

A statistical model for removing inter-device differences in spectroscopy

Lu Wang,^{1,6} Jong Soo Lee,² Pierre Lane,³ E. Neely Atkinson,¹ Andres Zuluaga,⁴
Michele Follen,⁵ Calum MacAulay,³ and Dennis D. Cox^{1*}

¹Department of Statistics, Rice University, 6100 Main St., Houston, TX 77005 USA

²Department of Applied Economics and Statistics, University of Delaware, 206 Townsend Hall, Newark, DE 19716 USA

³Department of Cancer Imaging, British Columbia Cancer Research Centre, 600 West 10th Ave., Vancouver, British Columbia V5Z4E6 Canada

⁴Clearview App, Inc. 3900 Essex Ln Ste 250 Houston, Texas 77027-5181 USA

⁵Department of Obstetrics and Gynecology, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso, TX 79905 USA

⁶lu.wang@rice.edu

*dcox@rice.edu

Abstract: We are investigating spectroscopic devices designed to make *in vivo* cervical tissue measurements to detect pre-cancerous and cancerous lesions. All devices have the same design and ideally should record identical measurements. However, we observed consistent differences among them. An experiment was designed to study the sources of variation in the measurements recorded. Here we present a log additive statistical model that incorporates the sources of variability we identified. Based on this model, we estimated correction factors from the experimental data needed to eliminate the inter-device variability and other sources of variation. These correction factors are intended to improve the accuracy and repeatability of such devices when making future measurements on patient tissue.

©2014 Optical Society of America

OCIS codes: (120.4800) Optical standards and testing; (170.4440) ObGyn; (170.6510) Spectroscopy, tissue diagnostics; (030.6600) Statistical optics.

References and links

1. J. A. Freeberg, D. M. Serachitopol, N. McKinnon, R. Price, E. N. Atkinson, D. D. Cox, C. MacAulay, R. Richards-Kortum, M. Follen, and B. Pikkula, "Fluorescence and reflectance device variability throughout the progression of a phase II clinical trial to detect and screen for cervical neoplasia using a fiber optic probe," *J. Biomed. Opt.* **12**(3), 034015 (2007).
2. J. M. Yamal, G. A. Zewdie, D. D. Cox, E. N. Atkinson, S. B. Cantor, C. MacAulay, K. Davies, I. Adewole, T. P. Buys, and M. Follen, "Accuracy of optical spectroscopy for the detection of cervical intraepithelial neoplasia without colposcopic tissue information; a step toward automation for low resource settings," *J. Biomed. Opt.* **17**(4), 047002 (2012).
3. B. Yu, H. L. Fu, and N. Ramanujam, "Instrument independent diffuse reflectance spectroscopy," *J. Biomed. Opt.* **16**(1), 011010 (2011).
4. B. Yu, H. Fu, T. Bydlon, J. E. Bender, and N. Ramanujam, "Diffuse reflectance spectroscopy with a self-calibrating fiber optic probe," *Opt. Lett.* **33**(16), 1783–1785 (2008).
5. J. L. Devore, *Probability and Statistics for Engineering and the Sciences* (Cengage Learning, 2011, 8th ed.), Chap.11.
6. B. G. Lipták, *Instrument Engineers' Handbook, Volume Two: Process Control and Optimization*. (CRC Press, 2010.)

1. Introduction

The goal of our group is to develop technology to diagnose pre-cancerous and cancerous cervical lesions with *in vivo* spectroscopic measurements. This technology utilizes reflectance and fluorescence spectroscopy to assess tissue emission intensity at certain excitation, emission and remittance wavelengths [1]. We call the current generation of these devices

Multispectral Digital Colposcope 3 (MDC3). Details may be found in [2–4]. One issue with the MDC3 is that the shape and amplitude of the recorded spectra are different across devices for the same target. Therefore, device specific “correction factors” are needed to make the measurements consistent across devices.

For purposes of calibration and quality assurance, we make measurements on a daily basis from square Teflon tiles doped with two types of inorganic fluorophores. There are 4 such tiles that we refer to as “standards”. We conducted an experiment in which we repeatedly measured the reflectance and fluorescence spectra from these standards. We present a model to estimate correction factors for each device and standard. This statistical model provides a unified way to incorporate the different sources of variation and to assess the accuracy of the estimates. The correction factors will be used to remove the systematic differences between devices.

2. Materials and methods

2.1 Device description

The MDC3 devices in this study consist of a seven-fiber optic cable. Light from the sample is received by a single central fiber surrounded by a ring of six illumination fibers. This simple configuration was adopted because in previous studies we found that small differences in probe assembly with complex designs resulted in differences that could not be corrected for. The light source is a Xenon arc lamp inside the device (Fig. 1). The output of the lamp is sent to a power meter, which measures the total energy of illumination light delivered to the sample, which is used to normalize the measurements. For the fluorescence measurements, optical filters are used to produce narrow band excitation light and to restrict the range of emission light measured by the spectrometer. There are three different fluorescence filter sets in the device. Additionally, for the white light reflectance measurements, broadband filters are used. Table 1 below shows the details of the four different filter sets. A single measurement produces spectra from all four filter sets by quickly rotating filter wheels as well as opening and closing shutters. A measurement from a standard of all four spectra takes about 1 second.

Table 1. Wavelength ranges for the excitation and detection light.

	Description	Excitation range (nm)	Detection range (nm)
Filter set 1	White reflectance	No filter	410 - 700
Filter set 2	Blue fluorescence	455 - 465	495 - 550
Filter set 3	Violet fluorescence	410 - 420	440 - 560
Filter set 4	UV fluorescence	340 - 360	420 - 580

We have four MDC3 devices and four measurement standards. The four MDC devices were manufactured in a single production run. The standards were ordered in two batches to identical specifications. In clinical use, we measure the standards prior to tissue measurement for quality assurance and standardization. Ideally, the devices and standards would be identical, but there are systematic differences.

2.2 Design of experiment

We conducted an experiment to assess repeatability of the standards measurements. In order to assess differences between both devices and standards, we measured each standard with each device. To account for time trends, we made repeated measurements over the day in three sessions: morning, mid-day and evening. The standard is kept in a plastic receptacle that has an opening into which the probe is inserted to make a measurement (Fig. 1). Because of possible effects from variable contact of the probe with the standard, we made five sets of five measurements for each session. For a given set the probe was kept in the receptacle to make five measurements; between sets, the probe was removed and reinserted. Therefore, there are

4 devices \times 4 standards \times 3 time sessions \times 5 sets \times of 5 measurements = 1200 spectra measured for each illumination detection filter set.



Fig. 1. (a) MDC3 device. (b) The spectroscopic probe inserted into the standard receptacle.

Data processing includes background subtraction and wavelength calibration (from pixel location in the spectrometer to wavelength in nanometers). To correct the measurements for variation in the intensity of illumination power all the measurements were to be divided by the power meter measurement. Due to missing values and quality control issues for some of the optical power measurements, 82 measurements were removed, leaving 1118 measurements.

2.3 Log additive model

We assume a linear response for the device that is wavelength dependent with no additive terms (it is assumed that offsets are removed by background subtraction). Illumination amplitude fluctuations are multiplicative factors that are constant across emission wavelengths. Taking logarithms turns these multiplicative relations into additive effects. Taking logarithms also reduces dynamic range and makes the variances more uniform. We posit the following log additive model for measurements from each filter set:

$$y_{ijk}(\lambda) = \log Y_{ijk}(\lambda) = \mu(\lambda) + ICFD_i(\lambda) + ICFS_j(\lambda) + IAC_{ijk} + \varepsilon_{ijk}(\lambda) \quad (1)$$

where

$Y_{ijk}(\lambda)$ = measured intensity at wavelength λ .

$i, j = 1, 2, 3, 4$ for 4 devices and 4 standards respectively.

$k = 1, \dots, N_{ij}$, indexes for measurement number for device i and standard j .

λ = wavelength in nanometers.

$\mu(\lambda)$ = grand mean for log of all measurements in a single filter set.

$ICFD_i(\lambda)$ = log of correction factor based on device i .

$ICFS_j(\lambda)$ = log of correction factor based on standard j .

IAC_{ijk} = log of amplitude correction factor.

$\varepsilon_{ijk}(\lambda)$ = unexplained variation.

There are no terms in the model for each session or the set within each session. Our results in section 3 suggested that the time effects are subsumed in the IAC_{ijk} terms. We will analyze

those to assess if there are any effects from session and set. Using this log additive model, it is straight forward to estimate the correction factors and to apply standard statistical methods to assess the accuracy of estimates.

2.4 Model estimation

To calculate the estimates of each term in Eq. (1) we use analysis of variance (ANOVA) [5] software. The first step is to fit the wavelength dependent correction factors ICFD and ICFS. Let $\hat{\mu}(\lambda)$ be the average of the log of all the measurement at each wavelength, then we have

$$IC\hat{F}D_i(\lambda) = \left(\frac{\sum_{j=1}^4 \sum_{k=1}^{N_{ij}} y_{ijk}(\lambda)}{\sum_{j=1}^4 N_{ij}} - \hat{\mu}(\lambda) \right), \quad IC\hat{F}S_j(\lambda) = \left(\frac{\sum_{i=1}^4 \sum_{k=1}^{N_{ij}} y_{ijk}(\lambda)}{\sum_{i=1}^4 N_{ij}} - \hat{\mu}(\lambda) \right) \quad (2)$$

with N_{ij} = number of measurements for device i and standard j . At each wavelength λ , put

$$\hat{\varepsilon}_{ijk}^{(1)}(\lambda) = y_{ijk}(\lambda) - \hat{\mu}(\lambda) - IC\hat{F}D_i(\lambda) - IC\hat{F}S_j(\lambda) \quad (3)$$

The log of amplitude correction is given by $IA\hat{C}_{ijk} = n_\lambda^{-1} \sum_\lambda \hat{\varepsilon}_{ijk}^{(1)}(\lambda)$ with n_λ being the number of emission wavelengths. The residual is $\hat{\varepsilon}_{ijk}(\lambda) = \hat{\varepsilon}_{ijk}^{(1)}(\lambda) - IA\hat{C}_{ijk}$. The corrected spectra are then given by

$$\log Y_{ijk}^{CF}(\lambda) = \log Y_{ijk}(\lambda) - IC\hat{F}D_i(\lambda) - IC\hat{F}S_j(\lambda) - IA\hat{C}_{ijk} \quad (4)$$

3. Results

We made repeated measurements as described above. Filter set 1, the white reflectance data, has the most complicated shape as shown in the left plot of Fig. 2. For the fluorescence filter sets, filter set 4 has the widest range of emission wavelengths and encompasses all of the structure observed in filter sets 2 and 3. Therefore, we only present the results of filter sets 1 and 4. Below are the plots of the original measurements after pre-processing:

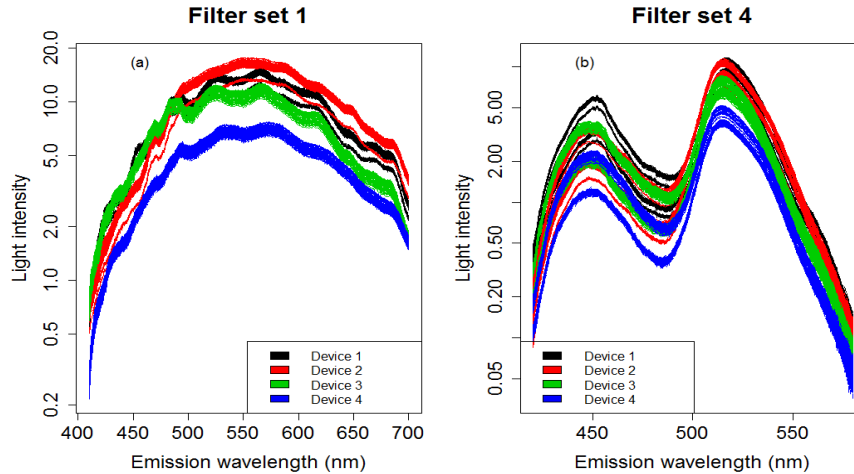


Fig. 2. Plots of the measured spectra color-coded by device.

Note that all plots are in semi-log scale, so the units on the vertical axis are real and not logarithmic. In Fig. 3, the left hand plots show the corrected spectra as given by Eq. (4) and

right plots show the residuals defined above. In the left two plots in Fig. 3, the measurements align very well: we mostly see the blue color for device 4 because it was plotted last.

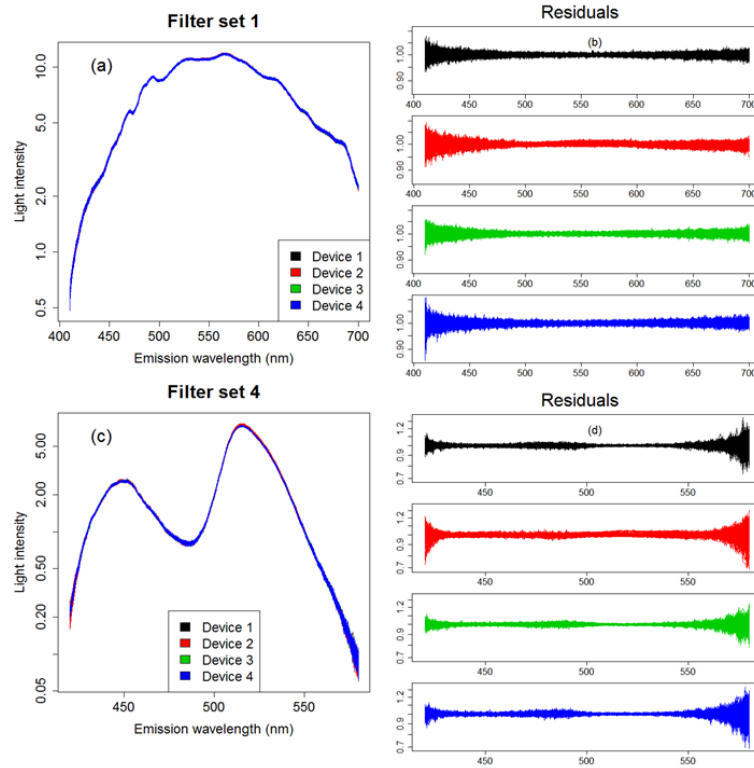


Fig. 3. (a) (b) - light intensities after applying the correction factors (Log10 plot with original y-axis scale); (c) (d) - residuals plotted by devices in same scale.

In the Table 2 below, we show the average of percentage of variance explained by each factor. The effects of device and standard explained most of the variation, which shows the correction factors are effective for removing most of the variation.

Table 2. Percentage of variance explained by each factor (averaged across wavelength)

% of variation	Filter set 1	Filter set 2	Filter set 3	Filter set 4
CFD	96.53%	34.69%	48.64%	70.40%
CFS	0.27%	64.50%	50.63%	27.28%
AC	3.15%	1.23%	1.10%	2.11%
Unexplained	0.07%	0.29%	0.10%	0.41%

Figure 4 shows the estimates of correction factors by device and standard and their corresponding 95% confidence intervals. The 95% confidence intervals are given by:

$$IC\hat{F}D_i(\lambda) \pm 1.96 \sqrt{(1 - w_i)^2 \text{var}(\bar{y}_i) + \sum_{j \neq i} w_j^2 \text{var}(\bar{y}_j)} \quad (5)$$

with

$$\bar{y}_i(\lambda) = \frac{\sum_{j=1}^4 \sum_{k=1}^{N_{ij}} y_{ijk}(\lambda)}{\sum_{j=1}^4 N_{ij}}, \quad w_i = \frac{\sum_{j=1}^4 N_{ij}}{\sum_{i=1}^4 \sum_{j=1}^4 N_{ij}},$$

$$\text{var}(\bar{y}_i) = \frac{\sum_{j=1}^4 \sum_{k=1}^{N_{ij}} \hat{\epsilon}_{ijk}^2(\lambda)}{\sum_{j=1}^4 N_{ij} - 1}.$$
(6)

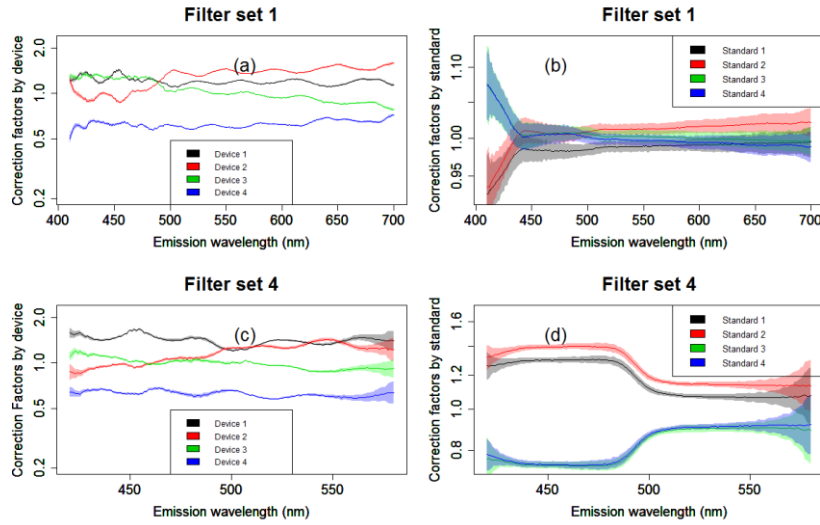


Fig. 4. Correction factors by device ($\exp(ICFD_i(\lambda))$) and standard ($\exp(ICFS_j(\lambda))$), with 95% confidence intervals.

The amplitude correction factors are very similar across all filter sets, but they show random patterns across session of the day and sets of measurements (Fig. 5).

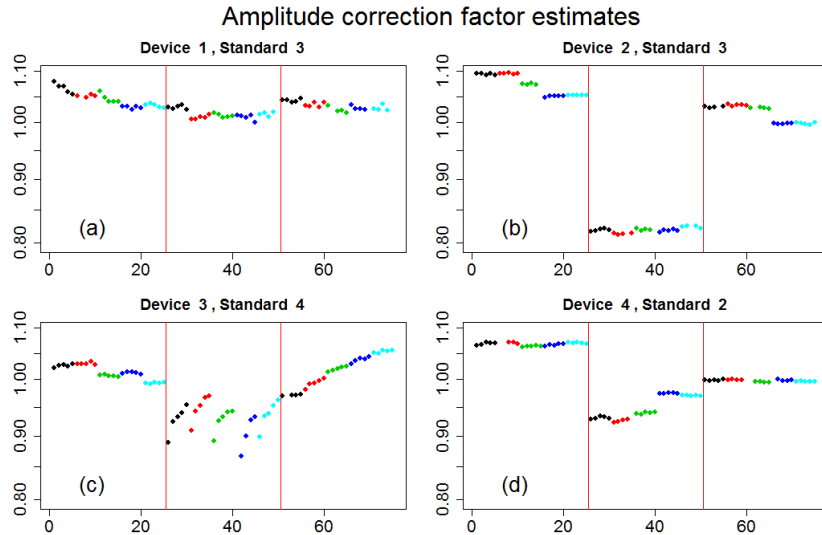


Fig. 5. Amplitude correction factors ($AC_{ijk} = \exp(IC_{ijk})$) estimates for filter set 1. The red lines separate the sessions. The color codes indicate sets of 5 measurements. Missing values are left blank.

In the plot above, we picked some device and standard combinations from filter set 1 to show how the amplitude correction estimates vary. The fluctuations in amplitude suggest that we cannot use these estimates to make any consistent corrections to patient measurements.

4. Discussion and conclusions

4.1 Application of the model

The log additive model we proposed provides a unified approach to estimate the correction factors along with accuracy assessments, i.e. confidence intervals. There are three general components to the model from the effects of the device, the standard and the amplitude fluctuations. The first two can be reliably used to correct the differences between devices and standards. After we applied the correction factor to the raw measurements, the curves align very well.

The plot of the correction factors for device and standard (Fig. 3) both show some overall patterns. Device correction factors oscillate along the emission wavelength for all filter sets and devices and some of them show upward or downward trends. The oscillating pattern of $CFD_i(\lambda)$ is likely due to the slight differences in lamp spectra and the spectral transmission of the components in the optical pathway, particularly the interference coatings on the hot mirror, excitation-emission filters, and the anti-reflection coatings on the projection and collimation lens elements. The correction factor for device 2 ($CFD_2(\lambda)$) has a different trend compared to the other three. Later we found that this was because one of the elements in the excitation projection lens was positioned in a backward position. This was then corrected.

Standard correction factors are quite different across filter sets. Correction factors for filter set 1 are all small in magnitude. Standards 1 and 2, which were ordered in the first batch, apparently have higher level of fluorophores than standards 3 and 4, which were in the later batch, and the relative concentrations of the fluorophores vary across the batches. We believe this variation reflects batch to batch differences that one can expect to occur.

The amplitude correction estimates AC_{ijk} are constant across wavelength and are highly consistent across filter sets. We only show some of them for filter set 1. They have distinct patterns over time. We believe that these are not due to any long term trends in the device (heating, system component expansion etc.), but likely due to fluctuations in the arc lamp output and arc position over time and so there is no way to obtain real corrections for these variations in patient measurements without additional information. Components to obtain this needed additional information are being added to the MDC devices.

One important application is to use these estimates for quality control purposes. Our recommendation is when taking a new standard measurement, take the log10 of the spectra and then subtract out both the device and standard correction factors and log10 of the grand mean, if the result is outside of a certain confidence range, we will classify the measurement as an outlier and recommend a repeat measurement. This confidence range is calculated as the six-sigma [6] interval for the amplitude correction estimates from our experiment for the given device and standard.

During the patient measurement in the clinic, we propose to make one standard measurement right before the patient *in vivo* tissue measurement. For this standard measurement, if we don't reject it based on the above criteria, the patient tissue measurement will be multiplied by 10^{-CFD} to obtain the device corrected spectrum.

4.2 Further investigations

One problem with this model is that the residual variance is not constant across all emission wavelengths. We could use a wavelength dependent weight to refine the model fitting with a two-stage algorithm.

There is also an issue with updating the correction factor estimates. Since we suggest taking a standard measurement right before the patient measurement and we believe the

device performance may change over time, it may give greater accuracy if our estimated correction factors evolve accordingly.

Acknowledgments

The authors gratefully acknowledge the support from NIH-NCI Grant No. P01-CA82710 from the National Cancer Institute. The authors appreciate the suggestions from Dr. Scott B. Cantor which enhanced the clarity of the manuscript.