Improved accuracy in time-resolved diffuse reflectance spectroscopy.

Alerstam, Erik; Andersson-Engels, Stefan; Svensson, Tomas

Published in:
Optics Express

DOI:
10.1364/OE.16.010440

2008

Citation for published version (APA):
Improved accuracy in time-resolved diffuse reflectance spectroscopy

Erik Alerstam, Stefan Andersson-Engels and Tomas Svensson
Department of Physics, Lund University
Sweden
erik.alerstam@fysik.lth.se

Abstract: Significant improvements in the accuracy of time-resolved diffuse reflectance spectroscopy are reached by using a Monte Carlo scheme for evaluation of measured photon time-of-flight distributions. The use of time-resolved diffusion theory of photon migration, being the current standard scheme for data evaluation, is shown defective. In particular, the familiar problem sometimes referred to as absorption-to-scattering coupling or crosstalk, is identified as an error related to the breakdown of the diffusion approximation. These systematic errors are investigated numerically using Monte Carlo simulations, and their influence on data evaluation of experimental recordings are accurately predicted. The proposed Monte Carlo-based data evaluation avoids these errors, and can be used for routine data evaluation. The accuracy and reproducibility of both MC and diffusion modeling are investigated experimentally using the MEDPHOT set of solid tissue-simulating phantoms, and provides convincing arguments that Monte Carlo-based evaluation is crucial in important ranges of optical properties. In contrast to diffusion-based evaluation, the Monte Carlo scheme results in optical properties consistent with phantom design. Since the MEDPHOT phantoms are used for international comparisons and performance assessment, the performed characterization is carefully reported.

© 2008 Optical Society of America

OCIS codes: (170.5280) Photon migration; (300.6500) Spectroscopy, time-resolved; (170.6510) Spectroscopy, tissue diagnostics; (170.7050) Turbid media; (290.1990) Diffusion; (160.4760) Optical properties

References and links
1. Introduction

Photon time-of-flight spectroscopy (TOFS), also known as time-resolved spectroscopy, is a well established technique for characterization of scattering media. Its ability to accurately and quantitatively assess both absorption and scattering is extensively utilized in the field of biomedical optics. Applications include in vivo determination of physiological and optical parameters of muscle [1], the human brain [2], breast tissue [3], and human prostate tissue [4]. Furthermore, TOFS is an important tool in the emerging field of optical tomography [5]. A slightly differ-
ent area of application is pharmaceutical analysis, where TOFS has been used for analysis of chemical composition and physical properties [6, 7, 8].

Modeling of light propagation in scattering materials (i.e. photon migration) is a fairly complex matter. The photon time-of-flight (TOF) distribution, as recorded in TOFS, depends on refractive index, absorption and scattering coefficients, scattering anisotropy, sample geometry and boundary conditions, as well as on the size and location of source and detector. The complexity of this problem is reflected in the fact that Monte Carlo (MC) simulation is the gold standard for photon migration modeling. MC allows direct simulation of radiative transport theory, and is therefore considered highly accurate. The difficulties connected with the use of MC for data evaluation have, however, prevented it from becoming a tool for routine data analysis. This is especially the case in TOFS, where most photon migration modeling is based on the diffusion approximation of radiative transport theory. In simple geometries, diffusion theory supplies analytical expressions for photon TOF distributions, and dramatically simplifies data evaluation.

As early as 1990, Yoo et al. pointed out that time-domain diffusion theory is incapable of describing the propagation of light pulses in many materials of physical interest [9]. Since then several authors have investigated the validity of time-domain diffusion theory, and its dependence on how boundary conditions are taken into account [10]. Hielscher et al. concluded that the theory fails in reproducing the results from MC simulations of diffuse reflectance, and that the determination of reduced scattering coefficients suffer from particularly large errors [11]. Similar findings are reported by Cubbedu et al. who concluded that the model performance varies with the range of optical properties, as well as on experimental configuration [12]. Kienle et al. refined the diffusion models used for time-resolved diffuse reflectance, but reported that significant deviations from MC remains [13]. The three above cited publications clearly shows that diffusion modeling fails in describing time-resolved diffuse reflectance in important ranges of optical properties, regardless of how boundary conditions are treated. Despite these findings, diffusion modeling has remained the standard tool for data evaluation in TOFS.

In general, the validity of diffusion modeling decreases with increasing absorption, decreasing scattering, and decreased source-detector separation. Motivated by the high absorption and low scattering encountered in human prostate [4], we recently developed a scheme for Monte Carlo evaluation of TOFS data [14]. The developed scheme is referred to as White Monte Carlo (WMC), and is based on early ideas of the scalability of Monte Carlo simulations [15, 16, 17]. The speed and flexibility of the WMC approach makes it suitable for routine evaluation of TOFS in both reflectance and interstitial configurations, over a wide range of optical properties. The superior performance of WMC-based data evaluation has been carefully verified for the interstitial geometry, both theoretically and experimentally using liquid phantoms (intralipid and ink). More recently, WMC modeling has been used to significantly improve the accuracy of in vivo TOFS characterization of prostate tissue [18].

WMC evaluation is of interest also for the important case of diffuse reflectance, especially since Monte Carlo accurately can account for boundary effects. The present work aims at giving both experimental and theoretical evidence that WMC-based data evaluation significantly improves the performance of time-resolved diffuse reflectance (i.e. reflectance TOFS). It also aims to explain and quantify the model-related errors induced by diffusion modeling of diffuse reflectance. In particular, WMC is shown to avoid and explain the previously reported artifacts of TOFS referred to as absorption-to-scattering coupling or crosstalk [19, 20]. Experimental work is carried out on the solid phantoms prepared by Pifferi et al. that originally was intended for use in an international comparison of photon migration instrumentation (the MEDPHOT phantoms) [19]. Since WMC modeling is shown superior to diffusion modeling, the characterization of these phantoms is an important part of the results of the present article. Several authors have stressed the importance of calibrated and characterized reference phantoms [19, 21, 22], and we argue that the data presented here is, up to now, the most accurate assessment of the optical properties (absorption and reduced scattering) of these phantoms.
2. Materials and methods

2.1. Time-of-flight instrumentation

Photon time-of-flight experiments were conducted using a compact (50x50x30 cm³) and portable time-domain photon migration instrument primarily intended for spectroscopy of biological tissues in clinical environments. Detailed information on the instrumentation can be found in previous publications [4]. Briefly, the system is based on pulsed diode laser technology and time correlated single photon counting (TCSPC). Four pulsed diode lasers (at 660, 786, 830, and 916 nm) generate 70 ps FWHM pulses (average power 1-2 mW). Light is injected into the sample and collected using 600 μm GRIN optical fibres. The collected light is sent to a fast MCP-PMT connected to a TCSPC-card that records the time-of-flight histograms with 24.4 ps time resolution. The instrument response function (IRF) is measured by putting the two fibre ends face-to-face with a thin paper coated on both sides with black toner. The total broadening of the system yields an IRF of approximately 100 ps FWHM.

2.2. The MEDPHOT phantom kit

This study utilized the MEDPHOT phantom kit, a collection of 32 solid cylinders (4.5 cm thick, 10.5 cm diameter) with different scattering and absorption properties. These epoxy-based solid phantoms were fabricated as a part of the MEDPHOT protocol, intended to be circulated among research groups, to allow comparison of the performance of different instruments. The phantoms were manufactured to combine four concentrations of scatterer (TiO₂ powder) with eight concentrations of absorber (toner) in linear and equally spaced steps. The phantoms are labeled with a letter and a number, where the letter indicate the nominal scattering (A, B, C and D, corresponding to \( \mu_s' = 5, 10, 15 \) and 20 cm⁻¹, respectively, at 800 nm) and the number indicate the nominal absorption (1, 2, 3, 4, 5, 6, 7, 8 corresponding to \( \mu_a = 0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35 \) cm⁻¹, respectively, at 800 nm)[19]. The refractive index of the resin matrix was assumed to be \( n = 1.55 \) [23]. Also (owing the Monte-Carlo based evaluation method), the anisotropy factor of the TiO₂ had to be estimated. Based on integrating sphere measurements by Swartling et al. [24], on the same brand of TiO₂ powder (T-8141; Sigma-Aldrich, St. Louis, Missouri) the anisotropy factor, \( g \), was assumed to be 0.75, which is large enough for the similarity relation to apply, i.e. independence of \( g \) on the derived \( \mu_s' \) [25, 26].

Unfortunately the phantom A8 was missing when the phantoms were shipped to us and our measurements hence do not include this phantom.

2.3. Experimental

The measurements were performed by putting the optical fibres (guided by thin stainless steel tubes) in contact with the sample. The fibre separation was fixed at \( \rho = 15 \) mm (center to center), and the positioning of the fibre pair on the phantom was random (somewhere in the middle of the phantom). The space between the fibres was occupied by a simple light-trap (black paper folded several times in contact with the phantom somewhere in between the fibres) to minimize the possibility of light-leakage into the collecting fibre. For the same reason, all adjacent surfaces were covered in black paper. Data were collected for 30 seconds for each measurement.

To minimize the temporal drifts, all measurements were conducted in a temperature stabilized lab where the system had been running for several hours prior to each measurement session. The temperature in the lab, as well as inside the system, was monitored to ensure stability during the measurements. The IRF was recorded approximately every 15 minutes during measurements as well as prior to and after each session. The phantoms were measured in random order independently on three occasions (2007-12-21 and twice 2008-01-08, 1.5 hours apart), denoted run 1, run 2 and run 3 in chronological order (represented by circles, diamonds and squares respectively throughout the figures in this work).
2.4. Modeling

This work utilizes two different forward models for photon propagation for data evaluation and mutual comparison, the Monte-Carlo based White Monte Carlo (WMC) model [14] and the diffusion approximation of radiative transport theory, utilizing the extrapolated boundary condition (EBC diffusion) [10, 13]. Since the fibre separation is significantly smaller than the phantom dimensions, it is assumed that the phantoms can be treated as semi-infinite.

The WMC model has been extensively explained in [14]. Briefly, the scheme is based on the scalability of Monte-Carlo simulations in certain geometries, e.g. infinite and semi-infinite homogeneously scattering media. Hence a single simulation, comprising several billion photons, is performed and the resulting photon distribution can be scaled to the desired $\mu_s'$ and $\mu_a$. The approach provides a fast and accurate equivalent to traditional Monte Carlo that can be used as a forward photon propagation model. The input parameters for the database-simulation used in this work were: Semi-infinite media, $\mu_{s,\text{max}}=90$ cm$^{-1}$, $t_{\text{max}}=2$ ns, NA=0.29, $n=1.55$, $g=0.75$ and $6 \times 10^9$ photons. During this simulation, the photons are simulated at $\mu_s = \mu_{s,\text{max}}$ in an absorptionless media ($\mu_a = 0$). Photons are terminated when they escape the media or when the simulated time-of-flight exceeds $t_{\text{max}}$. This results in a time interval $[0, t_{\text{max}}]$ where the time-of-flight distribution is valid. However, during spatial (i.e. $\mu_s$) scaling from $\mu_{s,\text{max}}$ to $\mu_s$ this time interval also scales as $[0, \alpha t_{\text{max}}]$ where $\alpha = \mu_{s,\text{max}} / \mu_s$. This implies that WMC will be a less advantageous when the resulting pulses are broad in time, i.e. at high scattering and low absorption.

During WMC evaluation of time-of-flight data, the fitting procedure is based on an exhaustive search over a pre-defined $\mu_s'$ interval with a finite resolution, $\Delta \mu_s'$. For each $\mu_s'$ value, the optimal values of $\mu_a$ and a free amplitude parameter, $k$ are determined using a Marquard-Levenberg minimization of the error norm:

$$\tilde{\chi}^2(\mu_s') = \min \{\chi^2(k, \mu_a, \mu_s')\}.$$  

(1)

In this work, a $\mu_s'$ resolution of $\Delta \mu_s' = 0.05$ cm$^{-1}$ was used.

The impulse response of a semi-infinite media, modeled using the diffusion approximation of radiative transport theory with the extrapolated boundary condition (EBC-diffusion) is given by Eq. 2 [10, 13].

$$R(\rho, t) = a(n) \Phi(\rho, z = 0, t) + b(n) R_f(\rho, t),$$  

(2)

where $\rho$ is the source-detector separation, and the coefficients $a(n)$ and $b(n)$ are dependent on the refractive index of the medium. The reflectance (Eq. 2) is the sum of two terms, the fluence rate, $\Phi$, and the flux, $R_f$, weighted by $a$ and $b$. The two terms are given in Eq. 3 and Eq. 4 respectively.

$$\Phi(\rho, z = 0, t) = \frac{c'}{(4\pi Dc')^{3/2}} \exp(-\mu_a c' t) \left[ \exp\left( -\frac{r_1^2}{4Dc' t} \right) - \exp\left( -\frac{r_2^2}{4Dc' t} \right) \right],$$  

(3)

$$R_f(\rho, t) = \frac{1}{2} \frac{c'}{(4\pi Dc')^{3/2}} t^{-5/2} \exp(-\mu_a c' t) \left[ z_0 \exp\left( -\frac{r_1^2}{4Dc' t} \right) + (z_0 + 2zb) \exp\left( -\frac{r_2^2}{4Dc' t} \right) \right],$$  

(4)

Here, $r_1^2 = z_0^2 + \rho^2$ is the squared distance to the positive source, $r_2^2 = (z_0 + 2zb)^2 + \rho^2$ the squared distance to the negative source, and $c'$ is the speed of light in the medium. In this work, the diffusion coefficient, $D$, is defined in the absorption-independent way [27, 28, 29, 30], $D = (3\mu_s)^{-1}$. The two distances $z_0$ and $zb$ related to the source mirroring of EBC are given by:

$$z_0 = \frac{1}{\mu_a + \mu_s'},$$  

(5)
where \( R_{\text{eff}}(n) \) is the effective reflection coefficient which is dependent on the refractive index of the medium. Using Eq. 2.3.5 and 2.3.7 of [10] to calculate \( R_{\text{eff}} \) for \( n = 1.55 \) yields \( R_{\text{eff}} = 0.599 \). The \( a \) and \( b \) coefficients are calculated using Eq. 7 of [13], resulting in \( a(1.55) = 0.094 \) and \( b(1.55) = 0.251 \).

When the diffusion approximation is employed for data evaluation, an optimal fit between experimental data and \( kR(\rho, t) \) is reached iteratively by employing Levenberg-Marquardt optimization (in which \( \mu_s' \), \( \mu_a \), and the free amplitude parameter \( k \) are adjusted).

It should be noted that during the iterative data-fitting procedure of both the EBC-diffusion and the WMC model, the impulse responses provided by the models are convoluted with the recorded IRF. To minimize the effect of an uncertainty in the recorded IRF all data below 20% of the peak intensity is not used during the fitting procedure. In addition, during EBC-diffusion fitting, all data below 80% on the leading edge is disregarded as diffusion modeling is known to be a poor model for early photons [9, 12, 13]. The remaining data used during fitting is said to be within the fitting range. For simplicity, fitting ranges are denoted by the above stated percentages, e.g. EBC-diffusion uses a 80/20 fitting range while WMC uses a 20/20 fitting range.

During evaluation of experimental data the data points were weighted with the square-root of the signal given the normal distributed noise. During diffusion-based evaluation of WMC data, all data points were equally weighted (the noise characteristics of WMC simulations are not directly comparable to that of experimental recordings).

Finally, note that the two above mentioned models apply for reflectance in semi-infinite geometries, while the phantoms have a finite diameter and a finite thickness. However, since the fibre separation is significantly smaller than the phantom dimensions, it is assumed that the phantoms can be treated as semi-infinite. In fact, Monte Carlo simulation in an infinite geometry for a worst case scenario (\( \mu_s' = 3 \) cm\(^{-1} \) and \( \mu_a = 0 \) cm\(^{-1} \), i.e. both lower scattering and lower absorption than our phantoms) shows that none of the photons that are detected at \( \rho = 15 \) mm and fall within the 20/20 fit range would have been affected by changing the geometry from a semi-infinite half-space to the actual phantom geometry. This was concluded by tracking \( 2 \times 10^8 \) photons and recording their maximum depth and maximum radial excursion (\( 4 \times 10^5 \) photons contributed to the TOF distribution at \( \rho = 15 \) mm). Note, however, that if longer time-of-flights are included (e.g. when the fit range is extended), phantom boundaries may become an issue.

## 3. Results
Motivated by the fact the optical properties of the MEDPHOT phantoms commonly are characterized at around 800 nm, the presentation of experimental results in this section focus on the measurements at 786 nm. The results do however apply to all four measured wavelengths, and all data is enclosed in the Appendix. Note also that the simulation results apply to all wavelengths.

### 3.1. WMC versus the EBC diffusion model
In order to quantify the errors related to the use of EBC diffusion modeling, the corresponding TOF distributions were fitted to the TOF histograms delivered by the WMC model. The WMC-model, being equivalent to traditional Monte Carlo, is considered the gold standard, i.e. the optical properties used by the WMC model, \( \mu_{WMC} \), are considered true optical properties. To provide results comparable to our experimental results a 80/20 fitting range was used, and the temporal channel width was 24.4 ps (equal to the experimental channel width, see Sect. 2.1). Furthermore, to make the results applicable to actual measured data, both the WMC and
diffusion-based TOF distributions were convoluted (during the fitting procedure) with an IRF from our system. The same IRF was used in all convolutions (note that this eliminates the IRF measurement uncertainty that may aggravate real measurements). To verify that the results obtained were not dependent on the specific IRF used, the procedure was repeated with a 100 ps FWHM Gaussian, representing an almost ideal IRF, showing only minor differences.

The parameter space covered, i.e. the parameter spaced used to extract TOF distributions from the WMC-model, was:

$$0.01 \leq \mu_a \leq 0.75 \text{ [cm}^{-1}]$$
$$2 \leq \mu'_s \leq 18 \text{ [cm}^{-1}]$$

The fibre separation and refractive index were fixed at $\rho = 15 \text{ mm}$ and $n = 1.55$.

The relative error in derived optical properties is defined as:

$$\Delta \mu = \frac{\mu_D - \mu_{WMC}}{\mu_{WMC}},$$ (7)

where $\mu_D$ is the optical properties derived when employing data evaluation based on the EBC diffusion model. The corresponding error map is shown in Fig. 1. The results are similar to those presented for infinite media in Ref. [14], indicating that the presence of a boundary does not significantly worsen the performance of diffusion based modeling when using a reasonable boundary condition. In this context, it is also interesting to study the case of an ideal TOFS-system, i.e. one that exhibits an infinitely short IRF so that true impulse responses can be studied. An error map for impulse responses is shown in Fig. 2.

![Fig. 1. Relative errors due to EBC diffusion modeling when using a 80/20 evaluation range, and involving a convolution with an IRF recorded by our TOFS system. The dashed line indicates the zero relative error and hence the border between overestimation and underestimation of the derived parameters. For most optical combinations of $\mu'_s$ and $\mu_a$ overestimations of the derived optical properties occur while minor underestimations occasionally are observed when the absorption is low.](image)

When comparing Fig. 1 and 2, one should take into consideration that the effective fitting ranges of the two are different, despite both of them using an 80/20 fit range. Regardless, they exhibit similar behaviour. This indicate that the performance of EBC-diffusion based evaluation is not heavily affected by the presence of a reasonably short, non-ideal IRF (as long as the IRF is recorded accurately).

### 3.2. Experimental results

The experimental results for $\lambda = 786 \text{ nm}$ are presented in Fig. 3, and includes evaluations based on both WMC (a), and EBC diffusion modeling (b). The three runs, 1, 2, and 3, are presented as
Fig. 2. Relative errors due to EBC diffusion modeling when using a 80/20 fit range for the impulse response. The dashed line indicates the zero relative error and hence the border between overestimation and underestimation of the derived parameters. As in Fig. 1, overestimations of the derived optical properties dominate. The apparent instability in derived $\mu_a$ for high absorption and low scattering media is most likely related to the few number of data points eligible for evaluation resulting from the short pulses exhibited by such media.

circles, diamonds and squares respectively, illustrating the high reproducibility of our measurements. Note that a single outlier is present (A7, run 2), presumably related to an experimental error. The mean scattering and absorption of all the constant-level series, as obtained from

Fig. 3. A so called accuracy plot (see Ref. [19]) showing derived $\mu'_s$ and $\mu_a$ at 786 nm for run 1 (circles), run 2 (diamonds) and run 3 (squares), using (a) WMC evaluation, and (b) diffusion evaluation. From the design of the phantoms, points are expected to fall on a grid. The mean scattering and absorption values for the WMC measurements are shown as dashed lines in both figures. In addition, the mean WMC-derived optical properties with corrections according to the error-map (Fig. 1) are shown in (b) as red crosses.

WMC-based data evaluation, are presented as dashed lines (vertical lines for constant scattering series and horizontal lines for constant absorption series). For comparison, these lines are also presented in part (b) where the outcome of diffusion modeling is presented. It is apparent that the optical properties derived using the WMC approach exhibit linear behavior, while the linearity is disrupted by model-related errors when data evaluation is based on EBC diffusion modeling. This erroneous behavior was predicted in Sect. 3.1. All the WMC-derived data (for all four wavelengths) are explicitly given in Tab. 1 and 2 in order to allow other researchers to
directly compare their results to those presented here.

It is important to note that the overall pattern of derived optical properties deviates from the perfect grid-like pattern expected from the phantom design. An obvious example is the B1 phantom, exhibiting significantly lower scattering than other B-phantoms. In addition, the variations in scattering among the D-phantoms are surprisingly large. The good reproducibility shows that this deviation is related to phantom manufacturing rather than measurement uncertainty. This conclusion is further supported by noting that correlated systematic variations from phantom design have been reported by Pifferi et al. (their work is based on transmission measurements and diffusion modeling, resulting in patterns similar to that shown in Fig. 3(b)).

To verify the quantification of the diffusion related errors as presented in Fig. 1, the mean WMC-derived optical properties of each phantom were multiplied by the corresponding (interpolated) relative error. The resulting optical properties constitute a prediction of the outcome of diffusion modeling, and are presented in Fig. 3(b) as red crosses. The agreement appears very good, showing that the error map is highly relevant and applicable to actual measurements.

Linearity plots of the derived \( \mu_a \) for the four constant scattering series (A-D) are shown in Fig. 4. The WMC-derived \( \mu_a \) exhibits linear behavior, while the EBC-diffusion derived data features a non-linear increase. This is more evident in the lower scattering series (e.g. A phantoms) as the relative error decreases with increasing scattering (as previously shown in Sect. 3.1). Note that at higher scattering (e.g. D phantoms), the obtained \( \mu_a \) pattern can easily be mistaken for a linear increase.

The corresponding linearity plots for the derived \( \mu_s' \), for the constant absorption series (1-8), are shown in Fig. 5. As expected from the phantom design, both diffusion and WMC modeling results in a linear increase in derived \( \mu_s' \). However, in good agreement with previous results reported in Ref. [14], the parameters derived from the diffusion model exhibit an offset-like behavior that causes the extrapolated scattering at zero nominal scattering to significantly deviate from zero. The severity of this overestimation is directly related to the amount of absorber added to the phantom. Due to the imperfections in phantom manufacturing it is, however, difficult to analyze these linearity plots in more detail. An example of this issue is the non-zero offset resulting from the linear extrapolation of the \( \mu_s' \) values of the 8-phantoms (highest absorption) even for WMC modeling (this offset is further discussed in Sect. 4).

### 3.3. Model stability with regards to fit range

The importance of the fit range was studied by evaluating experimental data using both EBC diffusion and WMC for different fit range settings. All data points beyond the trailing edge point corresponding to 20% of the peak maximum were always disregarded. The start of the fit range was varied from 10% to 20% of the peak maximum. The results showed that the choice of fit range has a significant impact on the derived optical properties. A fit range of 15% resulted in the most consistent and accurate results across all phantoms.
Fig. 5. Linearity plot of the derived $\mu'_{s}$ (at 786 nm) for the constant absorption series (1-8). All WMC-derived values (for all runs) are shown, as well as the best linear fit to these points. The mean values obtained from diffusion modeling are shown as red crosses.

The fit range was varied between 1% on the leading edge to 50% on the trailing edge (of peak maximum). The results for the phantoms A1, A7 B4 and D4 (the 786 nm measurement, run 1) are shown in Fig. 6. For diffusion evaluation, the trend is that severe overestimations occur when early data are included. The overestimation decrease (or turn to underestimations) as fewer early data points are included in the fit. The magnitude of the overestimation decreases with increasing scattering and decreasing absorption, as seen also in Fig. 1, while the negative slope with respect to the early fit range limit remains. Using WMC data evaluation, both derived scattering and absorption exhibits insensitivity to the early fit range limit as long as the peak of the data is included. With a few exceptions, such as the D4-phantom (run 1) where a slight positive slope is present, this desirable behavior was seen in almost all measurement. Note that the occurrence of minor slopes appears to be measurement specific, rather than phantom or modeling specific. The phenomenon is observed mainly in highly attenuating phantoms, and may be attributed to e.g. light leakage into the collecting fibre, uncertainties in the recorded IRF or temporal drifts.

4. Discussion

Time-resolved measurements of diffuse reflectance (i.e. time-of-flight spectroscopy, TOFS) are frequently used for characterization of turbid materials. Although the diffusion approximation of light propagation is the current standard for data evaluation in TOFS, this work clearly shows that the accuracy can be significantly improved if refined models (e.g. WMC) are used.

Although the limited validity of diffusion models has been known for many years [9], little has been done to understand and overcome the corresponding errors in diffusion-based data evaluation. This is presumably related to the difficulty in finding alternatives suitable for practical use. The WMC approach proposed in the present paper is a competitive scheme for routine data evaluation, and is shown to avoid the errors of diffusion-based data evaluation. Since it is based on Monte Carlo simulation, being the gold standard for modeling of light propagation within the field of biomedical optics, it is difficult to identify obvious further improvements in terms of model correctness. It should, however, be noted that scalability requirement of the WMC limits its applicability to e.g. infinite or semi-infinite geometries. More complex geometries can be handled by a similar approach, involving multiple MC simulations at different $\mu'_{s}$.
Fig. 6. Derived $\mu'$ and $\mu_a$ as a function of the early fit range limit for the Phantoms A1, A7, B4 and D4 (786 nm data from run 1). The small black and red dots indicate WMC and diffusion evaluation respectively. Each point illustrates the derived $\mu'$ and $\mu_a$ when excluding all earlier the data points during the nonlinear fitting procedure, i.e. the fitting range extend from that point (inclusive) to 20% of the peak maximum at the trailing edge. The corresponding TOF histograms are shown for reference. A1 and A7 illustrates the performance in low scattering A phantoms while B4 and D4 exemplifies behavior in the B, C and D phantoms.
[15, 17]. Higher order approximations of the radiative transport equation may in the future provide additional means for implementing refined data evaluation in TOFS.

The value of the WMC approach, and the deficiency of diffusion modeling, is highlighted in several ways. The three following paragraphs discuss important aspects of this finding:

First, the diffusion model is compared numerically to Monte Carlo simulations in Sect. 3.1. There, error-maps are used to provide a quantitative measure of the errors induced by diffusion evaluation. In contrast to previous studies, where Monte Carlo and diffusion are compared at a single or at few combinations of $\mu_a$ and $\mu'_s$ [11, 13], these maps shows the errors in an entire range of optical properties. In addition, Fig. 3(b) shows how the error-maps can be used to predict the erroneous outcome of diffusion evaluation of experimental data. The error-maps can be considered to provide estimates of relative errors and a guidance to whether diffusion is applicable (their potential value as a tool for correction of diffusion-based data evaluation is not fully examined). It is also interesting to note that the error-maps for semi-infinite (reflectance) geometries, as presented in the present article, resemble those presented for infinite geometries in Ref. [14] (despite differences in refractive index, anisotropy factor and fit range). Hence, one can argue that the poor performance of diffusion theory in semi-infinite media should be assigned to the breakdown of the diffusion approximation itself, rather than to inappropriate account for boundary conditions. Hielscher et al. came to a similar conclusion when investigating different boundary conditions, as they all failed to predict Monte Carlo derived results [11]. Logically, if the diffusion approximation breaks down in a certain region of optical properties even in infinite media [14], one should not expect a performance improvement when investigating the more complex case of geometries with boundaries.

Second, the performance of WMC and diffusion evaluation are compared experimentally using the MEDPHOT set of tissue-simulating phantoms. Fig. 3. These phantoms were manufactured to produce four levels of reduced scattering, and eight levels of absorption (32 phantoms in total). In a so called accuracy plot shown in Fig. 3, derived values $\mu'_s$ and $\mu_a$ are expected to fall on a grid consisting of equidistant vertical and horizontal lines (different spacing between $\mu_a$ and $\mu'_s$ levels). As seen in Fig. 3(b), diffusion evaluation clearly fails to reproduce a grid-like pattern. The discrepancy between observed and expected pattern has been reported in several publications, and is often referred to as absorption-to-scattering coupling (or crosstalk) [19, 12, 20]. In contrast, WMC evaluation produces data points on a grid that is consistent with the nominal values of the phantoms. Hence, the above mentioned artifacts are related to the breakdown of the diffusion approximation, and can be completely eliminated by employing the WMC approach. The minor and irregular deviations from a perfect grid that remains even when WMC is employed is most likely related to imperfections in phantom manufacturing (i.e. the actual optical properties are not as intended). As discussed in Sect. 3.2, convincing evidence for this conclusion is (i) the good reproducibility as shown in Fig. 3(a), and (ii) that the observed deviations correlate to inter-phantom variations as reported by Pifferi et al. despite difference in measurement geometry, instrumentation and operators (see Fig. 6(c) in Ref. [20]). Note that this holds also for the surprisingly large variations in $\mu'_s$ for the D-phantoms. The apparent difficulty in the manufacturing of phantoms suggests that intra-phantom heterogeneity cannot be ruled out (and since heterogeneity would decrease reproducibility, this issue may deserve further attention). Furthermore, it is important to remember that phantom imperfections affect the linearity plots shown in Figs. 4 and 5. This is particularly obvious in part (8) in Fig. 5, where the scattering of B8 is slightly above the average B scattering level while the D8 phantom exhibits a scattering slightly lower than the average D phantom (see Fig. 3(a)), misleadingly suggesting a significant scattering offset (note that this plot also suffers from the lack of the A8 phantom). Nonetheless, the linearity plots clearly shows the deficiency of diffusion evaluation, while the outcome of WMC evaluation is highly satisfying. Note, however, that the outcome of diffusion evaluation unfortunately may be mistaken for a linear increase in derived $\mu_a$ or $\mu'_s$ together with a scattering offset.
Third, the two evaluation schemes are compared by investigating their robustness with respect to fit range. As clearly seen in Fig. 6, and as expected for a correct model, the WMC evaluation is largely independent of the selection of fit range (as long as the peak is included). In great contrast, diffusion evaluation is often highly sensitive to the selection of fit range start time. Since the influence of the fit range selection varies with the range of optical properties (and differs between $\mu_a$ and $\mu'_s$), this figure also illustrates that finding an optimal fit range is not possible. In general, if diffusion is used for data evaluation, it appears wise to exclude a majority of the early data points while still keeping the peak data point. This is in agreement with the findings of Cubbedu et al. [12]. It is interesting to note that both WMC and diffusion evaluation becomes unstable when the fit range start time is selected so that the peak reflectance data point is excluded from data evaluation. Since most of the information on scattering is found in the early part of the TOF distribution, this may come as no surprise. However, Kienle et al. conducted a similar (but theoretical) study of diffusion stability with respect to fit range, showing that stable evaluation can be achieved even when excluding the peak [13]. The failure of both models to do so (in the present experimental study) might indicate a systematic error in the measurements, such as an inaccurate IRF recording procedure.

A question related to the fit range aspect discussed in the previous paragraph, is whether an investigation of fit range sensitivity can reveal measurement quality. If the fit range selection has a systematic influence on WMC-based evaluations, this may indicate systematic measurement problems such as temporal drifts or light leakage (assuming that the WMC model is valid, which can be questioned in, for example, heterogeneous materials). In fact, a slight fit range dependence, such as that shown for phantom D4 in Fig. 6, was sometimes observed for the more highly attenuating phantoms. This problem may be assigned to light leakage, as this effect becomes increasingly important for phantoms with higher attenuation. In addition, this difficulty may influence the reproducibility (the highly attenuating D phantoms exhibit a slightly lower reproducibility than the other groups).

Finally it should be noted that the present work is concerned with time-of-flight spectroscopy and time-resolved diffusion theory. Frequency domain photon migration instrumentation, involving lower detection frequencies than TOFS, may be less sensitive to the breakdown of the diffusion approximation [31].

5. Conclusion

The present work presents experimental and theoretical evidence that the transition from diffusion models to WMC significantly improves the accuracy of time-resolved diffuse reflectance spectroscopy in ranges of optical properties of great interest to the biomedical optics community. Of particular importance is (i) that WMC evaluation eliminates the previously reported and familiar artifacts of TOFS known as absorption-to-scattering coupling or crosstalk, and (ii) that the evaluation outcome is largely independent of the fit range setting. In this work, the artifacts are identified as being due to the breakdown of the diffusion approximation. The use of error-maps allows accurate prediction of the errors related to diffusion evaluation, and is a valuable tool when determining the validity of diffusion theory. While the use of refined data evaluation is shown crucial in certain ranges of optical properties, it is also shown that diffusion model can be used successfully as long as scattering is sufficiently high. Since the breakdown is gradual, and depends on optical properties and measurement geometry, it is difficult to generalize when diffusion modeling should be avoided.

The present article also provides the first characterization of the MEDPHOT phantoms that is consistent with the nominal optical properties. The derived optical properties of these phantoms are therefore carefully stated for the wavelengths 660, 786, 830 and 916 nm. Also, we argue that the observed deviations from the phantom design are due to imperfections in phantom manufacturing. These imperfections must be considered when using these phantoms for performance assessment.
Acknowledgment

We are grateful to the biomedical optics group at Politecnico di Milano for generously letting us use the MEDPHOT phantoms. We gratefully acknowledge the financial support through the EC grant Nano-UB Sources.

Appendix

Table 1. Derived reduced scattering coefficients [cm$^{-1}$]. * Due to the limited time range where WMC is valid, these measurement had to be evaluated using a reduced fit range (20/40).

<table>
<thead>
<tr>
<th></th>
<th>run1</th>
<th>run2</th>
<th>run3</th>
<th>run1</th>
<th>run2</th>
<th>run3</th>
<th>run1</th>
<th>run2</th>
<th>run3</th>
<th>run1</th>
<th>run2</th>
<th>run3</th>
</tr>
</thead>
<tbody>
<tr>
<td>660 nm</td>
<td>4.75</td>
<td>5.00</td>
<td>5.00</td>
<td>3.70</td>
<td>4.10</td>
<td>4.00</td>
<td>3.35</td>
<td>3.70</td>
<td>3.65</td>
<td>3.00</td>
<td>3.25</td>
<td>3.20</td>
</tr>
<tr>
<td>786 nm</td>
<td>5.40</td>
<td>5.25</td>
<td>5.15</td>
<td>4.30</td>
<td>4.30</td>
<td>4.15</td>
<td>3.65</td>
<td>3.90</td>
<td>3.75</td>
<td>3.05</td>
<td>3.35</td>
<td>3.25</td>
</tr>
<tr>
<td>830 nm</td>
<td>5.55</td>
<td>5.85</td>
<td>5.85</td>
<td>4.50</td>
<td>4.65</td>
<td>4.70</td>
<td>3.95</td>
<td>4.20</td>
<td>4.10</td>
<td>3.50</td>
<td>3.65</td>
<td>3.50</td>
</tr>
<tr>
<td>916 nm</td>
<td>5.55</td>
<td>5.75</td>
<td>5.65</td>
<td>4.40</td>
<td>4.70</td>
<td>4.55</td>
<td>3.95</td>
<td>4.25</td>
<td>4.15</td>
<td>3.10</td>
<td>3.60</td>
<td>3.60</td>
</tr>
</tbody>
</table>

* Due to the limited time range where WMC is valid, these measurement had to be evaluated using a reduced fit range (20/40).
Table 2. Derived absorption coefficients [cm$^{-1}$]

<table>
<thead>
<tr>
<th></th>
<th>660 nm</th>
<th>786 nm</th>
<th>830 nm</th>
<th>916 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>run1</td>
<td>run2</td>
<td>run3</td>
<td>run1</td>
</tr>
<tr>
<td>A1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>A2</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>A3</td>
<td>0.15</td>
<td>0.15</td>
<td>0.14</td>
<td>0.13</td>
</tr>
<tr>
<td>A4</td>
<td>0.20</td>
<td>0.21</td>
<td>0.20</td>
<td>0.17</td>
</tr>
<tr>
<td>A5</td>
<td>0.25</td>
<td>0.28</td>
<td>0.26</td>
<td>0.23</td>
</tr>
<tr>
<td>A6</td>
<td>0.34</td>
<td>0.35</td>
<td>0.34</td>
<td>0.30</td>
</tr>
<tr>
<td>A7</td>
<td>0.41</td>
<td>0.37</td>
<td>0.38</td>
<td>0.34</td>
</tr>
<tr>
<td>B1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>B2</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>B3</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.12</td>
</tr>
<tr>
<td>B4</td>
<td>0.22</td>
<td>0.22</td>
<td>0.21</td>
<td>0.19</td>
</tr>
<tr>
<td>B5</td>
<td>0.30</td>
<td>0.30</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>B6</td>
<td>0.34</td>
<td>0.35</td>
<td>0.34</td>
<td>0.30</td>
</tr>
<tr>
<td>B7</td>
<td>0.42</td>
<td>0.43</td>
<td>0.42</td>
<td>0.37</td>
</tr>
<tr>
<td>B8</td>
<td>0.51</td>
<td>0.50</td>
<td>0.52</td>
<td>0.44</td>
</tr>
<tr>
<td>C1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>C2</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>C3</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>C4</td>
<td>0.24</td>
<td>0.23</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td>C5</td>
<td>0.30</td>
<td>0.30</td>
<td>0.29</td>
<td>0.26</td>
</tr>
<tr>
<td>C6</td>
<td>0.36</td>
<td>0.35</td>
<td>0.36</td>
<td>0.31</td>
</tr>
<tr>
<td>C7</td>
<td>0.46</td>
<td>0.45</td>
<td>0.42</td>
<td>0.38</td>
</tr>
<tr>
<td>C8</td>
<td>0.58</td>
<td>0.52</td>
<td>0.50</td>
<td>0.41</td>
</tr>
<tr>
<td>D1</td>
<td>0.02*</td>
<td>0.02*</td>
<td>0.02*</td>
<td>0.02</td>
</tr>
<tr>
<td>D2</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>D3</td>
<td>0.16</td>
<td>0.16</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>D4</td>
<td>0.24</td>
<td>0.24</td>
<td>0.23</td>
<td>0.20</td>
</tr>
<tr>
<td>D5</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.25</td>
</tr>
<tr>
<td>D6</td>
<td>0.37</td>
<td>0.36</td>
<td>0.36</td>
<td>0.33</td>
</tr>
<tr>
<td>D7</td>
<td>0.41</td>
<td>0.41</td>
<td>0.42</td>
<td>0.37</td>
</tr>
<tr>
<td>D8</td>
<td>0.47</td>
<td>0.47</td>
<td>0.46</td>
<td>0.41</td>
</tr>
</tbody>
</table>