

# Estimation of mammographic density on an interval scale by transillumination breast spectroscopy

## Kristina M. Blackmore

Mount Sinai Hospital  
Samuel Lunenfeld Research Institute  
60 Murray Street  
Toronto, Ontario M5T 3L9  
Canada

## Samantha Dick

Princess Margaret Hospital, UHN  
Ontario Cancer Institute  
Division Biophysics and Bioimaging  
610 University Avenue, Rm. 8-303  
Toronto, Ontario M5G 2M9  
Canada

## Julia Knight

Mount Sinai Hospital  
Samuel Lunenfeld Research Institute  
60 Murray Street  
Toronto, Ontario M5T 3L9  
Canada  
and  
University of Toronto  
Dalla Lana School of Public Health  
155 College Street  
Toronto, Ontario M5T 3M7  
Canada

## Lothar Lilje

Princess Margaret Hospital, UHN  
Ontario Cancer Institute  
Division Biophysics and Bioimaging  
and  
University of Toronto  
Department of Medical Biophysics  
610 University Avenue  
Toronto, Ontario M5G 2M9  
Canada

## 1 Introduction

Emerging research into certain disease risk factors has led to an increase in risk awareness among the medical community and general population.<sup>1-3</sup> Consequently, there is a growing desire to use risk factors to monitor and maintain health and possibly delay or prevent future disease. Current examples of commonly monitored risk factors include blood pressure, cholesterol, weight or body mass index (BMI) and prostate specific antigen (PSA), to name but a few.<sup>4-7</sup> However, to be effective in monitoring health, these risk factors should be attainable noninvasively, or through sampling of body fluids only, and they should also be measurable at a frequency enabling a quantitative assessment of change due to environ-

**Abstract.** Transillumination breast spectroscopy (TiBS) uses nonionizing optical radiation to gain information about breast tissue morphological and structural properties. TiBS spectra are obtained from 232 women and compared to mammographic density (MD) quantified using Cumulus. The ability of TiBS to estimate MD is assessed using partial least-squares (PLS) regression methods, which requires TiBS spectra as input ( $X$ ) and Cumulus MD as target ( $Y$ ) data. Multiple PLS models are considered to determine the optimal processing technique(s) for the input ( $X$ ) and target ( $Y$ ) data. For each model, the association between TiBS estimated MD ( $\hat{Y}$ ) and Cumulus MD ( $Y$ ) is established using Spearman's rank correlation and linear regression analysis. The model that best estimates MD has the fewest assumptions regarding target ( $Y$ ) and spectral ( $X$ ) processing. The Spearman's correlation coefficient between predicted MD and Cumulus MD for this model is 0.88, with a regression slope ( $\beta$ ) of 0.93 (95% CI 0.83–1.02) and an  $R^2$  of 0.78. The approximation of individual MD was within 10% of Cumulus MD for the majority of women (80%), without stratification on age, body mass index (BMI), and menopausal status. TiBS provides an alternative to mammography assessed MD enabling frequent and earlier use of MD as a risk marker in preventive oncology. © 2008 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3041498]

Keywords: mammographic density; breast spectroscopy; partial least squares; risk assessment; breast cancer.

Paper 08298R received Aug. 27, 2008; revised manuscript received Oct. 17, 2008; accepted for publication Oct. 17, 2008; published online Dec. 16, 2008.

mental impact, preventive interventions, and/or aging.

In the field of preventive oncology, there is a shift toward the development of quantitative models for risk assessment.<sup>8</sup> With respect to breast cancer, numerous studies have demonstrated that mammographic density (MD), obtained from standard x-ray mammography, is a strong independent risk factor for the disease.<sup>9-19</sup> As a result, MD is often used in epidemiological and pharmaceutical studies as a surrogate marker for breast cancer risk when measurements are executed on an interval scale, using threshold modeling or similar means of quantification.<sup>12,20,21</sup> However, with the exception of clinical trials, MD is generally not reported on an interval scale to the physician and/or to the patient as such an analysis would require a trained individual to extract this information from analogue or digital mammograms. Instead, radiologists commonly only classify MD as low, medium, or high, rendering it

Address all correspondence to: Lothar Lilje, Department of Medical Biophysics, Ontario Cancer Institute, Princess Margaret Hospital, 610 University Avenue, Rm. 7-416, Toronto, ON, Canada M5G 2M9. Tel: 416-946-4501 x5743; Fax: 416-946-6529; E-mail: llilje@uhnres.utoronto.ca

insufficient for breast health/risk monitoring. Furthermore, because mammography involves exposure to ionizing radiation, and its sensitivity in premenopausal women is reduced, it is not appropriate for young women or women at increased risk (i.e., mutation carriers) and not for repeated assessments, as would be required, for example, in chemoprevention trials.

Transillumination breast spectroscopy (TiBS) is a nonimaging, noninvasive optical technique that provides information about bulk breast tissue properties (light scattering, water, lipid, and hemoglobin content) based on the spectral dependence of a photon's probability to pass through this tissue.<sup>22-27</sup> Recently, we demonstrated the ability of TiBS to differentiate women with >75% MD from those with <75% MD with an associated area under the curve (AUC) of 0.92 using receiver operator curve (ROC) analysis, indicating good sensitivity and specificity of this technique in assessing MD on a categorical scale.<sup>28</sup> In this analysis, we evaluated the ability of TiBS to predict MD on an interval scale, as required for quantification of change, for example, in chemoprevention trials, among 232 pre- and postmenopausal women. MD was obtained from analogue mammograms using a computer-assisted threshold program<sup>12</sup> (i.e., Cumulus) and subsequently used as the target data ( $Y$ ) in the training of a partial least-squares (PLS) regression based model,<sup>29</sup> with the TiBS spectra acting as the input data ( $X$ ).

## 2 Materials and Methods

### 2.1 Study Population

Study participants were recruited between March 1, 2000, and September 30, 2004, from the Marvella Koffler Breast Imaging Centre at Mount Sinai Hospital in Toronto. This study was approved by the Research Ethics Boards of the University of Toronto, Mount Sinai Hospital, and the University Health Network. Inclusion criteria were an analogue standard screening mammogram within approximately 12 months prior to recruitment, but not exhibiting any radiological suspicious lesions ( $N=232$ ). Exclusion criteria included prior fine-needle aspiration, core biopsies, or any other type of breast surgery including breast reduction or augmentation and any type of tattoos on the breast(s). Additionally, women showing left and right asymmetry in MD, based on classification of mammograms by an expert radiologist (>25% difference), were excluded, thus limiting participants to women whose breast tissue retained symmetric tissue aging.

Information concerning participants' age, menopausal status, height, and weight were collected by means of a self-administered questionnaire. Postmenopausal status was defined as having had no menstrual period for at least 12 months. Height and weight were used to calculate BMI defined as weight in kilograms divided by the square of the height in meters.

### 2.2 Quantification of MD from Mammograms

Mammographic breast tissue density (MD) was used as the gold standard to evaluate the potential for TiBS to estimate breast cancer risk. All film mammograms in cranial-caudal view (2 films  $\times$  232 volunteers,  $N=464$ ) were digitized using a Lumisys Digital Scanner (Kodak, Rochester, New York, USA) at 12-bit resolution and a pixel pitch of 260  $\mu\text{m}$ . Digi-

tized mammograms were examined using Cumulus,<sup>12</sup> an interactive density-threshold software. For each image, the trained reader interactively selects two pixel level values using a special outlining tool. The first level selected (*iEDGE*) separates the outer edge of the breast from the image background. The second level (*iDY*) defines the edges of x-ray dense tissue regions, where all pixels within the outlined region of interest are considered to represent mammographic densities. Additionally, the program enables the user to outline the pectoral muscle thus excluding contributions of the x-ray dense muscle from MD. Pixels between the delineated pectoral muscle and the edge threshold (*iEDGE*) represent total breast area, while those above the density threshold (*iDY*) represent dense tissue areas. The percent dense area (MD) is defined by the ratio of dense area(s) to the total breast area, multiplied by 100.

MD measurements on all mammograms were performed by two individual raters (KMB, LL) after being trained in the use of Cumulus by an expert rater (NB). Mammograms were presented in a randomized order over eight different sessions ( $N=58$  mammograms per session). To assess the reliability of MD measurements, a randomly selected subset of mammograms (15%) were interspersed throughout the eight sessions ( $N=72$  or nine films per session). For the trained raters, the reading of MD was repeated three times (reads 1 to 3) for the entire data set with a period of at least 1 month separating each read. The expert rater only read the mammograms once, including repeats. Table 1 displays the breakdown of study volunteers, as well as MD, demographic, reproductive, and anthropometric information.

### 2.3 Optical Setup and Procedure

The instrumentation used to gather transillumination spectra was previously described in detail.<sup>30</sup> A 50-W halogen lamp served as broadband light source. UV, part of the visible spectrum, and mid-IR radiation were eliminated using a cut-on and a heat rejection filter, respectively. A total power of 0.25 W, covering the 550 to 1300-nm bandwidth, was delivered to the skin via a 5-mm light guide. Transmitted light was collected via a 7-mm-diam optical fiber bundle (140 Si/Si fibers, 200  $\mu\text{m}$  core diameter, numerical aperture: 0.36, P & P Optica, Kitchener, Canada), pointed coaxially toward the light source. The interoptode distance was provided by a caliper holding the optodes. The source fiber was placed against the skin on the top surface of the breast with minimal compression. Wavelength-resolved detection was achieved using a spectrophotometer (Kaiser Optical Systems, Inc., Ann Arbor, MI, USA) with holographic transillumination grating (15.7 rules/mm blazed at 850 nm) and a CCD (Photometrics, Tucson, AZ, USA) with a spectral resolution of better than 3 nm from 625 to 1060 nm ( $N=436$  wavelengths). These wavelengths were selected as they include the absorption spectra of the primary tissue chromophores, namely, hemoglobin (deoxy- and oxy-), water, and lipid. Health Canada Investigational New Device Class II approval was obtained.

All measurements were taken in the dark, with the participant seated comfortably in an upright position and the breast resting on the support platform containing the attached caliper. A total of eight measurements in cranial-caudal projection were taken per individual, four per breast (center: midline

**Table 1** MD, demographic, reproductive, and anthropometric information of all participants with analogue mammograms.

	Premenopausal <sup>a</sup> (N=109)	Postmenopausal <sup>b</sup> (N=123)	All Women (N=232)
Total participants	232		
Total mammograms	464		
MD (%) mean (sd)	35.5 (19.7)	23.1 (18.6)	28.9 (20.1)
Age (years) mean (sd)	45.9 (4.1)	55.4 (6.3)	50.9 (7.1)
BMI (kg m <sup>-2</sup> ) mean (sd) <sup>c</sup>	26.3 (6.4)	26.7 (5.1)	26.6 (5.7)

<sup>a</sup>Premenopausal women had their last menstrual period less than 1 month before mammography and TiBS measurements.

<sup>b</sup>Postmenopausal women had their last menstrual period at least 12 months before mammography and TiBS measurements.

<sup>c</sup>Premenopausal N=106; postmenopausal N=121.

close to the pectoral muscle; medial: 2 cm from the inner edge; distal: 2 cm behind the nipple; lateral: 2 cm from the outer edge). With the measurement of four positions on each breast, an ovoid shaped volume of approximately 23 cm<sup>3</sup> for a 5-cm-thick breast is sampled.<sup>30,31</sup> Numerical modeling demonstrated that 70% of the total breast tissue contributed to 98% of the optical signal. Temporal and spatial reproducibility of the optical measurements is good, as previously addressed.<sup>30,31</sup> The entire TiBS procedure takes approximately 15 min.

#### 2.4 Preparation of Spectra for Data Analysis

All spectra were corrected for variations in the wavelength-dependent signal transfer function of the optical system and the thickness of the interrogated tissue, such that all spectra used in further data processing are independent of the instrument and interoptode distance resulting in units of optical density per centimeter (OD cm<sup>-1</sup>) for the spectra. For correction of the signal transfer function, spectra were referenced to a 1-cm-thick ultra-high-density polyurethane transmission standard (Gigahertz Optics, Munich, Germany). Its optical properties (OD ~1.8 to 2.3 over the wavelength range of interest) were measured separately using an integrating sphere diffuse reflectance setup.<sup>30</sup>

#### 2.5 PLS Analysis

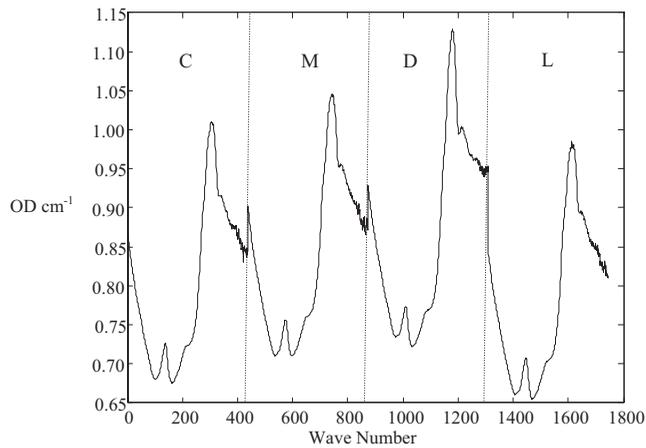
Prior to PLS analysis, the error in MD measurements for each rater and between raters (intra- and interrater error, respectively) was determined as this can directly affect the quality of PLS algorithm training and hence the strength of the attainable correlation between TiBS spectra and MD (see the following). Intraclass correlation coefficients (ICCs) were calculated to assess intrarater repeatability for each read (1 to 3) for the two trained raters (KMB and LL) and for the expert rater (NB, one read only). To assess interrater error, two methods were used. First, the mean absolute MD difference between each trained rater and that of the expert was calculated for each mammogram and each read using a mixed linear model

(PROC MIXED in SAS 9.1). Second, an interclass correlation coefficient was calculated between each trained rater and the expert rater for each read.

Training of the PLS algorithm was executed on a subset of randomly selected spectra ( $N_T=348$  or 75%) and the predictive power of the algorithm tested by including the remaining 25% of spectra in a validation data set ( $N_V=116$ ). The PLS function in MATLAB<sup>29</sup> extracts a common spectrum, called a **b** vector (% cm OD<sup>-1</sup>), from the entire training spectral data set that when multiplied as a cross-product with an individual's spectrum ( $X_i$ ) (OD cm<sup>-1</sup>) produces a scalar or the target, MD ( $Y_i$ ) where  $Y_i=X_i\mathbf{b}$ . The **b** vector identifies those wavelengths, and indirectly also morphological and structural traits, that contribute positively, negatively, or not appreciably to MD ( $Y_i$ ). The 75 versus 25% ratio for training and validation ensures that the training set covers a sufficient range of the variation within the population without over training the system, thereby permitting the PLS algorithm to retain validity on the validation set.<sup>29</sup> Because the PLS algorithm uses absolute intensities, spectra were not mean centered as in our previous analysis.<sup>28</sup>

Since only MD of the entire breast (global assessment) is an established marker for breast cancer risk,<sup>9-19</sup> the four spatially collected spectra per breast were considered as a single vector ( $\mathbf{X}_i$ ) for PLS training with the corresponding Cumulus MD as the target ( $Y_i$ ). Therefore, spectra from all quadrants on each breast (center, medial, distal, lateral) were appended to create a single spectrum to approximate global information (Fig. 1). Hence, for each individual two spectra with corresponding MD (left and right breasts were treated as separate events) were used as the input ( $\mathbf{X}_i$ ) and target ( $Y_i$ ) variables, respectively.

We considered multiple PLS models (Table 2) to determine which spectral processing technique on the input data ( $\mathbf{X}_i$ ) and/or the target data ( $Y_i$ ) best predicts individual global MD ( $\hat{Y}_i$ ). For the target variable ( $Y_i$ ), we examined a model that included only Cumulus MD of the expert rater (NB), as well as models that used the average Cumulus MD of all three raters, to approximate the "gold standard" or best available



**Fig. 1** Example of an appended spectrum from a participant where Cumulus MD ( $Y_i$ )=1.6%. Order of spectra is center (C), medial (M), distal (D), lateral (L). Each spectrum is measured in OD  $\text{cm}^{-1}$  over 436 wavelengths (625 to 1060 nm).

read (see the following results). We also considered models that excluded mammograms (and associated appended spectra) with greater than 10% absolute disagreement in Cumulus MD ( $n=32$ ) between any two raters and models where the Cumulus MD of each rater was not averaged and hence each appended spectrum was used up to three times in the training of our PLS algorithm. In some models, detection shot or CCD readout noise seen in the TiBS spectra ( $X_i$ ) was reduced using either one of two MATLAB defined functions: Boxcar (window  $\delta\lambda=7$  nm) or Savitsky-Golay (Savgol) smoothing and differentiation with a filter width  $\delta\lambda=9$  nm or a third-order polynomial.<sup>29</sup>

Spectral processing and PLS algorithm training was executed in MATLAB version 12.0 and statistical analyses were conducted using SAS version 9.1. For each PLS model, the association between the PLS predicted MD ( $\hat{Y}_i$ ) and the target, Cumulus MD ( $Y_i$ ), was established using Spearman's rank correlation and linear regression analysis [where Cumulus MD ( $Y_i$ )=independent variable ( $x$ ) and PLS MD ( $\hat{Y}_i$ )=dependent variable ( $y$ )] for both the training and validation data sets. For each model, we also examined the residuals [i.e., Cumulus MD ( $Y_i$ ) minus PLS predicted MD ( $\hat{Y}_i$ )] plot-

**Table 2** Results of linear regression analysis [Cumulus MD ( $Y_i$ )=independent variable ( $x$ ) and PLS predicted MD ( $\hat{Y}_i$ )=dependent variable ( $y$ )] and Spearman's rho correlation coefficients for different PLS models for the training and validation data sets.

Model	MD ( $Y_i$ )		Spectra ( $X_i$ ),		N	Data Set	Intercept	95% CI	$\beta$	95% CI	$R^2$	Spearman's rho		
	Averaged	Excluded	Shot-Noise	Minimized										
1	Exert only				348	Training	0.90	-1.70	3.50	0.98	0.90	1.06	0.64	0.80
					116	Validation	1.09	-3.55	5.73	0.98	0.84	1.12	0.62	0.80
2	X				348	Training	0.50	-1.27	2.28	0.99	0.94	1.04	0.80	0.90
					116	Validation	1.15	-2.09	4.39	0.93	0.83	1.02	0.78	0.88
3	X	X			319	Training	0.35	-1.12	1.82	0.99	0.95	1.04	0.86	0.94
					113	Validation	1.21	-2.11	4.53	0.93	0.83	1.02	0.77	0.88
4	X		X (Savgol/Boxcar)		348	Training	0.40	-1.21	2.02	0.99	0.94	1.04	0.83	0.92
					116	Validation	1.87	-1.26	4.99	0.90	0.81	0.99	0.78	0.89
5	X	X	X (Boxcar)		319	Training	0.40	-1.22	2.01	0.99	0.94	1.04	0.84	0.92
					113	Validation	1.77	-1.30	4.84	0.92	0.83	1.00	0.80	0.90
6	X	X	X (Savgol)		319	Training	0.38	-1.22	1.99	0.99	0.94	1.04	0.84	0.92
					113	Validation	1.73	-1.48	4.95	0.92	0.82	1.01	0.78	0.88
7					1044	Training	0.17	-0.79	1.12	1.00	0.97	1.02	0.89	0.95
8		X			957	Training	0.10	-0.79	0.99	1.00	0.97	1.02	0.92	0.96
					339	Validation	8.38	4.65	12.11	0.67	0.57	0.77	0.61	0.79
9		X	X (Savgol/Boxcar)		957	Training	0.27	-1.20	1.75	0.99	0.95	1.04	0.84	0.92
					339	Validation	2.96	-0.26	6.17	0.88	0.79	0.97	0.74	0.84

ted against Cumulus MD ( $Y_i$ ). For statistical testing,  $p$  values  $<0.05$  were considered significant.

### 3 Results

#### 3.1 Intra- and Interrater Error in Cumulus MD ( $Y_i$ )

The intrarater ICC for the expert (NB) was 0.97. For each trained rater (KMB and LL), the intrarater ICC improved with each read and was highest for the third read (0.96 and 0.93, respectively). The mean absolute difference in Cumulus MD for all mammograms between each rater and the expert also decreased with each read (i.e., from read 1 to read 3). The mean absolute difference between rater 1 and the expert and rater 2 and the expert for the third read was 6.0% (95% CI: 5.4%–6.4%) and 5.4% (95% CI: 4.9%–6.4%), respectively, and both were significantly different from zero ( $p < 0.0001$ ). The corresponding interrater ICC for rater 1 was 0.92, while for rater 2 it was 0.93.

#### 3.2 PLS Predicted MD ( $\hat{Y}_i$ )

Because the third read showed the highest intra- and interrater ICC and the smallest absolute difference in Cumulus MD between each trained rater and the expert, the target ( $Y_i$ ) used in all PLS modeling was MD from Cumulus read 3 for the trained raters.

Table 2 displays the results of linear regression analysis and Spearman's rank correlation coefficients for the association between Cumulus MD ( $Y_i$ ) (the independent variable,  $x$ ) and PLS predicted MD ( $\hat{Y}_i$ ) (the dependent variable,  $y$ ) for both the training and validation data sets.

Training the PLS algorithm using Cumulus MD of the expert rater only (model 1) or Cumulus MD of all three raters (not averaged) (models 7 to 9) as the target ( $Y_i$ ), yielded the lowest  $R^2$  values (0.61 to 0.74) and Spearman's rank correlation coefficients ( $\rho$ : 0.80 to 0.84) for the validation data set. Furthermore, for models 7 to 9 the slope of the regression line ( $\beta$ ) in the validation set was significantly less than 1.0, suggesting a bias in model development. Averaging Cumulus MD of all three raters (model 2) improved both the  $R^2$  value and Spearman's  $\rho$  correlation coefficient of the validation set (i.e., versus models 1, 7 to 9) and the 95% CI of the regression slope included 1.0. Conversely, excluding mammograms (and associated spectra) that displayed greater than 10% disagreement between any two raters (model 3) or reducing detection or CCD readout shot noise using either the Boxcar or Savgol MATLAB functions (models 4 to 6) only improved the results for the training data set, not the prediction of MD ( $\hat{Y}_i$ ) (i.e., versus model 2). Because model 2 had the least assumptions regarding spectral and target processing and because it demonstrated both a large  $R^2$  and Spearman's correlation coefficient in the validation data set, it was considered the most parsimonious model. The PLS  $\mathbf{b}$  vector associated with this model is shown in Fig. 2. Figure 3(a) is a scatter plot of PLS predicted MD ( $\hat{Y}_i$ ) versus Cumulus MD ( $Y_i$ ) (model 2, validation set only) and Fig. 3(b) is a plot of the residuals [i.e., Cumulus MD ( $Y_i$ ) minus PLS predicted MD ( $\hat{Y}_i$ )] versus Cumulus MD ( $Y_i$ ). Overall, for the majority of women (80%) the estimation of individual MD ( $\hat{Y}_i$ ) was within 10% of Cumulus MD ( $Y_i$ ), although for the entire data set it ranged between

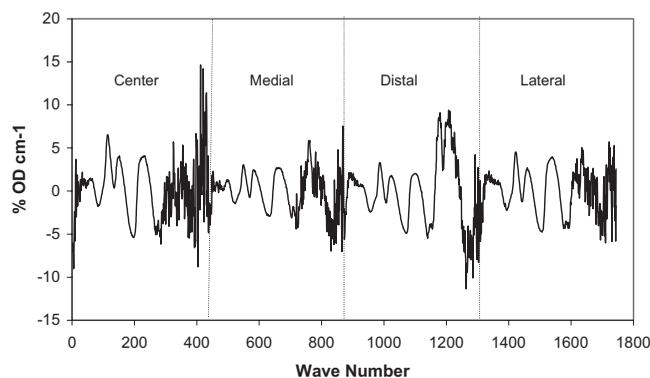


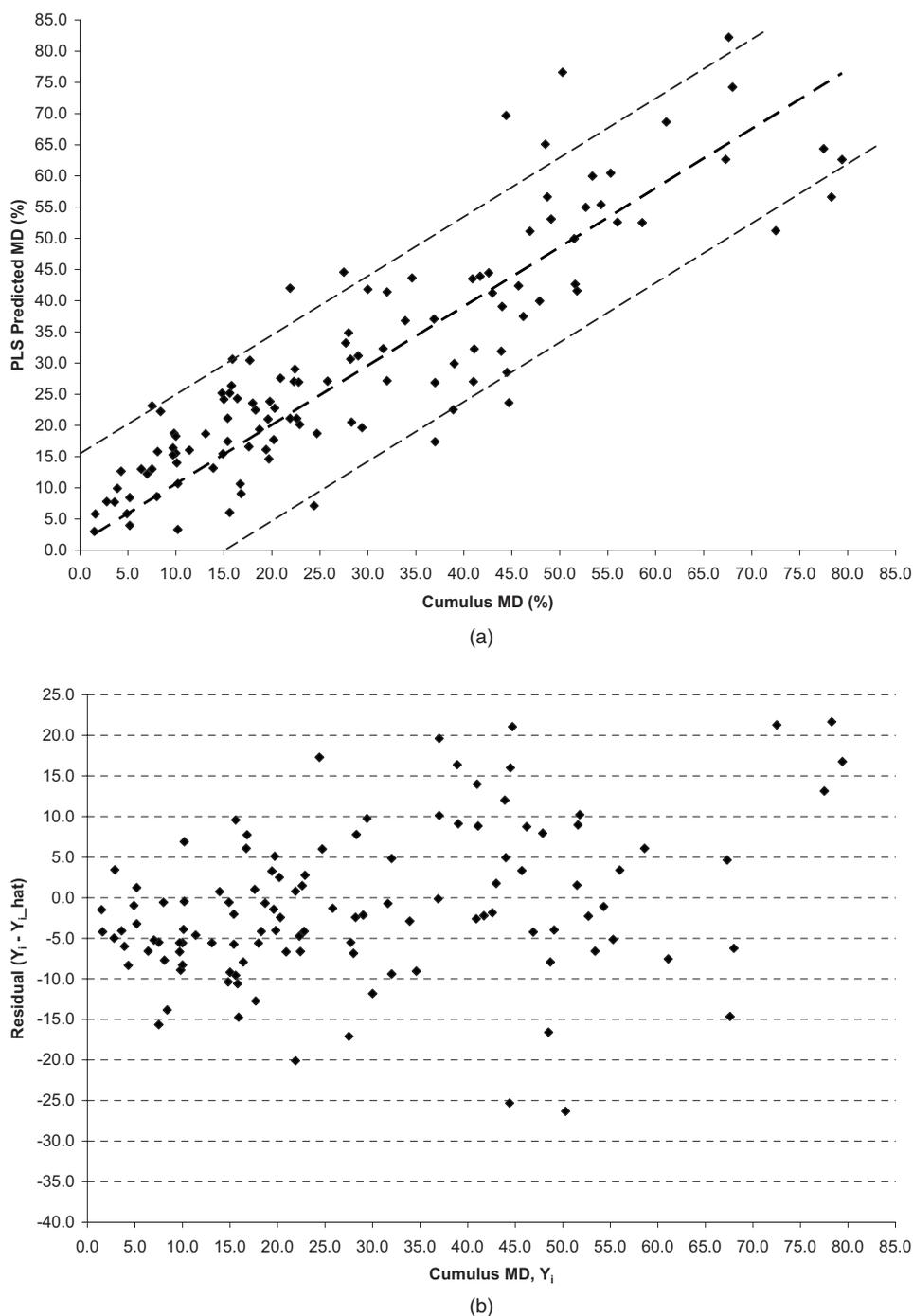
Fig. 2 Resulting PLS  $\mathbf{b}$  vector for model 2 showing relative weights in % OD  $\text{cm}^{-1}$  versus wavelength number for each measurement position.

–26 and +22% [Fig. 3(b)]. The slope of the residuals ( $Y_i - \hat{Y}_i$ ) for model 2 was  $\beta = 0.16\%$  (95% CI: 0.08% to 0.24%), significantly larger than zero ( $p < 0.0001$ ), suggesting that as Cumulus MD ( $Y_i$ ) increases so does the error associated with the prediction of MD ( $Y_i - \hat{Y}_i$ ) [Fig. 3(b)].

### 4 Discussion

The performance of the PLS algorithm is given by the actual strength of the correlation between the target ( $Y_i$ , Cumulus MD) and the input ( $\mathbf{X}_i$ , TiBS spectra) on which it is trained, as well as the accuracy of the target data and the quality of the spectral data. To assess the influence of target accuracy and spectral quality on PLS prediction of MD ( $\hat{Y}_i$ ) various models were evaluated. Among these models, we considered model 2 to be the “best” model because it had the fewest assumptions regarding spectral ( $\mathbf{X}_i$ ) and target ( $Y_i$ ) processing, the  $R^2$  and Spearman  $\rho$  values for the validation set were among the highest (0.78 and 0.88, respectively), and the results for the training and validation data sets were comparable, suggesting no over- or undertraining of the PLS algorithm.

Although we used MD from the final Cumulus read as the target in PLS training, as it demonstrated the highest intra- (0.93 to 0.97) and interrater ICC (0.92 and 0.93) of all three reads, it was not exact. In addition, the absolute difference in Cumulus MD between each trained rater and the expert rater ranged between 5 and 6%, which was significantly different from zero. Employing MD of the expert rater only as the target (model 1) resulted in a less favorable PLS algorithm; both the  $R^2$  values and correlation coefficients for the training data set were small and the prediction of MD in the validation set limited. Conversely, models that employed the average MD of all three raters as the target (models 2 to 6) demonstrated improved results for both the training and validation sets. Hence, there is a potential for bias to be introduced in the estimation of MD when MD from only one or different raters is used as the target in PLS training. Although, the expert rater (NB) has demonstrated a high odds ratio (OR) between MD derived by himself and breast cancer incidence,<sup>12</sup> it is possible that the trained raters (KMB, LL) utilized image attributes (i.e., coarseness, brightness, contrast, noise, etc.) when mea-



**Fig. 3** (a) Plot of PLS predicted MD ( $\hat{Y}_i$ ) versus Cumulus MD ( $Y_i$ ) (model 2, validation set) showing regression line and 95% CI for individual PLS predicted MD ( $\hat{Y}_i$ ); (b) plot of residuals ( $Y_i$  minus  $\hat{Y}_i$ ) versus Cumulus MD ( $Y_i$ ) (model 2, validation set).

suring MD, that were also more common to the TiBS spectra, but that may potentially impact the TiBS-derived MD relationship with breast cancer risk.

For some mammograms, the absolute difference in Cumulus MD between raters exceeded 10%. Although, excluding these mammograms (and associated appended spectra) (models 3, 5, and 6) resulted in a slight improvement in PLS training, it did not improve the estimation of MD in the validation set. This is because removal of these mammograms likely

resulted in exclusion of more cases with higher MD compared to those with lower MD, producing a more homogeneous training set that no longer covered the variance seen in the validation spectra data set. Additionally, despite the fact that participants were selected based on bilateral symmetry of MD by an expert radiologist, left-right symmetry of Cumulus MD was within 10% for only 80% of mammogram pairs (data not shown). Consequently we treated MD from the left and right breasts as separate events, rather than averaging them. This

decision was also based on the fact that Cumulus MD for each breast was derived separately and that TiBS spectra were obtained from each breast. Treating each breast as a separate event would enable future use of TiBS in women with bilateral variations in MD.

The source of the spectral noise at short wavelengths (i.e., wave number 0 to 50; Fig. 1) is due to limited photons traversing several centimeters of tissue, while the noise at long wavelengths (i.e., wave number 400 to 450, Fig. 1) is due to the low quantum efficiency of the CCD detector. Although some of this noise was translated to the PLS  $\mathbf{b}$  vector (Fig. 2), reducing spectral noise using smoothing functions (either Boxcar or Savgol, models 4 to 6), only improved training of the PLS algorithm and not the estimation of MD in the validation data set. Hence, the noisy spectral components did not contribute significantly to the sensitivity/specificity of PLS predicted MD. Instead, it is likely that the noisy components of the  $\mathbf{b}$  vector are associated with the step function at longer wavelengths, which likely resulted from appending individual spectra into a single spectrum (Fig. 1). We did consider training the PLS algorithm on each position individually, thereby avoiding the contribution of additional noise to the  $\mathbf{b}$  vector; however, the prediction of MD did not improve (data not shown). Furthermore, even though an appended spectrum was used, the PLS  $\mathbf{b}$  vector still captured variations in the contribution of absorbers to each measurement position due to the different anatomical structures present in each breast quadrant (Fig. 2). Removing very noisy spectra altogether and/or limiting the wavelength range to omit noisy spectral components may be beneficial and should be explored in future work.

The estimation of MD was poorest when Cumulus MD of each rater contributed independently to the training of the PLS algorithm, but the same appended TiBS spectrum was used each time in training the algorithm (models 7 to 9). Although the 75 versus 25% split for the training and validation sets, respectively, was maintained in these models, overtraining of the algorithm still occurred such that the resulting PLS algorithm no longer retained validity on the validation set (i.e., compared to the other models, the  $R^2$  values and correlation coefficients were much larger for the training set relative to the validation set).

Despite the inherent limitations in the definition of our PLS target ( $Y_i$ ) and the contribution of noise to some spectra ( $X_i$ ) and the resulting  $\mathbf{b}$  vector, the  $R^2$  and Spearman's correlation coefficients for the validation sets did not vary greatly between models ( $R^2$ : 0.61 to 0.80; Spearman's rho: 0.79 to 0.90). This suggests that the limited accuracy of the target and the quality of the spectral data did not contribute substantially to PLS prediction of MD ( $\hat{Y}_i$ ) by TiBS and that estimation of MD might be improved with additional data. Most encouraging is the observation that in the majority of cases the regression slopes ( $\beta$ ) were not significantly different from 1.0 (except in models 7 to 9) indicating that TiBS derived MD does not demonstrate a significant bias. However, even our optimal PLS algorithm (model 2) tended to underestimate MD in women with MD >45%, as suggested by our residual analysis. This is likely due to the fact that only a limited number of women with high MD were available on which to train our PLS algorithm; this should be considered a priority in future studies.

## 5 Conclusions

We showed that TiBS can estimate MD on an interval scale within 10% of Cumulus measured MD in the majority of women without stratification on age, BMI, and menopausal status. The limitations in the generation of a gold standard (i.e., Cumulus MD), which was used as the target for PLS training, does not appear to restrict the overall strength of the attainable correlation between Cumulus MD and PLS predicted MD, possibly indicating that a similar association ( $R^2 \geq 0.90$ ) can be demonstrated in larger multicenter studies. Hence, TiBS has the potential to become a noninvasive, nonionizing-radiation-based method to determine MD without the requirement for highly trained individuals, such as a radiologist and/or expert raters of mammograms. Compared to mammography, it can be applied at higher sampling intervals and can potentially be used to detect changes in MD and possibly the rate of this change. This latter application of TiBS would be important for monitoring high-risk populations (i.e., mutation carriers) and/or women enrolled in chemoprevention trials. While TiBS can determine MD values, its ability to demonstrate changes relative to actual breast cancer risk must still be demonstrated, which would require a prospective longitudinal study.

### Acknowledgments

The authors wish to express their gratitude to Dr. Norman Boyd (NB) for his help in the use of the Cumulus program and for his contribution to the development of the ideas that were the basis of this analysis. The authors also thank Dr. Martin Yaffe for the use of the Cumulus program and Dr. Gina Lockwood for discussion relating to the statistical analysis of the data. The authors express their sincere thanks to all study participants for their time and effort. The study was funded in part by the Susan Komen Foundation (BCTR0402530) and the Canadian Breast Cancer Research Alliance (017467).

### References

1. P. Greenwald, "The future of cancer prevention," *Semin Oncol. Nurs.* **21**(4), 296-298 (2005).
2. B. K. Rimer and J. M. Gierisch, "Public education and cancer control," *Semin Oncol. Nurs.* **21**(4), 286-295 (2005).
3. K. Viswanath, "Science and society: the communications revolution and cancer control," *Nat. Rev. Cancer* **5**(10), 828-835 (2005).
4. M. G. Myers, "Blood pressure measurement and the guidelines: a proposed new algorithm for the diagnosis of hypertension," *Blood Press. Monit.* **9**(6), 283-286 (2004).
5. R. G. Bowden, P. M. Kingery, and D. M. Rust, "Assessing risk using different cholesterol-screening methods," *Public Health* **118**(3), 225-229 (2004).
6. M. de Onis, "The use of anthropometry in the prevention of childhood overweight and obesity," *Int. J. Obes.* **28**(S), S81-S85 (2004).
7. A. Caplan and A. Kratz, "Prostate-specific antigen and the early diagnosis of prostate cancer," *Am. J. Clin. Pathol.* **117**(S), S104-S108 (2002).
8. G. H. Sakorafas, E. Krespis, and G. Pavlakakis, "Risk estimation for breast cancer development; a clinical perspective," *Surg. Oncol.* **10**(4), 183-192 (2002).
9. N. F. Boyd, B. O'Sullivan, J. E. Campbell, E. Fishell, I. Simor, G. Cooke, and T. Germanson, "Mammographic signs as risk factors for breast cancer," *Br. J. Cancer* **45**(2), 185-193 (1982).
10. J. N. Wolfe, A. F. Saftlas, and M. Salane, "Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study," *AJR, Am. J. Roentgenol.* **148**(6), 1087-1092 (1987).
11. A. M. Oza and N. F. Boyd, "Mammographic parenchymal patterns: a

- marker of breast cancer risk," *Epidemiol. Rev.* **15**(1), 196–208 (1993).
12. N. F. Boyd, J. W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B. Miller, G. A. Lockwood, D. L. Tritchler, and M. J. Yaffe, "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study," *J. Natl. Cancer Inst.* **87**(9), 670–675 (1995).
  13. N. F. Boyd, G. A. Lockwood, J. W. Byng, D. L. Tritchler, and M. J. Yaffe, "Mammographic densities and breast cancer risk," *Cancer Epidemiol. Biomarkers Prev.* **7**(12), 1133–1144 (1998).
  14. M. J. Yaffe, N. F. Boyd, J. W. Byng, R. A. Jong, E. Fishell, G. A. Lockwood, L. E. Little, and D. L. Tritchler, "Breast cancer risk and measured mammographic density," *Eur. J. Cancer Prev.* **7**(S1), S47–55 (1998).
  15. P. M. Vacek and B. M. Geller, "A prospective study of breast cancer risk using routine mammographic breast density measurements," *Cancer Epidemiol. Biomarkers Prev.* **13**(5), 715–722 (2004).
  16. J. A. Harvey and V. E. Bovbjerg, "Quantitative assessment of mammographic breast density: relationship with breast cancer risk," *Radiology* **230**(1), 29–41 (2004).
  17. V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," *Cancer Epidemiol. Biomarkers Prev.* **15**(6), 1159–1169 (2006).
  18. N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, S. Minkin, and M. J. Yaffe, "Mammographic density and the risk and detection of breast cancer," *N. Engl. J. Med.* **356**(3), 227–236 (2007).
  19. M. Yaffe and N. Boyd, "Mammographic breast density and cancer risk: the radiological view," *Gynecol. Endocrinol.* **21**(S1), 6–11 (2005).
  20. N. F. Boyd, "Effects at two years of a low-fat, high-carbohydrate diet on radiologic features of the breast: results from a randomized trial. Canadian Diet and Breast Cancer Prevention Study Group," *J. Natl. Cancer Inst.* **89**(7), 488–496 (1997).
  21. J. A. Knight, L. J. Martin, C. V. Greenberg, G. A. Lockwood, J. W. Byng, M. J. Yaffe, D. L. Tritchler, and N. F. Boyd, "Macronutrient intake and change in mammographic density at menopause: results from a randomized trial," *Cancer Epidemiol. Biomarkers Prev.* **8**(2), 123–128 (1999).
  22. H. Key, E. R. Davies, and P. C. Jackson, "Optical attenuation characteristics of breast tissues at visible and near-infrared wavelengths," *Phys. Med. Biol.* **36**(5), 579–590 (1991).
  23. B. J. Tromberg, O. Coquoz, J. B. Fishkin, T. Pham, E. R. Anderson, J. Butler, M. Cahn, J. D. Gross, V. Venugopalan, and D. Pham, "Non-invasive measurements of breast tissue optical properties using frequency-domain photon migration," *Philos. Trans. R. Soc. London, Ser. B* **352**(1354), 661–668 (1997).
  24. V. Quaresima, S. J. Matcher, and M. Ferrari, "Identification and quantification of intrinsic optical contrast for near-infrared mammography," *Photomed. Photobiol.* **67**(1), 4–14 (1998).
  25. A. E. Cerussi, A. J. Berger, F. Bevilacqua, N. Shah, D. Jakubowski, J. Butler, R. F. Holcombe, and B. J. Tromberg, "Sources of absorption and scattering contrast for near-infrared optical mammography," *Acad. Radiol.* **8**(3), 211–218 (2001).
  26. N. Shah, A. Cerussi, C. Eker, J. Espinoza, J. Butler, J. Fishkin, R. Hornung, and B. Tromberg, "Noninvasive functional optical spectroscopy of human breast tissue," *Proc. Natl. Acad. Sci. U.S.A.* **98**(8), 4420–4425 (2001).
  27. T. Durduran, R. Choe, J. P. Culver, L. Zubkov, M. J. Holboke, J. Giammarco, B. Chance, and A. G. Yodh, "Bulk optical properties of healthy female breast tissue," *Phys. Med. Biol.* **47**(16), 2847–2861 (2002).
  28. K. M. Blackmore, J. A. Knight, R. Jong, and L. Lilge, "Assessing breast tissue density by transillumination breast spectroscopy (TIBS): an intermediate indicator of cancer risk," *Br. J. Radiol.* **80**(955), 545–556 (2007).
  29. B. M. Wise, "PLS Tool Box Tutorial: MatLab Version 6," Eigenvektor Research Inc., Seattle, WA (2000).
  30. M. Simick, "Near infrared transillumination spectroscopy of breast tissue for correlation with mammographic density," M.Sc. thesis in Department of Medical Biophysics, University of Toronto, Toronto (2002).
  31. M. K. Simick, R. Jong, B. Wilson, and L. Lilge, "Non-ionizing near-infrared radiation transillumination spectroscopy for breast tissue density and assessment of breast cancer risk," *J. Biomed. Opt.* **9**(4), 794–803 (2004).