Development of quality control and instrumentation performance metrics for diffuse optical spectroscopic imaging instruments in the multi-center clinical environment


Event: SPIE BiOS, 2013, San Francisco, California, United States
Development of Quality Control and Instrumentation Performance Metrics for Diffuse Optical Spectroscopic Imaging Instruments in the Multi-Center Clinical Environment


Beckman Laser Institute and Medical Clinic, University of California Irvine, 1002 Health Sciences Road, East Irvine CA 92617

ABSTRACT

Instrument equivalence and quality control are critical elements of multi-center clinical trials. We currently have five identical Diffuse Optical Spectroscopic Imaging (DOSI) instruments enrolled in the American College of Radiology Imaging Network (ACRIN, #6691) trial located at five academic clinical research sites in the US. The goal of the study is to predict the response of breast tumors to neoadjuvant chemotherapy in 60 patients. In order to reliably compare DOSI measurements across different instruments, operators and sites, we must be confident that the data quality is comparable. We require objective and reliable methods for identifying, correcting, and rejecting low quality data. To achieve this goal, we developed and tested an automated quality control algorithm that rejects data points below the instrument noise floor, improves tissue optical property recovery, and outputs a detailed data quality report. Using a new protocol for obtaining dark-noise data, we applied the algorithm to ACRIN patient data and successfully improved the quality of recovered physiological data in some cases.

Keywords: quality control, ACRIN, diffuse optical spectroscopic imaging, frequency domain photon migration, multi-center trial, standardization, instrument testing

1. INTRODUCTION

1.1. Diffuse Optical Spectroscopic Imaging of Breast Tumor Response to Neoadjuvant Chemotherapy

Diffuse Optical Spectroscopic Imaging (DOSI) is a powerful tool for non-invasively measuring the concentrations of oxygenated hemoglobin (ctO2Hb), deoxygenated hemoglobin (ctHHb), water (ctH2O), and lipid (%lipid) in thick tissues.[1] Because of the high sensitivity of near-infrared (NIR, 650-1000nm) absorption spectra to the states of cells, vessels and matrix[2], tumor NIR absorption properties are being investigated as biomarkers that characterize tumor response to cancer therapy. Non-invasive DOSI endpoints based upon quantitative measures of tumor optical/functional properties can be obtained rapidly with no risk or discomfort to patients. The optical biomarkers generated by DOSI and related optical imaging technologies could be used as surrogate markers of pathologic response which is an established indicator of long-term patient survival.[3, 4] Our long-term goal is to provide oncologists with a simple, risk-free bedside tool that helps to inform medical decisions on chemotherapy regimen, duration, and surgical timing, thereby maximizing therapeutic response and minimizing unnecessary toxicity.

Significant preliminary evidence suggests DOSI could meet this important clinical goal. Using a bedside-capable DOSI instrument, we reported the first in-human use[5], multi-patient study[6] and comparison with MRI[7] that suggested optical imaging technologies could monitor tumor response to neoadjuvant (i.e., pre-surgical) chemotherapy. Similar contrast mechanisms have been studied by other groups using related diffuse optical imaging approaches.[8-11] We recently reported in the largest known published clinical study (36 tumors) strong correlations between DOSI-measured optical biomarkers and the degree of pathological response.[12] New optical imaging biomarkers are also being discovered, such as “metabolic flares” (i.e., changes in ctO2Hb 1 day after therapy), baseline oxygen saturation (stO2) and spectral/spatial heterogeneity, that may allow for early prediction of final tumor pathological responses[13-15].
1.2. Quality Control in Multicenter Studies

Although these findings have generated considerable excitement, none of them have been validated in a multi-center environment. To meet this challenge, we are currently engaged in a multicenter clinical trial sponsored by the American College of Radiology Imaging Network (ACRIN, #6691). The primary aim of the ACRIN study is to evaluate whether a quantitative DOSI tissue optical index (TOI), defined as \( \text{ctHHb} \times \text{ctH2O}/(\%\text{lipid}) \), can predict neoadjuvant chemotherapy pathologic complete response (pCR) by the mid-point of the therapy. TOI has been shown to be a high-contrast index that localizes enhanced metabolic events such as tumors in breast tissue.[12, 16] The secondary aims of the ACRIN study are to examine additional quantitative DOSI endpoints at multiple timepoints during neoadjuvant chemotherapy and to correlate DOSI with other standard of care imaging and/or any magnetic resonance imaging (MRI).[12, 14] We currently have 5 identical DOSI devices, termed the “Laser Breast Scanner (LBS),” in place at five locations in the United States (Dartmouth, the University of Pennsylvania, Massachusetts General Hospital, UC San Francisco, and UC Irvine) that are measuring breast tumor response over the course of neoadjuvant chemotherapy treatment. Because data for the study is compiled from multiple devices and multiple operators, standardized methods for measurement and data processing are vital. To achieve this standardization, we have created a uniform training system for operators[17], a standardized measurement protocol[18], and a robust calibration and device performance evaluation system using tissue simulating phantoms.[19, 20]

What we have lacked up to this point is a systematic method for objectively evaluating data quality and rejecting data that cannot be reliably processed due to noise or measurement errors. Data quality can vary widely across patients, measurement conditions and even over the course of individual measurement sessions. Sources of noise from the instrument, operator and tissue must all be characterized. The wide range of tissue physiology across the numerous patients in the ACRIN 6691 study leads to data with a wide range of signal-to-noise ratios (SNR). Operator measurement errors such as poor probe-tissue contact, damaged optical/electrical connections and excess RF noise also lead to data with inadequate SNR or introduce artifacts. Currently, quality control is performed by the operator on a case-by-case basis. This can be effective, but it is time-intensive, operator dependent, difficult to standardize and mainly performed post data acquisition. An automated quality control system is desirable to effectively and uniformly filter out bad data while it is acquired, thus allowing for on-the-fly corrections to be made. In this paper we present a new quality control algorithm that achieves this goal and demonstrate its effectiveness on clinical data obtained in the ACRIN # 6691 trial.

![Diagram](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)
1.3. DOSI Measurement and Data Processing

The DOSI devices featured in the ACRIN trial use integrated frequency domain photon migration (FDPM) and steady-state (SS) reflectance methods to obtain broadband absorption coefficient ($\mu_a$) spectra separate from tissue scattering. The absorption spectra are then fit with known tissue extinction spectra (hemoglobin, water and lipids) to obtain tissue chromophore concentrations. FDPM data is collected by intensity-modulating laser diodes at frequencies swept from 50 to 600 MHz. Six wavelengths (660, 690, 785, 810, 830 and 850 nm) are used with fiber optic cables that deliver the light to the breast; diffuse reflectance is detected using an avalanche photodiode (APD) placed 28mm from the source optical fiber. The APD records the amplitude and phase shift of the collected light at each modulation frequency. SS data is also collected via diffuse reflectance, but the light source is a steady state broadband lamp and the detector is a grating-based spectrometer, which records intensity across 650-1000 nm. The entire data collection process is executed automatically by a GUI-based software package called EZDOS.[20] Both datasets are then processed together using the procedure diagrammed in Figure 1. First, FDPM data is fit to the diffusion equation using a semi-infinite boundary condition to obtain $\mu_a$ and reduced scattering coefficient ($\mu_s'$) values at the six discrete FDPM wavelengths.[21, 22] The discrete $\mu_s'$ spectrum is then fit to a power law to obtain a full $\mu_s'$ spectrum over the 650-1000 nm regime. Thus, the SS broadband reflectance is now only a function of absorption. The FDPM $\mu_a$ values are then used to appropriately scale the SS intensity data to obtain absolute reflectance values. Using this reflectance spectrum and the $\mu_s'$ spectrum, a numerical solution for $\mu_a$ is found at every wavelength. Because the final spectra are largely dependent on the initial FDPM amplitude and phase information, we have decided to focus the quality algorithm on this portion of the data.

2. DEVELOPMENT OF AN AUTOMATED QUALITY CONTROL ALGORITHM

The current implementation of the automated quality control algorithm focuses on FDPM data only at this time and is applied in two parts: (a) individual laser diode modulated amplitudes and (b) collective optical property consistency.

Figure 2. Modulated dark amplitude measured on the UCI DOSI instrument participating in the ACRIN trial. The modulation of each laser diode produces different levels of RF noise, resulting in a unique noise floor for each diode.
2.1. Filtering Phase and Amplitude

Each laser diode FDPM data (i.e., multi-frequency amplitude and phase) is first cleaned and filtered. The algorithm first characterizes measurement noise. It is necessary to take a “dark measurement” to find the true noise floor at each modulation frequency (similar to the “dark noise subtraction” measurement routinely employed in conventional spectroscopy). The “RF dark noise” is achieved by placing the DOSI probe on a sheet of black rubber to block all laser light sources and running five standard frequency sweeps. Multiple sweeps provide a mean and standard deviation of the “dark RF amplitude.” It is important to modulate the laser diodes under the same conditions as the tissue measurement when taking a dark measurement because the system noise floor of the system depends upon the specific laser diode and RF power delivered. Figure 2 provides a sample case; note that each diode has a different frequency-dependent noise floor. The EZDOS software has been modified to prompt the operator to take a dark measurement during the calibration procedure. After the dark measurement, data is collected normally.

When the data is processed (i.e., calibrated and then fit to the diffusion model), the noise-filtering portion of the algorithm compares the amplitude at each measurement data point to the dark amplitude at the same frequency for the same laser diode. Amplitudes at a given frequency that are within two standard deviations of the corresponding dark amplitude are excluded from the subsequent FDPM $\mu_a$ and $\mu_s'$ fits. The effect of noise-filtering on a typical noisy measurement taken with the ACRIN DOSI system at UC Irvine (UCI) is shown in Figure 3. This process has the advantage of selecting a unique frequency range for each laser diode and each measurement and can be done automatically. Operator-performed quality control generally picks one continuous frequency range for the entire data set, so the algorithm is much more sensitive to changes in noise levels throughout the measurement and is able to independently determine the appropriate frequency range for each laser diode.

![FDPM Fits On Noise-Filtered Data](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

Figure 3. FDPM fits on noise-filtered data from a noisy measurement. The noisy data (grey Xs) is effectively rejected while higher SNR data (black points) is kept, which is most apparent in the phase. Note that the algorithm is able to remove noise that is interspersed throughout the data, not just in one continuous block.

2.2. Filtering Optical Property Values

The second phase of the quality control algorithm is designed to filter out FDPM data that is skewed by sources of error not corrected by noise filtering. The $\mu_s'$ values over the entire 650-1000 nm wavelength range are calculated from the slope of the six FDPM $\mu_s'$ values, so they are strongly affected if the value for one diode is suspect. The algorithm takes
advantage of this fact by comparing power law fits of different combinations of FDPM $\mu'_a$ discrete spectra values. Normally, the processing code runs a least-squares power law fit on the six $\mu'_a$ values using a Levenberg-Marquardt algorithm and uses the resulting fit to calculate the full $\mu'_a$ spectrum; thus we are constraining the scattering to fit this model. If the average squared residual value of this fit is above a preset threshold, the quality control algorithm runs six additional fits using each possible combination of five $\mu'_a$ values (a different diode excluded each time). If the average squared residual value for one of these fits is less than half the value for all the other fits (including the original six-diode fit), then this fit is used to calculate the $\mu'_a$ spectrum, and the diode that was excluded from the fit is also excluded from the broadband $\mu_a$ fit. If no fit meets this criterion then the original six-diode fit is used. A comparison of a $\mu'_a$ spectrum with and without eliminating a diode is shown in Figure 4. The diode-elimination process is limited in that it requires that only one diode has skewed data, although the algorithm could in principle be expanded to additional wavelengths. Generally because measurements with two or more poor FDPM data rarely fit well using any combination of diodes, and should simply be flagged as suspect. Data with only one bad diode, on the other hand, can often yield high quality fits if the diode is removed.

![Figure 4. Optical property fits without (left) and with (right) automated diode elimination. The algorithm correctly eliminates the outlier and substantially improves the quality of the chromophore fit.](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

2.3. Quality Control Reporting

The progress of the QC algorithm is summarized in a detailed data report that is output at the conclusion of data processing. The purpose of this report is to better inform the operator about the exact steps taken by the algorithm and about which data files are suspect, thus allowing for a review after all data has been processed. The following information is provided in the data report:

- Files with diodes with less than 25% of points above the noise floor are flagged. The diodes are identified.
- Files with successful diode elimination are flagged, and the eliminated diode is identified.
- Files with unsuccessful diode elimination attempts are flagged.
- For each file, a full spreadsheet of which points were above the noise floor and which were below is saved.
- The total number of points above the noise floor, fraction of points above the noise floor, and average amplitude for each diode in each file are recorded.
- The noise floor used by the algorithm is saved in a file.
The data report allows the operator to see where the algorithm was able to improve data quality and where it identified problematic data that it cannot fix. In the following section, we present two cases illustrating the effectiveness of the quality control algorithm and the data reporting system.

3. QUALITY CONTROL ON ACRIN 6691 PATIENT MEASUREMENTS

3.1. ACRIN Clinical Data Acquisition

The general framework for DOSI data acquisition has been previously described and has been standardized for the ACRIN #6691 trial.[12] The patient data used in the examples below were taken from the ACRIN #6691 study. All subjects provided informed written consent, and all patient measurements were performed in strict adherence to the master clinical protocol (#6691) approved by ACRIN that was subsequently reviewed and approved by the review boards for each participating research center.

3.2. Test on Noisy Clinic Data

Figure 5 shows a pair of 2-dimensional maps of tissue optical index (TOI), defined as (ctHHb · ctH2O)/(%lipid), that was generated by manually scanning the DOSI probe across a patient’s breast.[18] The two images are constructed using the same data; the left is processed without any quality control and the right is processed using the quality control algorithm. The areola (which typically shows high TOI due to the glandular tissue) and tumor are indicated by the two circular regions in both images. In the right image, stars mark points that the algorithm flagged for low amplitude; circles mark points that were flagged for low amplitude and that had a diode removed from broadband processing; triangles mark points that were flagged for low amplitude and that had poor scattering fits that were unable to be fixed via diode removal. The rectangular sections in the left image are characteristic of data points with poor chromophore fits resulting in either extremely high or zero concentrations (i.e., unrealistic physiology). These regions are completely resolved after application of the quality control algorithm; only one point still fits poorly after the noise filtering and diode elimination steps and the corrected points now have physiologically consistent and realistic values. Before applying the quality control algorithm, nearly a quarter of the points in the tumor region were non-evaluable and would have to have been discarded. For smaller tumors (limited number of points) or assessment of tumor spatial heterogeneity, losing these points this could be problematic. The quality control algorithm allows us to process the noisy points and keep the dataset. We note that in the past manual filtering would have resolved some of these errors, but the quality control algorithm resolves the errors in a consistent and objective manner.

3.3. Unresolvable Data

In contrast to the above section, Figure 6 shows a pair of TOI maps from a different patient in the ACRIN clinical trial. The maps are once again constructed from the same data; the right image uses the quality control algorithm and the left image does not. Tumor and areola regions are marked, triangles mark points with poor scattering fits that could not be improved via diode removal and stars mark points with a diode removed. No points in this data were flagged for low amplitude. The combination of poorly fitting points and low noise is a clear indicator of experimental error. In this case, it was determined that at numerous times throughout the measurement, the operator was removing the probe from the patient’s skin before every diode was modulated, resulting in bad data on the diodes that were modulated last. Although the algorithm is unable to resolve such an error, the information in the data report serves as an easy indicator that the dataset should be excluded from the study. Upon subsequent review, it was determined that an instrument timing malfunction was the cause of low data quality.
Figure 5. Effect of quality control (QC) algorithm on TOI maps for a patient measurement. Stars mark points flagged for low amplitude; circles mark points that were flagged for low amplitude and that had a diode removed from broadband processing; triangles mark points that were flagged for low amplitude and that had poor scattering fits that were unable to be fixed via diode removal. The algorithm fixes the poor data points allowing the entire tumor area to be evaluated.

Figure 6. Example of QC algorithm not able to resolve problems in TOI maps. Triangles mark points with poor scattering fits that could not be improved via diode removal and stars mark points with a diode removed. Numerous points flagged for bad fitting but not for high noise indicate other sources of error.
4. DISCUSSION

The above examples serve as a proof of concept that the quality control algorithm can effectively, consistently and objectively improve data and fit quality by filtering out noise as well as and identify data with unresolvable sources of instrument/operator error. Without quality control, both sets of data would be deemed non-evaluable, and the patients would have to be excluded from the study. In the 41-patient study described by Udea et al. (ref [8]), six patients had to be excluded from the analysis due to poor SNR. We applied the current quality control algorithm to these excluded patients and found that three of them could be corrected. The quality control algorithm would have thus been able to reduce the number of excluded patients in this study by half. It is important to note that our previous subjective operator-based quality control was not able to adequately clean the noise in these excluded patient’s data; in contrast the quality control algorithm could have cleaned this data. Patient accrual can be difficult and is often the limiting factor to completing studies like ACRIN #6691. The more sophisticated data processing performed by the quality control algorithm ensures that if data is at all useable, it is not necessarily rejected from the study.

In addition to increasing patient accrual, the quality control algorithm also provides an efficient system for identifying and characterizing different sources of error. In the data presented in Section 3.3, the large number of flagged points quickly alerts the operator of potential problems with the data set. The data report allows us to know exactly which points to investigate as well as details on why those points were flagged; for this particular data, we noticed that there were no low-amplitude points and that every diode that fit poorly was modulated at the end of each measurement sequence. This led us to the conclusion that the probe was being lifted off the surface of the tissue too early yielding points that did not fit the semi-infinite model but had high amplitudes due to reflections off the tissue surface. We were then able to immediately contact the collaborators who had taken the measurements and inform them of the error, ensuring that no further measurements were compromised. An oversight process such as this one is vital to a multi-center trial such as ACRIN #6691 because it provides a fast, consistent way of monitoring measurements from every sight, and gives us the information we need to help every operator take accurate, repeatable measurements. The newly added dark measurement is fast, easy to perform, and is the only additional measurement step required to conduct the entire procedure; all other steps are done automatically by the processing code. We are constantly striving to reduce site and operator dependence in our measurements, and the quality control algorithm and data reporting system do so extremely effectively.

5. CONCLUSION

We have developed a novel automated quality control algorithm for DOSI instruments that improves data fitting, provides a standardized method of identifying poor quality data, and increases measurement quality and repeatability across the ACRIN sites. Since its implementation in late 2012, we have already demonstrated the algorithm’s effectiveness in several sets of patient data. As we continue to include the dark measurement step in the patient measurement procedure we will be able to compile a larger body of quality control processed data, which we will use to further analyze the algorithm’s effectiveness.

6. ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health under grants P41EB015890-33 (Laser Microbeam and Medical Program: LAMMP), U54-CA136400 (NTR), R01-CA142989, P30-CA62203 (University of California, Irvine Cancer Center Support Grant). Support also comes from the American College of Imaging Radiology Network (ACRIN) under Trial #6691 from U10 CA80098 and U10 CA079778. BLI programmatic support from the Beckman Foundation is acknowledged. We also thank all who helped to perform the setup for this study and ensuing tests and measurements at each ACRIN study site.
REFERENCES


[17] [ACRIN 6691 Training Videos], (2012).


