples. The Gaussian model relied explicitly on features at 2.282 and 2.532 μm though these were only two of 10 features simultaneously fit with a quadratic baseline to the range of 2.230–2.603 μm of the spectra. The wavelet model relied on bitumen features selected at 2.274, 2.396 and 3.725 μm that could be isolated from the baseline and noise. Both models yielded similar dispersion in their estimates of TBC (±~1–2%) while the wavelet model proved to be more robust when applied to the validation and blind data suites. We also considered the effects of using the L2 optimization (classical least-squares) and L1 optimization (minimization of largest outlier) schemes for both models. Both schemes produced similar results for the model suite of samples for TBC but the L1 was superior when applied to the validation and blind data suites. The wavelet model using the L1 optimization appeared to be quite robust producing estimates of TBC (±~1.0–1.7%).

1. Introduction

Oil sands are mixtures of quartz, clay, water, bitumen and accessory minerals. The Athabasca oil sands are a vast reserve of bitumen (very heavy oil) estimated to contain about 1.7 trillion barrels of which 170 billion are economically recoverable. About 35 billion barrels are accessible using surface mining techniques. Current production of oil from oil sands is approaching one million barrels a day.

With the recovery methods most commonly used with surface mining, oil sand ore is excavated, then mixed with heated water to form an aqueous slurry from which oil is encouraged to separate in large settling vessels. The separation process is enhanced by the use of chemical additives, aeration, and mechanical and thermal energy, as well as careful control of throughput rates and water additions.

The efficiency of the bitumen recovery process is a function of bitumen content, the type and amount of clays present, the degree of weathering of the oil sand sample before processing and the temperature of the slurry [1,2]. The ore deposits are not homogeneous and the ore can display considerable variability in water, clay and bitumen content and in particle size distribution [3,4]. Because bitumen recovery is related to the bitumen, fines and clay content in ore, the efficiency of extraction operations can potentially be improved with timely estimates of the physical and chemical properties of oil sands. Hence, there is a need for a robust method to estimate total bitumen content in real time.

This paper investigates the use of diffuse reflectance spectroscopy for the estimation of bitumen content in oil sand ores. Infrared reflectance spectroscopy has been used in a number of oil sand studies to investigate for example variations in oil sand slurry [5], the mineralogy of oil sands [6], bitumen characteristics [7], and bitumen content in ground ore [8]. Syncrude Canada has been using a broad band analysis of reflected light [9] (a ratio of peak bitumen absorption to baseline) in the short-wave infrared wavelengths (SWIR: 1.60–3.00 μm) as the oil sand as it is conveyed to be processed. As part of this work, we simulated the broadband method reported by Thompson (1983) [10] and obtained a precision in the estimation of TBC of ±~3% based on a small sample. The preci-

* Corresponding author. Tel.: +1 250 853 3280; fax: +1 250 721 6216.
E-mail address: benoit.rivard@uvic.ca (B. Rivard).

0016-2361/$ - see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.fuel.2009.03.027

sion currently achieved commercially with a two-wavelength instrument is in the order of 1–1.5% though little has been published recently.

With the goal of improving the current method we examine the selection of different bitumen (or other component) spectral features towards the development of a more robust and accurate model of TBC. To this end, hyperspectral reflectance spectra (8 cm^{-1} resolution) were obtained that covered both the short-wave infrared and thermal infrared (TIR: 3.00–30.00 μm). Identification of usable bitumen spectral features in Athabasca oil sands in the SWIR has already been undertaken [5,6,9–11,12]. Laboratory spectroscopic studies on oil sands in the thermal infrared have also been undertaken [12–18], though little research has been conducted on its use in a real-time estimation of TBC. Here we examine the combination of selected spectral features to improve the modeled determination of TBC and we select features using two methods: the first uses classical Gaussian singlets (Gaussians) with an appropriately chosen baseline to model the hyperspectral spectra; the second is a wavelet method that relies on the difference of Gaussian wavelet [19,20] to model the power or influence of bitumen at different scales of the spectra. Both methods do not require a priori knowledge of the location of absorption features associated with the presence of bitumen. We aim to obtain the best determination of the TBC for a suitable suite of test and validation oil sands samples using Gaussians and wavelets. In a subsequent study these results are used to define broadband models for the rapid estimation of TBC.

2. Experimental section

2.1. Sample suite

A total of 51 samples were collected by mine geologists from various mines of the Athabasca oil sands of Alberta, Canada. They represent samples from routine mining operations collected over a variety of spatial locations (39 hand samples from different mine faces) and depth (12 samples from different depth of a drill core). Samples weighting from 0.5 to 1 kg were stored in plastic bags at a temperature of −5 °C to prevent loss of water and aromatic components. Prior to spectral measurements the samples were allowed to equilibrate to room temperature and then placed in a tray to emulate samples observed on a conveyor belt. Thus the sample were neither cut nor pressed but occurred in a fragmented state. Routine laboratory analyses were conducted to measure the bitumen, water and solid contents in each sample. The bitumen content ranged from 3% to 16% with water content varying from 0.5% to 12%.

The samples were divided into a modeling suite to establish TBC predictive models, and a validation suite and blind suite to evaluate the models. The modeling suite was taken entirely from one suite of mines and comprised 12 samples from drill cores of various depth and nine samples from mine faces. All validation samples are from a separate mine (15 samples in total) from which the blind samples (another 15 in total) were also collected. Using samples from different mines for modeling and validation served as a natural division of the data for the analysis and provided a test of the general applicability of the derived models. The distribution of the bitumen content was fairly uniform over the sample range for the model suite and blind suite but clustered into two main wt% for the validation suite (at ~7% and 15%). The model and validation suite were obtained and analyzed in 2006, the blind suite was obtained and analyzed in 2007.

2.2. Spectral measurement

Laboratory measurements were undertaken on all samples using a Bomen MB102 Fourier Transform Infrared (FTIR) spectrometer equipped with a Mercury/Cadmium/Telluride (MCT) detector. The MCT detector measured reflected thermal energy at wave numbers ranging from 500 to 6000 cm^{-1} (1.67–20 μm) that covers the major SWIR and TIR regions. The instantaneous field of view was a circle with diameter of 20 mm. The light source was a globar positioned at an angle 35° incident to the sample surface with a viewing angle at 35° providing a diffuse bidirectional reflectance measurement. Reflectance spectra were obtained from the ratio of each measurement to that of an illuminated diffuse gold panel of known reflectance taken with the same geometry and acquired after the measurements of each sample. Measurements at six different locations on each sample were obtained. Each measurement represents the co-addition of 32 individual spectra at 8 cm^{-1} spectral resolution. The final spectra considered in the TBC modeling were generated by averaging the six measured spectra. The signal-to-noise ratio (S/N) of the final spectra were estimated to be ~300 at 1.67 μm and >>300 at 20 μm (Fig. 1).

2.3. Gaussian decomposition of SWIR spectra

For the Gaussian analysis the spectra were first converted from reflectance to absorbance and fitting was performed using the commercial package Peakfit, Version 4.12 (SeaSolve Software Inc.). Gaussian fitting to the spectra was done in wavelength space conducted over the range 2.231–2.603 μm. Fig. 2 shows the average absorption spectra for the model, validation and blind data suites.

All spectra were analyzed in two steps. In the first, Peakfit was allowed to make an initial guess as to the number of peaks and peak locations on the basis of inflection points within the range of 2.230–2.603 μm of the spectra for a given baseline type, e.g. constant, linear or quadratic, and with Savitsky–Golay smoothing and amplitude rejection threshold. The amplitude rejection threshold serves as a threshold above which Peakfit will consider peaks as being real and is used by Peakfit when selecting portions of the spectra to which it fits the baseline. Setting this too low will allow small spurious features to be modeled as peaks while setting it too high will reject real peaks. From the initial guess of peak locations, a set of fixed peak locations was selected that coincided with the average location of particular peaks that occurred in all spectra. There were peaks in some spectra not common to all but these were small enough to ignore without introducing significant error into neighboring peaks or the baseline. In the second step, the fitting was performed as before except the number of peaks allowed and their locations were constrained to the previously determined average locations, i.e., peak amplitudes and widths and baseline types were allowed to vary. The smoothing and amplitude rejection levels were set to the most common values seen in the first step, i.e., Savitsky–Golay smoothing of 16% window size and amplitude rejection threshold of 12% of the maximum amplitude. Table 1 provides a summary of the peak locations and their possible origin.

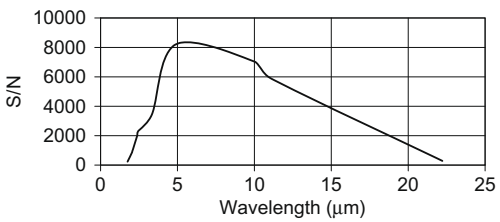

Fig. 1. Average single-scan S/N ratio measured over five days using a 100% reflectance diffuse gold panel and globar illumination. S/N ratio for the diffuse gold panel was adequate, exceeding 1000 over the entire spectral region detected (1.67–20.0 mm). It was estimated for a single scan while measuring a diffuse gold panel, a target of high and uniform reflectance.

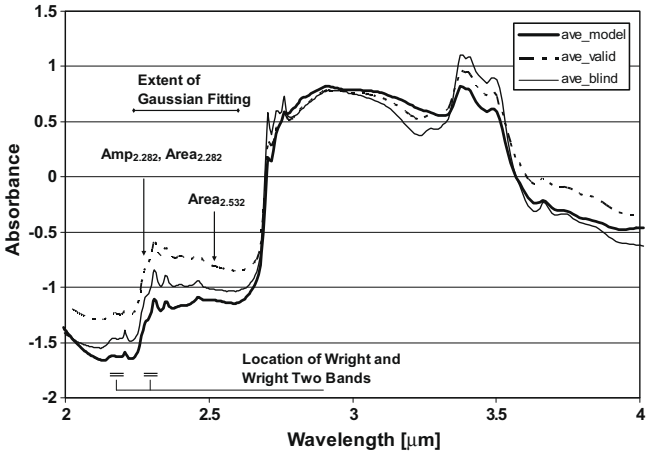

Fig. 2. Average absorption spectra of model suite (21 samples), validation suite (15 samples) and blind suite (15 samples). Sections for Gaussian fitting were initially considered if they had a well bounded baseline, i.e., flat or monotonically decreasing.

For each peak fitted with a Gaussian for a given baseline type, estimates were found of the Gaussian full-width at half-maximum (FWHM), the Gaussian area as determined by numerical integration ($Area_{int}$) and by an analytical solution ($Area_{ana}$), and the Gaussian amplitude (Amp). In most cases $Area_{int}$ and $Area_{ana}$ were essentially the same but differed when the peaks were relatively low amplitude ($Area_{ana}$ tending to be larger). For each type of baseline the coefficients of the polynomial terms defining the baseline were also determined, i.e., B_{const} (0th order term), B_{xterm} (1st order term) and B_{x2term} (2nd order term). In the case of the linear and quadratic models the value of the baseline at the peak centre, B_{peak}, was also determined. Fig. 3 shows an example of a Peakfit solution using the selected peak centers as listed in Table 1 and a quadratic baseline.

2.4. Wavelet decomposition of SWIR and TIR spectra

Wavelet analysis was conducted on the FTIR spectra in order to improve the detection of weak spectral features. The general principle behind this technique [19] is that any signal sampled in a regular manner, say a reflectance spectrum in wavelength in this case (not absorbance), may be represented as a sum of similar wave-like functions (wavelets). As a first step, the mean reflectance of the spectrum calculated over the entire spectral range is subtracted at each band. This mean subtraction is required to ensure that the wavelets are comparable between scales, and that the wavelets from different spectra are comparable at the same scale. The original spectrum can be reconstructed by adding the mean back to the modified spectrum. To compute the wavelets, the original spectrum ($x_{n'}$, n' = spectral band $1 \ldots N$) is convolved with a fixed-shape signal (a wavelet base, ψ) at various scales (S) to extract the spectral information:

$$W_n(S) = \sum_{n'=0}^{N-1} x_{n'} \psi \left[\frac{(n'-n)\delta t}{S} \right] \quad (1)$$

where N is the number of bands and δt is the minimum band interval. The input spectrum must contain a constant band interval and for this study we used as an input the reflectance as a function of wavenumber. For this study, the shape of the wavelet base chosen is the 2nd order derivative of Gaussian (DOG [29]) as explained below. By varying the wavelet scale S and translating the wavelet base (which acts like a filter) along the local band n, the convolution will decompose an original spectrum into various wavelet power (W) spectra at different scales. At any given scale, the amplitude of the wavelet base required to capture the amplitude of features in the original spectrum will be recorded as the power (W). In this study the power spectrum for ten scales was calculated.

The reflectance spectra of oil sand samples encompass a range of electronic and vibrational processes for minerals, water and organic compounds [30,31]. These processes give rise to absorption

Table 1
Summary of average Gaussian peak locations and interpretation.

Section	Average peak location [μm]	Origin	Physical interpretation	Strength[a]	Shape[b]	References
B(2.231–.603 μm)	**2.282**	clay, bitumen	asymmetric CH_2 stretch + asymmetric CH_2, CH_3 bend	m	p	[11]
			OH stretch + Al-OH bend			[21]
			OH stretch + Fe-OH bend			[22]
	2.311	clay, bitumen	asymmetric CH_2 stretch + asymmetric CH_2, CH_3 bend	vw	sh	[11]
			asymmetric C–O stretch			[22]
			Mg-OH stretch			[23]
	2.315	clay, bitumen	asymmetric CH_3 stretch + symmetric CH_3 bend	vw	sh	[11]
			asymmetric C–O stretch			[22]
			Mg-OH stretch			[23]
	2.319	clay, bitumen	symmetric CH_2 stretch + asymmetric CH_2,CH_3 bend	s	p	[11]
			asymmetric C–O stretch			[22]
			Mg-OH stretch			[23]
			OH stretch + Mg-OH bend			[22]
	2.327	clay	OH stretch + Mg-OH bend	w	sh	[22]
	2.353	clay, bitumen, water	symmetric CH_3 stretch + symmetric CH_3 bend	s	p	[11]
			carbonate absorption			[24]
			H_2O, OH			[25]
			asymmetric C–O stretch			[22]
	2.357	bitumen	symmetric CH_3 stretch + symmetric CH_3 bend	vw	sh	[11]
	2.405	clay, water	metal-OH	vw	c	[26]
			OH			[27]
			adsorbed H_2O			[28]
	2.464	Unknown		w	p	
	2.532	bitumen, atmosphere	symmetric + asymmetric	vw	c	[22]
			C–O stretch			[27]
	2.603[c]	unknown				

[a] Strength: vs, very strong; v, strong; m, medium; w, weak; vw, very weak.
[b] Shape: p, identifiable peak; sh, shoulder; c, baseline. Strength and shape are based on the appearance of the original spectra at the peak centers, not the appearance of the individual peaks.
[c] Only used for the constant and linear baseline models. The bold entries indicate the Gaussian peaks used in the final Gaussian (quadratic baseline) models.

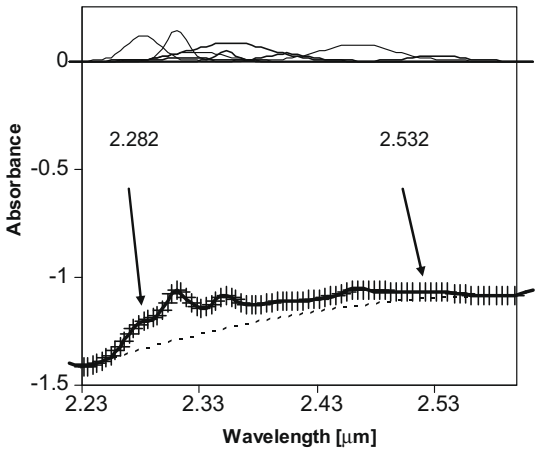

Fig. 3. Example of Gaussian solution for model sample AS (9.88% TBC) derived from Section B. The quadratic baseline (dashed line) and Gaussians for the 10 peaks used in this model are shown (light solid lines). The model derived from the sum of these peaks and baseline is shown as +s overlaid against the actual spectrum (solid line). The two labeled positions indicate the Gaussian peaks used in the final Gaussian models.

features or reflectance peaks depending on the spectral range of observation or physical properties of the sample. Individual spectral features can be represented by Gaussian or quasi-Gaussian functions. Torrence and Compo [19] recommends that the shape of a wavelet base (ψ) reflect the property of features present in the original spectrum. Thus we made use of the 2nd order derivative of Gaussian (DOG) as a base since absorption features of minerals and many other compounds can be approximated by a Gaussian waveform. DOG has a zero mean power, a mathematical prerequisite for the wavelet base used to perform the continuous wavelet transform on discrete data series such as the spectra of this study. The DOG base captures positive and negative oscillations (e.g., peaks and troughs) of the spectrum as separate peaks and troughs in the wavelet power spectra. The DOG base provides sufficient resolution in wavelength for locating features and their inflection points.

An example of the wavelet power for one sample (C712, 16.26% TBC) is provided in Fig. 4. The lower valued scales are sensitive to the small scale structure of the spectra while the higher valued scales are sensitive to the large scale structure (i.e., the baseline). To determine the spectral features sensitive to the variation of bitumen content, we examined the correlation coefficient ($R_{feature}$) between the wavelet power at each scale and each wavelength with the known TBC value for the entire suite of model samples. The resulting correlation coefficient at each wavelength and scale can be visualized with the use of a scaleogram shown in Fig. 5 which conveys the amplitude of the correlation coefficient in grey tones. The wavelet feature regions that are most strongly correlated or anti-correlated with TBC, $R_{feature} > 0.85$ and $R_{feature} < -0.85$, respectively, have been highlighted by contours to identify the locations and width of features with highest correlation with TBC. Seven spectral feature regions shown on Fig. 5 have a value of $R_{feature}$ higher than 0.9 or lower than −0.9. In each of these regions the feature with highest correlation with TBC was selected and is listed and identified in Table 2. Six of the seven features (ID: F1 = 1.710 μm, F2 = 1.754 μm, F5 = 2.274 μm, F6 = 2.396 μm, F8 = 3.672 μm, F9 = 3.725 μm) are negatively related to TBC and correspond to bitumen absorption features in the reflectance spectra and in the wavelet power at the proper scale. Only one of the seven features (F4) is positively correlated to TBC, which may be due to the absorption feature of clay in reflectance spectra.

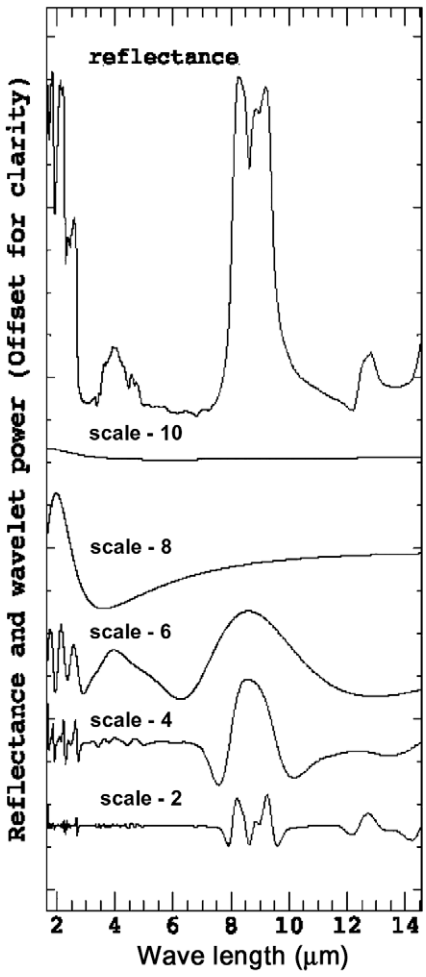

Fig. 4. Example of single wavelet power spectrum resulting from analysis of reflectance spectra for the model sample C712 (16.26% TBC). Lower valued scales represent the variation of the spectrum over small scales. Larger scales represent the large structure usually identified with the baseline.

3. Analysis and results

3.1. Gaussian estimation of TBC

An estimate of TBC based on the fitted Gaussian parameters for a given baseline was derived assuming a linear model correlating these parameters with TBC. Two optimization schemes for fitting a linear model were considered: the traditional least-squares optimization (L2 norm) and the minimization of outliers (L1 norm). For both optimizations the linear model was constructed by first selecting the Gaussian parameter that had the highest individual correlation with TBC. Then, additional parameters not already included in the current model were added to the model one at a time in order of decreasing correlation of the individual parameters with TBC. An *F*-test at the 95% confidence limit was conducted to determine if the additional parameter provided a statistically significant improvement over the current model [34]. After all of the unused parameters were tested the parameter that provided the most significant improvement was then added to the current model. The process was repeated until no further significant improvement in R^2_{Model} occurred.

Several models possessed very similar R^2_{Model} values. The most robust model should provide the best estimate of both the validation and blind bitumen values as measured by the mean error μ_{Valid} and dispersion σ_{Valid} between their respective estimated and true

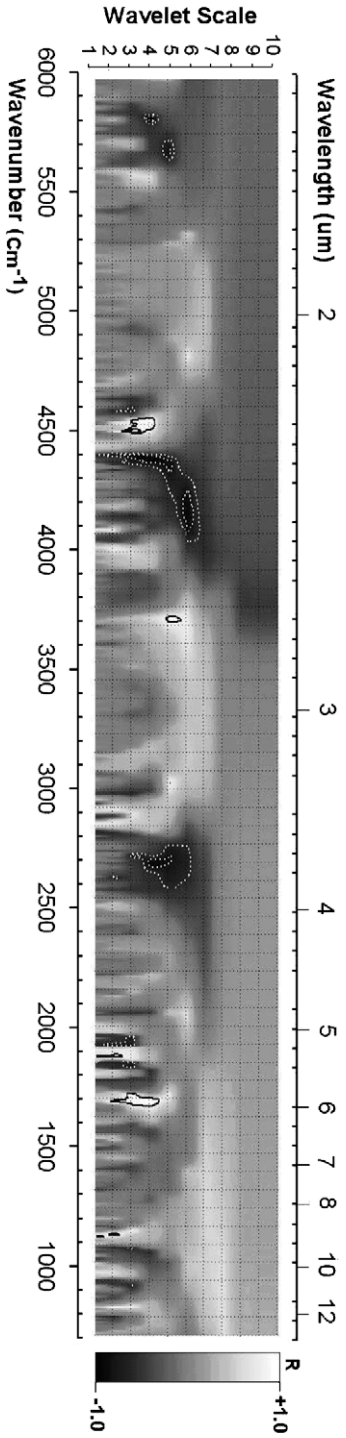

Fig. 5. Scaleogram of the correlation coefficient, $R_{feature}$, between true TBC and wavelet power. The correlation coefficient was established for the wavelet power at each scale and each spectral band (e.g. wavelength) and the TBC for all model samples. Dash contours: $R_{feature} = -0.85$ to -0.9 and $R < -0.9$; Solid contours: $R_{feature} = 0.85$ to 0.9 and $R > 0.9$. The computation of wavelets was conducted in wavenumbers and the non-linear display of wavelength is shown for reference.

values. Hence, the model that produced the highest R^2_{Valid} as well as the lowest μ_{Valid} and σ_{Valid} was selected as being the most robust model. The definitive check on the robustness of the model can then be made by considering how well the solution extends to the blind bitumen values. Using the L2 optimization the most robust Gaussian model of TBC was found to be:

$$\mathrm{TBC}(\%) = -17.84 + 0.84 * Area_{\mathrm{int},2.532} + 1.40 * Area_{\mathrm{int},2.282} + 3.98 * Amp_{2.282} \quad (2a)$$

fit to Section B with a quadratic baseline. Similarly, using the L1 optimization we found:

$$\mathrm{TBC}(\%) = -15.07 + 0.72 * Area_{\mathrm{int},2.532} + 1.26 * Area_{\mathrm{int},2.282} + 5.12 * Amp_{2.282} \quad (2b)$$

Fig. 6 shows the estimated TBC derived from Eqs. 2a and 2b for the model, validation and blind data suites. Table 3 provides a summary of statistics of the fit to both the Gaussian and wavelet models.

3.2. Wavelet estimation of TBC

From the initial set of seven features that best correlated with TBC we examined models of all possible linear combinations of up to five features (~number of model samples$^{1/2}$). The performance of each combination of features was initially judged using an adjusted R^2_{Model} defined as:

$$\mathbb{R}^2_{Model} = 1 - \frac{(n-1)(1-R^2_{Model})}{n-p-1}, \quad (3)$$

where n is the number of observations, p is number of independent variables, and R^2_{Model} is the coefficient of determination. Among all possible models, the model that maximized the R^2_{Model} criterion was considered the best model. R^2_{Model} is used to avoid a bias towards solutions with larger numbers of independent variables (features) that may over fit the data.

Features F5 (2.274 μm), F6 (2.396 μm) and F9 (3.725 μm) were present in the top 10 performing models suggesting the final models need only contain these three bands. The most robust wavelet model of TBC, as derived from the wavelet power ($_sP_\lambda$) for scale (s) and center wavelength (λ) given by these three bands, using the L2 norm was determined to be:

$$\mathrm{TBC}(\%) = 11.84 - 51.08 * {}_4P_{2.274} + 21.65 * {}_6P_{2.396} - 136.84 * {}_4P_{3.725}. \quad (4a)$$

Using the L1 norm optimization the most robust wavelet model of TBC was found to be:

$$\mathrm{TBC}(\%) = 10.86 - 54.37 * {}_4P_{2.274} + 21.71 * {}_6P_{2.396} - 124.32 * {}_4P_{3.725}. \quad (4b)$$

Fig. 6 provides the relevant graphics showing the estimate of TBC derived from the wavelet models.

4. Discussion

4.1. Comparison of wavelet and Gaussian models

In the case of Gaussian fitting, the best model of TBC is formed from the linear combination of parameters from two absorbance peaks (2.282 and 2.532 μm) taken from a simultaneous fit of 10 peaks plus a quadratic baseline for a fixed level of smoothing, for a limited section of the spectrum. For the wavelet model a linear combination of the wavelet power from three locations (2.274, 2.396 and 3.725 μm) at different scales (4, 6, 4, respectively) calculated from the full spectrum, was effective. As such, the two methods are quite different in their approach: the wavelet acts as a series of filters sensitive to structure on a particular scale size while Gaussian fitting simultaneously fits all structures at all scales for a given noise level.

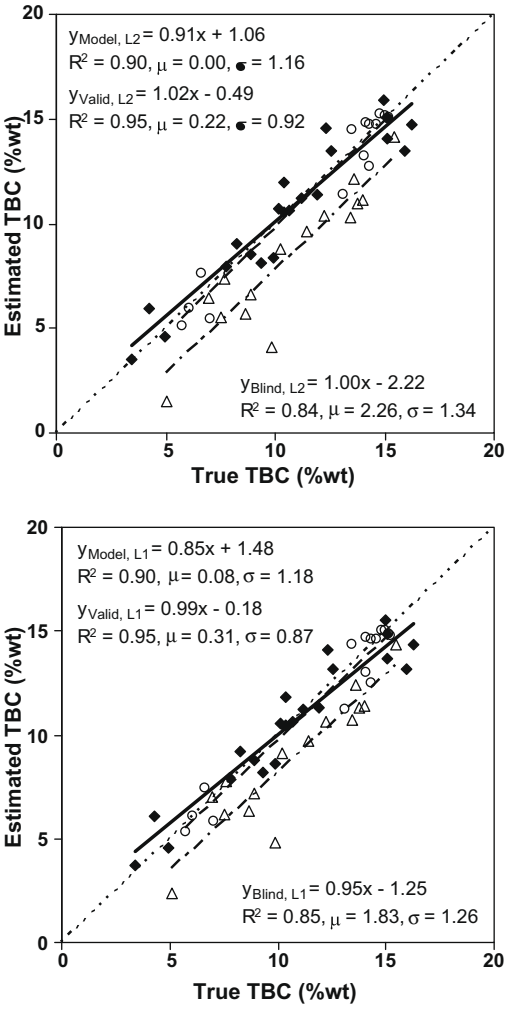

Fig. 6. Estimated TBC derived from L2 norm model (upper panel, Eq. (2a)) and L1 norm model (lower panel, Eq. (2b)) of Gaussian parameters fit to reflectance spectra in Section B assuming a quadratic baseline. The ideal solution would lie along the diagonal: Model – diamonds, solid line, Validation – open circles, short-dash line, Blind – triangles, long-short-dash line.

One advantage of the wavelet approach in this case is that it doesn't rely upon an *a priori* estimate as to where the baseline boundaries are but allows the wavelet to handle it as a large scale component(s). Furthermore, the noise is treated as a separate small scale component. On the other hand, the Gaussian model allows more flexibility in describing the baseline in terms of a variety of functions, not a single wavelet shape.

Both models make extensive use of features in the ~2.3 μm region. The Gaussian model's strongest correlation with TBC is with the amplitude of the 2.282 μm feature but this feature alone is not enough to generate a robust model of TBC. This particular feature is part of a complicated region in which the affect of other mineralogy, i.e., clays and/or water, must be properly removed. Furthermore, the variability in surface texture will affect the degree of volume scattering [35]. The agreement between the model and validation suites (see Fig. 6) suggests that the addition of the 2.532 μm feature is sufficient to adequately model all of these influences. However, the significant mean error (e.g. offset) in the estimated TBC values for the blind suite suggests that some other and unaccounted factor(s) may be at play when estimating the blind TBC values.

The wavelet model shares one feature with the Gaussian model, the scale 4 feature at 2.274 μm which equates to the Gaussian feature at 2.282 μm. The scale 6 wavelet feature at 2.396 μm is probably the same as the Gaussian feature at 2.405 μm but the feature was not selected for use in the Gaussian models. The third scale 4 wavelet feature at 3.725 μm is outside the range examined by Gaussian fitting. The second Gaussian feature at 2.532 μm corresponds with a weak positive correlation feature at small scales that was not selected for use in the wavelet models. The variability of the width and strength of the 2.396 μm wavelet feature (Fig. 5) as a function of scale suggests that it is part of a complex mix of bitumen and non-bitumen components. The 3.725 μm wavelet feature is quite distinct from the 2.274 μm and 2.396 μm features in two fundamental ways. Firstly, it is produced by different physical vibrational transitions (Table 2). Secondly, the baseline is flat in this part of the TIR (Fig. 4 – Scale 10 spectrum) and is dominated by surface scattering [35]. Given these differences, the role of the 3.725 μm feature may be to reduce any correlation between the 2.274 μm and 2.396 μm bitumen features in the wavelet models.

Table 2
Individual detectable features with wavelet power correlated/anti-correlated to model TBC values with absolute $R_{feature} > 0.85$.

Feature ID	Feature center (μm)	Feature range (μm)	Scale	$R_{Feature}$sign[a]	Origin	Physical interpretation	References
F1	1.710	1.703–1.719	4	–(X)	clay, bitumen	aromatic CH stretch + symmetric CH_2 stretch	[11]
						OH stretch	[32]
F2	1.754	1.742–1.766	5	–(X)	clay, bitumen	Al-OH bend	[11,32]
						OH stretch	[33]
F3	2.171	2.168–2.175	2	–	clay	OH stretch	[33]
F4	2.210	2.186–2.220	3	+(X)	clay	OH stretch	[33]
F5	**2.274**	2.258–2.319	4	–(X)	bitumen	asymmetric CH_2 stretch + asymmetric CH_2, CH_3 bend	[11]
F6	**2.396**	2.339–2.469	6	–(X)	clay, bitumen	symmetric CH_3 stretch + symmetric CH_3 bend	[11]
						OH stretch + Al-OH bend	[21]
						OH stretch + Fe-OH bend	[22]
F7	2.770	2.741–2.899	5	+		NH, OH stretch	[33]
F8	3.672	3.662–3.683	3	–(X)		O=CH	
F9	**3.725**	3.672–3.790	4	–(X)		O=CH	
F10	5.206	5.185–5.227	2	–	Unknown		
F11	5.334	5.312–5.357	2	+	Unknown		
F12	5.470	5.447–5.493	2	–	Unknown		
F13	5.865	5.787–6.946	3	+	clay, bitumen	$C\text{-}CH_3$, aldehyde, ketone, H-bonded acid, conjugated C=C and aromatic C=C, amide carbonyls	33
F14	8.940	8.879–9.001	2	+	clay, silicates		
F15	9.603	9.531–9.674	2	–	clay	aliphatic CH	

[a] (X): absolute value of correlation co-efficient > 0.90. The bold entries indicate the wavelet features used in the final wavelet models.

Table 3
Summary of errors for predicted versus measured TBC for model solutions for the Gaussian and wavelet models.

	Gaussian						Wavelet					
	Model		Validation		Blind		Model		Validation		Blind	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
L2	0.00	1.15	0.22	0.92	2.26	1.34	0.00	0.95	−0.64	1.01	−0.36	1.75
L1	0.08	1.18	0.31	0.87	1.83	1.26	0.21	0.97	−0.29	1.01	−0.09	1.63

4.2. Precision of Gaussian and wavelet models

It is clear from the fit to the model and validation data suites that the Gaussian model and wavelet model both yield comparable estimates of TBC (Table 3). The estimate of TBC for the blind data suite is not as good for both models, particularly so for the Gaussian model. For regression analyses where the errors are distributed normally the L2 optimization (classical least-squares) should produce the optimal solution [36]. For situations where there are errors in the data other than random errors, i.e., the reliability of the data values is not uniform, the L1 optimization may better remove this effect. Comparing the precision of the estimates for the model suite (model σ values in Table 3), the L2 optimization is marginally better than the L1 optimization for both the Gaussian (L1 $\sigma = 1.18$, L2 $\sigma = 1.15$) and wavelet (L1 $\sigma = 0.97$, L2 $\sigma = 0.95$) models. However the L1 optimization generally produces lower absolute errors (mean offset + dispersion) than the L2 optimization for both the validation (e.g. wavelet L1 $\mu = -0.29$, L2 $\mu = -0.64$) and blind data suites (e.g. Gaussian L1 $\mu = 1.83$ and $\sigma = 1.26$ while L2 has $\mu = 2.26$ and $\sigma = 1.34$). Hence, the L1 optimization appears to produce better estimates of TBC in general, which suggests there may be some small degree of variable weighting to the data points in the models.

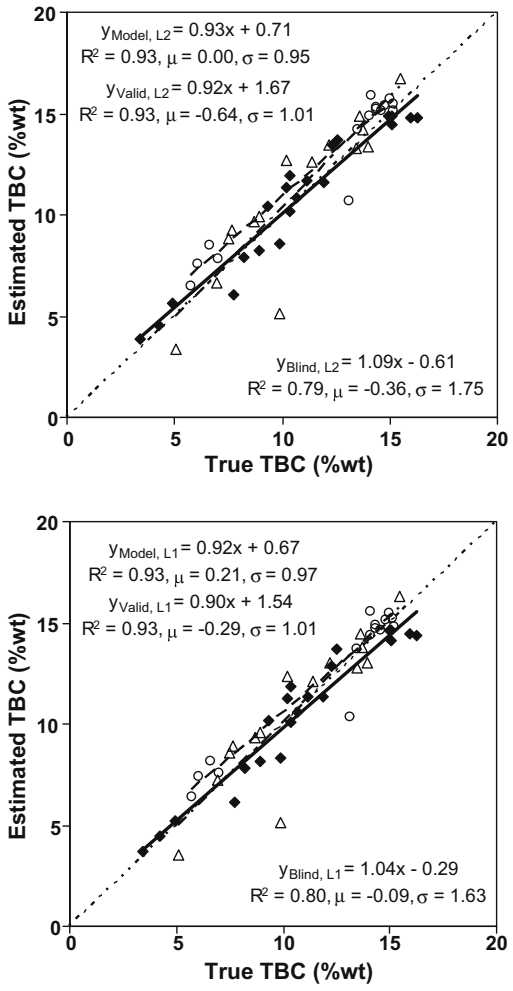

Fig. 7. Estimated TBC derived from L2 norm model (upper panel, Eq. (4a)) and L1 norm model (lower panel, Eq. (4b)) of wavelet parameters fit to reflectance spectra. The ideal solution would lie along the diagonal: Model – diamonds, solid line, Validation – open circles, short-dash line, Blind – triangles, long-short-dash line.

4.3. Robustness of Gaussian and wavelet models

The experiment was deliberately designed to test whether the models could be generalized to data taken from different mines. A cursory examination of Figs. 6 and 7 shows that the wavelet solution is superior in this regard. In Fig. 6 the Gaussian model works well in estimating TBC for the model and validation suite but there is obviously an offset in the estimated blind TBC. This offset is significant for both the L1 and L2 optimizations, though lower for L1 (Table 3). In the case of the wavelet modeling the general trend of increased dispersion in the sequence model-validation-blind is again seen, as was the case for the Gaussian models. However, the estimate of the blind TBC is much better.

The reason for this may be due to the use of the 3.725 μm feature in the wavelet models which may well account for different processes than the shorter wavelength features. The effect, however, may also be due to differences in how the two methods handle the effect of differing noise characteristics in the data suites. Fig. 8 is an example of the original input spectra for three samples, one drawn from the model, validation and blind data suites, with similar TBC values. Both the model and validation spectra exhibit similar noise characteristics while the blind spectrum is considerably noisier as seen in the tail of the spectrum from ~2.5 μm to the end. The smoothing which Peakfit performs as part of the fitting process was unable to overcome the additional noise in the blind test spectra. Thus, the increased noise in the blind suite will affect the 2.532 μm feature used in the Gaussian models. In addition, this increased noise will affect the Gaussian fitting of the baseline, i.e., the overall curvature of the spectrum, and thus the strength of the

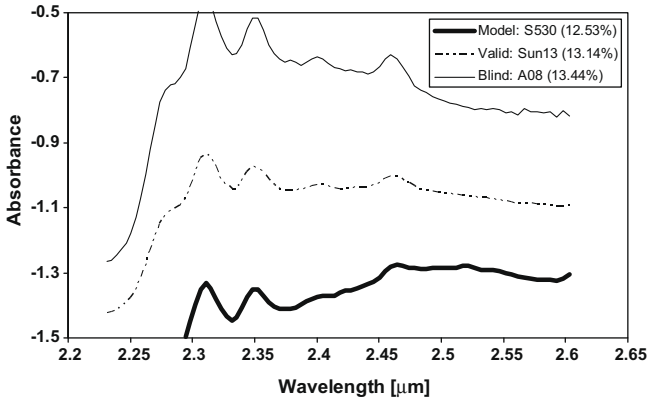

Fig. 8. Details of spectral region used in Gaussian modeling for samples drawn from the model, validation and blind data suites. The true TBC values are shown in parenthesis. Note the marked elevation in the noise level for the blind suite, particularly in the region of the 2.532 μm Gaussian feature. Baseline fitting in the Gaussian model will be sensitive to this variation.

Gaussian features. Both of these effects are intimately mixed together in the Gaussian modeling though only the strength of two Gaussian features are included in the Gaussian model.

Wavelet analysis on the other hand, decouples the influence of noise on the baseline fitting by considering their influence at different scales; noise as the highest frequency component (smallest scale) and the baseline as the lowest frequency component (largest scale). The bitumen features may then be successfully isolated at intermediate scales. The ability of the wavelet model to correctly handle the increased noise in the blind spectra, as well as the general variability of the noise albeit at a lower level in the model and validation spectra, while simultaneously modeling the baseline and bitumen features suggests that it is a robust technique for estimating TBC content in the reflectance spectra of Athabasca oil sands. The use of the L1 optimization will further enhance this model provided the distribution of the noise in the spectra does not change.

5. Conclusion

Modeling of the total bitumen content, TBC, in Athabasca oil sands was undertaken on the basis of its hyperspectral reflectance spectra. Two methods, Gaussian fitting and wavelet analyses, were investigated to identify useful bitumen features as well as the best removal of the baseline. The Gaussian model relied explicitly on features at 2.282 and 2.532 μm though these were only two of 10 features simultaneously fit with a quadratic baseline to the range of 2.230–2.603 μm of the spectra. The wavelet model relied on bitumen features selected at 2.274, 2.396 and 3.725 μm that could be isolated from the baseline and noise. Both models yielded similar dispersion in their estimates of TBC (±∼1–2%) while the wavelet model proved to be more robust when applied to the validation and blind data suites. We also considered the effects of using the L2 optimization (classical least-squares) and L1 optimization (minimization of largest outlier) schemes for both models. Both schemes produced similar results for the model suite of samples for TBC but the L1 was superior when applied to the validation and blind data suites. The wavelet model using the L1 optimization appeared to be quite robust producing estimates of TBC (±∼1.0–1.7%).

Acknowledgements

Syncrude, Suncor and the GEOmatics for Informed Decisions (Geoide) Network of Centres of Excellence of Canada contributed financially to the support of this research.

References

[1] Fong N, Ng S, Chung KH, Tu Y, Li Z, Sparks BD, et al. Bitumen recovery from model systems using a warm slurry extraction process: effects of oil sands components and process water chemistry. Fuel 2004;83:1865–80.
[2] Schramm LL, Stasiuk EN. Surfactants in Athabasca oil sands slurry conditioning, floatation recovery, and tailings processes. In: Schramm LL, editor. Surfactants: fundamentals and applications in the petroleum industry. Cambridge University Press; 2000.
[3] Bichard JA. Oil sands composition and behavior (Alberta oil sands technology and research authority. Edmonton; 1987.
[4] Hepler LG, His C. AOSTRA technical handbook on oil sands, bitumen's, and heavy oils (Alberta oil sands technology and research authority). Edmonton; 1989. p. 14 [chapter 2].
[5] Friesen WI. Appl Spectrosc 1996;50(2):1535–40.
[6] Cloutis EA, Gaffey M, Moslow T. Fuel 1995;74(2):874–9.
[7] Michaelian KH, Zhang SL, Hall RH, Bulmer JT. Fourier transform Raman spectroscopy of Syncrude sweet blend distillation fractions: the 200–1800 cm^{-1} region. Spectrochim Acta A Mol Biomol Spectrosc 2003;59(5): 895–903.
[8] Shaw RC, Kratochvil B. Anal Chem 1990;62:167–74.
[9] Dougan PD. Near-infrared reflectance analysis: its potential application in oil sand processing. AOSTRA J Res 1989;5:203.
[10] Thompson GR. Canadian Patent No. 1 139 702; 1983.
[11] Cloutis EA. Science 1989;245:165–8.
[12] Musser BJ, Kilpatrick PK. Energ Fuel 1998;12:715–25.
[13] Frakman Z, Ignasiak TM, Lown EM, Strausz OP. Energ Fuel 1990;4:263–70.
[14] Bukka K, Hanson FV, Miller JD, Obladi AG. Energ Fuel 1992;6:160–5.
[15] Calemma V, Iwanski P, Nali M, Scotti R, Montanari L. Energ Fuel 1995;9:225–30.
[16] Wilt BK, Welch WT, Rankin JG. Energ Fuel 1998;12:1008–12.
[17] Rose HR, Smith DR, Vassallo AM. Energ Fuel 1998;12:682–8.
[18] Permanyer A, Douifi L, Dupuy N, Lahcini A, Kister J. Fuel 2005;84:159–68.
[19] Torrence C, Compo GP. Bull Amer Meteor Soc 1998;79:61–78.
[20] Tan HW, Brown SD. J Chemometr 2002;16:228–40.
[21] Hunt GR, Salisbury JW, Lenhoff CJ. Mod Geol 1972;3:121–32.
[22] Clark RN, King TVV, Klejwa M, Swayze GA, Vergo N. J Geophys Res 1990;89:6329–40.
[23] Gupta RP. Remote sensing geology. New York: Springer-Verlag; 1990.
[24] Hunt GR, Salisbury JW. Mod Geol 1971;2:23–30.
[25] Hunt GR, Salisbury JW, Lenhoff CJ. Mod Geol 1974;5:15–22.
[26] Mustard JF. Am Mineral 1992;77:345–58.
[27] Hunt GR, Salisbury JW, Lenhoff CJ. Mod Geol 1973;4:237–44.
[28] Ross HP, Adler JEM, Hunt GR. Icarus 1969;11:46–54.
[29] Muraki S. IEEE Trans Vis Comput Graph 1995;1:109–16.
[30] Burns R. Mineralogical applications of crystal field theory. 2nd ed. Cambridge: Cambridge University Press; 1993.
[31] Salisbury JW, Walter LS, Vergo N, D'Aria DM. Infrared (2.1–25 mm) spectra of minerals. Baltimore: The John Hopkins University Press; 1991.
[32] van der Marel HW, Krohmer P. Contrib Mineral Petr 1969;22:73–82.
[33] Hunt GR. Geophysics 1977;42:501–13.
[34] Mendenhall W, Scheaffer RL, Wackerly DD. Mathematical statistics with applications. 2nd ed. Boston, MA: Duxbury Press; 1981.
[35] Moersch JE, Christensen PR. J Geophys Res 1995;100:7465–77.
[36] Press WW, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in Fortran: the art of scientific computing. 2nd ed. New York: Cambridge University Press; 1994.