PAPER

Material-agnostic characterization of spatially offset Raman spectroscopy in turbid media *via* Monte Carlo simulations†

Zuriel Erikson Joven, [ID] [a] Piyush Raj[a] and Ishan Barman [ID] *[a,b,c]

Spatially offset Raman spectroscopy (SORS) is a transformative method for probing subsurface chemical compositions in turbid media. This systematic study of Monte Carlo simulations provides closed-form characterizations of key SORS parameters, such as the distribution of spatial origins of collected Raman photons and optimal SORS geometry to selectively interrogate a subsurface region of interest. These results are unified across an extensive range of material properties by multiplying spatial dimensions by the medium's effective attenuation coefficient, which can be calculated when the absorption and reduced scattering coefficients are known from the literature or experimentation. This method of spatial nondimensionalization is validated *via* goodness-of-fit analysis on the aggregate models and by training a subsurface sample localization model on a heterogeneous population of materials. The findings reported here advance the understanding of SORS phenomena while providing a quantitative and widely applicable foundation for designing and interpreting SORS experiments, facilitating its application in disciplines such as biomedical, materials science, and cultural heritage fields.

Introduction

Spatially offset Raman spectroscopy (SORS) is a powerful technique for interrogating the subsurface chemical makeup of turbid media by measuring the shifts in frequency of scattered light.[1,2] Unlike conventional Raman spectroscopy, which primarily captures surface information, SORS utilizes spatial offsets between the excitation laser and collection optics to probe deeper layers in a manner similar to diffuse reflectance spectroscopy.[3–5] Such noninvasive and nondestructive sensing capabilities find applications in diverse fields, including medical diagnostics, materials science, and cultural heritage.[6–16]

As the use of SORS grows, there is an unmet need for widely applicable protocols to optimize experimental parameters and quantify results. The probability of spontaneous Raman scattering is on the order of 10^{-8}, and the Raman signal in SORS decays exponentially with increasing spatial offset.[17–19] The scarcity of collected Raman photons necessitates the deliberate placement of the source and detector probes to maximize the Raman signal recorded from specific subsurface regions, as indefinitely increasing laser power risks irradiation damage to the sample and feasible signal acquisition time is constrained based on the specific application.[20–24] Furthermore, while it is well established that the likelihood of a detected Raman photon originating from a specific layer generally decreases with depth (although not always monotonically), no widely applicable quantitative models describe the depth distribution of the spatial origins of collected Raman photons.[25–27] Such models are necessary for interpreting collected Raman spectra for the localization, not just detection, of subsurface Raman scatterers. Localization abilities are crucial in various SORS applications, such as tumor margin estimation, cartilage degradation assessment, and painted layer analysis.[28–31]

Prior works demonstrate progress towards the aforementioned goals but are often tailored towards a singular SORS application or are otherwise too contrived to be universally applicable. For example, Keller *et al.* presented histograms of Raman photon generation depth for various spatial offsets obtained from Monte Carlo simulations of SORS in a two-layered breast cancer tissue model.[32] Similarly, Zhang *et al.* presented Monte Carlo reconstructions of SORS spectra in human skin tissue and the relative contributions of each

[a]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA. E-mail: ibarman@jhu.edu
[b]Department of Oncology, Johns Hopkins University, Baltimore, Maryland 21287, USA
[c]The Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University, School of Medicine, Baltimore, Maryland 21205, USA
†Electronic supplementary information (ESI) available: Raw data, Monte Carlo code, additional Raman photon signal-to-noise ratio plots, and visualizations of collected Raman intensity distributions. See DOI: https://doi.org/10.1039/d4an01044b

tissue layer.[33] Studies of this type provide results that can guide other researchers in highly similar SORS applications but cannot be generalized to other applications with different material properties or non-planar geometry.[33–35] Mosca *et al.* reported that the transport length of the material defines the spatial scale of SORS phenomena and that multiplying the spatial offset by the reduced scattering coefficient, μ'_s, provides a nondimensional parameter that uniquely defines the collected Raman signal strength and the nondimensionalized 90th percentile probed depth *via* Monte Carlo simulations on non-absorbing samples with varying degrees of scattering.[25] Spatial nondimensionalization is an effective method for generalizing reported results to a wide range of materials; these results, however, assume zero absorption and stratified layers, limiting their applicability. Kotturi *et al.* analyzes surface-enhanced SORS detection capabilities for disc-shaped implants, a step towards applicability *in vivo* where detection targets are not infinitely wide planes.[36]

In this work, the concept of spatial nondimensionalization is extended to absorbing materials by multiplying spatial variables by the material's effective attenuation coefficient, μ_{eff}. This paper uses this new variation of spatial nondimensionalization to present, to the best of the authors' knowledge, the first closed-form expressions for key SORS experimental parameters as functions of the nondimensionalized spatial offset. Firstly, this study provides quantitative measures of location and dispersion for the depth distribution of the spatial origins of collected Raman photons for a given nondimensionalized spatial offset. One can calculate μ_{eff} from a material's absorption and reduced scattering coefficients, which can be determined from literature or *via* experimentation, multiply by the chosen spatial offset, and use the corresponding results presented here to quantitatively interpret the collected Raman spectra and estimate the depth of the Raman scatterers.[36–39] One could also use these results to choose an optimal spatial offset to maximize the relative Raman signal collected from a specific layer. This study then extends this optimization concept beyond layered geometry to subsurface spherical samples, providing optimal placement of the excitation laser and detection probe as a function of the sample's nondimensionalized diameter and nondimensionalized depth. Such optimization enables the chemical interrogation of clinically relevant spherical inclusions in subsurface layers, such as tumors and microcalcifications.[29,36–39] Finally, the power of spatial nondimensionalization is demonstrated by training a model to estimate the location of a subsurface spherical sample in a variety of turbid media from SORS measurements and evaluating its performance in materials outside of the training dataset.

Methods

Monte Carlo simulations

Numerical simulations of photon propagation in turbid media were performed using MCmatlab 4.4.5, a library in MATLAB (Mathworks) for Monte Carlo modeling of light transport in tissue.[40,58] The MATLAB code used in this study is freely accessible on GitHub (**https://github.com/zurieljoven/SORS-Simulations/**). These simulations were performed in homogeneous media assuming constant material properties. This assumption is justified by the relatively unchanging properties within the higher wavelength (700 nm onwards) regions commonly used in Raman spectroscopy in biological tissue.[18,20,41] Thirty-six materials were modeled by pairing a scattering coefficient (μ_s = 10, 20, 40, 60, 80, 100 cm^{-1}) with an absorption coefficient (μ_a = 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 cm^{-1}); all materials had a scattering anisotropy factor of g = 0.9.

As illustrated in Fig. 1, the distribution of spatial origins of collected Raman photons in a typical two-point SORS setup was obtained for each material and spatial offset in a three-step process modified from Wang *et al.*[42] First, the normalized fluence rate (NFR) distribution $\phi(x, y, z)$ was calculated for a Gaussian beam centered on the top of a 10 × 5 × 5 cm cuboid of uniform material, where (x, y, z) represent local position relative to the incident spot, with z denoting depth. The NFR reflects the likelihood of Raman scattering, assuming the probability of Raman photon generation is proportional to NFR. This assumption of linearity is used in previously published simulation studies to improve computational efficiency.[19,25,26,33,34,34,42] The 36 NFR distributions were calculated using MCmatlab simulations, with 10^8 photons launched from the incident laser and a voxel size of 0.2 × 0.2 × 0.2 mm.

Second, the escape functions $E(r_{detector}, \theta_{detector}, z_{detector})$ for each material, *i.e.*, the probability of a Raman photon being detected after generation at a given location relative to the detector fiber, was calculated. Because the simulated media are homogeneous, the distribution of escape function values is axisymmetric about the detector fiber. The escape function values $E(r_{detector}, z_{detector})$ were obtained by placing isotropic Raman annular sources of radii r = 0.1, 0.3, ..., 2.1 cm, radial thickness Δr = 0.2 cm, depths z = 0.05, 0.15, ..., 4.95 cm, and vertical thickness Δz = 0.1 cm, and dividing the number of collected photons by the number of launched photons.
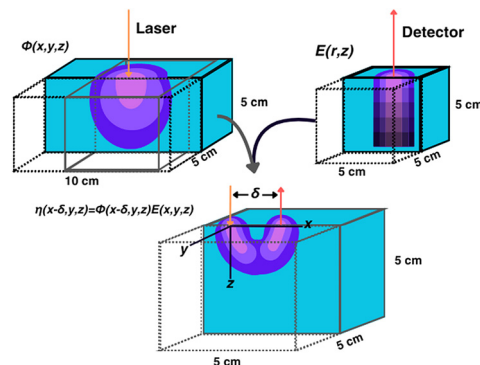

Fig. 1 Visualization of the calculation of the spatial distribution of collected Raman photon generation. Cyan represents the turbid media; purple represents magnitude of ϕ, E, or η (brightness indicates larger values).

MCmatlab simulations were performed for each of the 36 materials by isotropically launching 10^6 photons from each annulus in a $5 \times 5 \times 5$ cm cuboid of homogeneous media and a voxel size of $1 \times 1 \times 1$ mm. If no photons reached the detector, the simulation was repeated with 10^7 photons; if still no photons reached the detector, the corresponding escape function value was deemed negligible. The assumption of directional isotropy of Raman scattering is standard and validated by Keller et al.[26,34,43–45] The escape function values are then mapped to a Cartesian representation $E(x_{\text{detector}}, y_{\text{detector}}, z_{\text{detector}})$ with a voxel size of $0.2 \times 0.2 \times 0.2$ mm via bilinear interpolation in the $(r_{\text{detector}}, z_{\text{detector}})$ coordinate space.

Third, the distributions of the spatial origins of collected Raman photons $\eta(x, y, z)$ at a given spatial offset δ were calculated for each material via term-by-term multiplication of the NFR and escape function value distributions, offset by δ in the $+x$ direction:

$$\eta(x - \delta, y, z) = \phi(x - \delta, y, z)E(x, y, z).$$

η was calculated for a $5 \times 5 \times 5$ cm cuboid with the origin at the incident laser spot and a voxel size of $0.2 \times 0.2 \times 0.2$ mm and represents the strength of the collected Raman signal originating from (x, y, z).

Subsurface sample signal acquisition

For a spherical region of interest (ROI) of diameter D, centered at $(x_{\text{ROI}}, y_{\text{ROI}}, z_{\text{ROI}})$ within the media, the Raman photon signal-to-noise ratio $\text{SNR}_{\text{photon}}$ was defined as

$$\text{SNR}_{\text{photon}} = \frac{I_{\text{ROI}}}{I_{\overline{\text{ROI}}}},$$

where

$$I_{\text{ROI}} = \iiint_{(x,y,z)\in\text{ROI}} \eta(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

represents the signal (Raman signal from inside the ROI),

$$I_{\overline{\text{ROI}}} = \iiint_{(x,y,z)\notin\text{ROI}} \eta(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

represents the noise (Raman signal from outside the ROI), and

$$\text{ROI} = \left\{ (x,y,z) : \sqrt{(x - x_{\text{ROI}})^2 + (y - y_{\text{ROI}})^2 + (z - z_{\text{ROI}})^2} \leq D \right\}.$$

The $\text{SNR}_{\text{photon}}$ gives the ratio of collected Raman photons originating from within the ROI to those originating outside the ROI, assuming uniform turbidity and Raman cross-section across the ROI boundary. By the linearity of Raman scattering, the $\text{SNR}_{\text{photon}}$ for unequal Raman cross-section can be obtained via multiplication by the ratio of the Raman cross-sections. The contrast-to-noise ratio may also be orders of magnitude larger than $\text{SNR}_{\text{photon}}$ depending on the spectral shape of the Raman scatterers, e.g., if the Raman spectrum of the subsurface sample exhibits a tall and sharp peak where that of the surrounding media does not. This contrast-to-noise ratio can be deduced via linear combination of the constituent

Raman spectra, weighted by the corresponding I values multiplied by their respective Raman cross-sections. However, to maintain the broadest applicability of the reported results, this study was performed without consideration of specific Raman cross-section values or Raman spectra. For the same reason, as well as for computational efficiency, the deviations in turbidity effects caused by the subsurface sample were neglected, modeling the sample as the same material as the surrounding bulk media. Refractions and reflections due to refractive index mismatch across the ROI boundary were also neglected.

Spatial nondimensionalization

To generalize SORS phenomena across varying degrees of turbidity within a material, the following definition for the effective attenuation coefficient was adopted from prior literature:[46–49]

$$\mu_{\text{eff}} = \sqrt{3\mu_{\text{a}}(\mu_{\text{a}} + \mu_{\text{s}}')},$$

where

$$\mu_{\text{s}}' = \mu_{\text{s}}(1 - g)$$

is the reduced scattering coefficient, μ_{a} is the absorption coefficient, μ_{s} is the scattering coefficient, and g is the Henyey–Greenstein scattering anisotropy factor of the material. The calculated μ_{eff} values for the 36 materials used in this study are provided in Table 1.

Spatial nondimensionalization was performed by multiplying spatial dimensions by the relevant μ_{eff} value. In this paper, asterisks (*) denote nondimensionalized variables, and [–] denotes unitless space.

Results and discussion

Unified analysis of depth distribution of the spatial origins collected Raman photons

Fig. 2 illustrates the depth dependence of collected Raman intensity, $\eta(z)$, which resembles previously published trends.[25,26,33,34,42] Raman intensity generally decreases with depth, but this relationship is not necessarily monotonic nor uniform across varying values of μ_{s}, μ_{a}, or δ. In particular, at zero spatial offset, Raman intensity is greatest at $z = 0$ cm and appears to exponentially decay in deeper layers (Fig. 2A and B). However, at nonzero spatial offsets, intensity increases and

Table 1 Effective attenuation coefficients for each combination of absorption and scattering coefficient

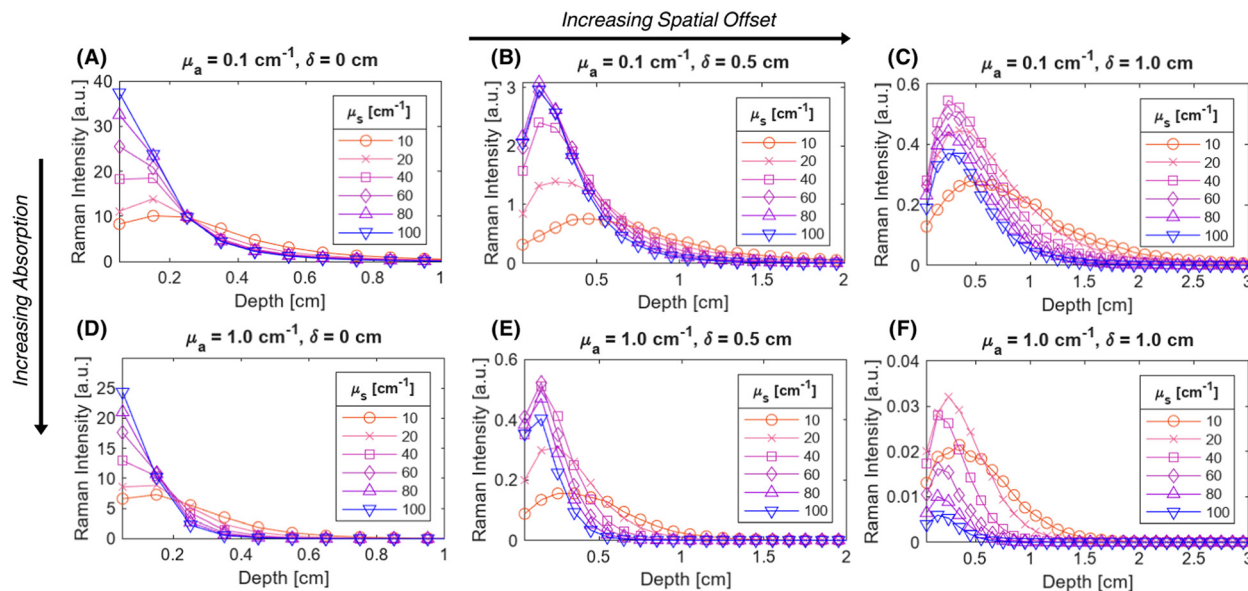| μ_{eff} [cm^{-1}]		μ_{a} [cm^{-1}]					
		0.1	0.2	0.4	0.6	0.8	1.0
μ_{s} [cm^{-1}]	10	0.58	0.85	1.30	1.70	2.08	2.45
	20	0.79	1.15	1.70	2.16	2.59	3.00
	40	1.11	1.59	2.30	2.88	3.39	3.87
	60	1.35	1.93	2.77	3.45	4.04	4.58
	80	1.56	2.22	3.17	3.93	4.60	5.20
	100	1.74	2.47	3.53	4.37	5.09	5.74

Fig. 2 Representative depth distributions of collected Raman intensity resemble left-truncated Gaussian distributions for various μ_s, μ_a, and δ. Depth distributions were obtained by integrating η over 1 mm thick layers, and are shown for $\mu_a = 0.1$ cm^{-1} (A–C) and 1.0 cm^{-1} (D–F), and for $\delta = 0$ cm (A and D), 0.5 cm (B and E), and 1.0 cm (C and F).

peaks at nonzero depth, followed by the same exponential decay, akin to a left-truncated Gaussian curve (Fig. 2C–F). The location of the peak of the distribution z_{peak}, as well as the degree of spread (which can be characterized by its full-width half maximum, or FWHM), vary with μ_s, μ_a, and δ. As scattering increases, the peak moves to shallower layers. When intensity is maximum at the surface, which occurs when $\delta \approx 0$, the negative slope is steeper for higher values of μ_s, implying that an extrapolated peak would occur at a more negative value; this supports the observation that μ_s and z_{peak} are inversely correlated. It is difficult to directly observe the effect of μ_s on the FWHM of the distribution for $\delta = 0$ cm, but the spread appears to decrease with increased scattering for nonzero spatial offset. Interestingly, while increasing μ_s (keeping μ_a, δ constant) appears to uniformly push the location of peak intensity to the left in Fig. 2, the value of peak intensity monotonically increases with μ_s for $\delta = 0$ cm (Fig. 2A and D) but increases then decreases for nonzero spatial offset (Fig. 2B–C, E and F). One explanation is that for low scattering media, the chance of sufficient scattering events occurring such that the photons, which were launched directly downwards, "turn around" towards the surface, will be lower than in highly scattering media. However, at much higher levels of scattering, there is a chance for photons to "get lost" and end up absorbed by the media or escape the surface at a different location than the detector. This phenomenon is especially prominent at higher values of μ_a and δ, since higher μ_a increases the chance of losing the photons to absorption, and higher δ requires the photons to travel a longer distance to reach the detector, which also increases the chance that absorption or irreversible escape at the surface will occur before detection (Fig. 2C, E and F). With increasing absorption, the entire dis-

tribution appears to shrink in both intensity and depth; for example, the maximum Raman intensity observed for $\mu_a = 1.0$ cm^{-1}, $\delta = 0.5$ cm was less than 20% of the maximum Raman intensity observed for $\mu_a = 0.1$ cm^{-1} at the same spatial offset (Fig. 2B and E). Additionally, for this spatial offset, the distributions appear to all settle around zero intensity between 1.5 and 2 cm for the lower absorption but settle between 0.5 and 1 cm for all high absorption cases except $\mu_s = 10$ cm^{-1} (Fig. 2B and E). This phenomenon, which is also observed for $\delta = 0$ and 1.0 cm, results in a decrease in the z_{peak} and FWHM with increasing absorption. Increased μ_a directly increases light attenuation, therefore intensity decays more rapidly at deeper layers, leading to a shallower z_{peak}, lower overall intensity values and a smaller spread in the distribution. Conversely, for fixed μ_s and μ_a, increasing spatial offset appears to stretch the distribution to deeper layers, increasing z_{peak} and FWHM. At nonzero spatial offsets, photons must also travel horizontally from source to detector to be collected; photons that travel close to the surface are likely to escape at the media–air interface before reaching the detector, but photons that travel at much deeper layers may never reach the detector before being absorbed. This results in a nonzero peak depth for Raman intensity that balances these counteracting phenomena (Fig. 2B, C, E and F). Additionally, at larger spatial offsets, each phenomenon has more chances to occur, but as photons escaping the surface is irreversible (whereas photons traveling deeper can scatter back to shallower layers), a greater weight of the collected Raman photon distribution shifts to the right tail of the depth distribution, resulting in a larger distribution spread.

In order to unify the interdependent effects of μ_s, μ_a and δ on the distribution parameters, the associated nondimensio-

nalized spatial offset $\delta^* = \delta \cdot \mu_{\text{eff}}$ was calculated for each distribution. The distributions were also nondimensionalized in depth by calculating $z^* = z \cdot \mu_{\text{eff}}$. Fig. 3 shows that distributions with similar δ^* values, even when the values of μ_s, μ_a and δ themselves differ, follow the same Raman intensity *versus* z^* distribution when normalized to unit peak; *i.e.*, δ^* uniquely determines the peak location and spread of the distribution. Distributions were grouped into bins of 0.1 [−] spatial offset, and (left-truncated) Gaussian curves were fit to the aggregate data from the normalized and nondimensionalized Raman intensity distributions (Fig. 3A–E.ii); good correspondence in the fitting is evidenced by all R^2 values being higher than 0.88 (Fig. 4A). Plotting the fit parameters for nondimensionalized peak location, z^*_{peak}, and nondimensionalized full-width half maximum, FWHM*, it is observable that both generally increase with δ^*, but at a decreasing rate (Fig. 4B–D). The same

can be said for the nondimensionalized "probed depth" quantity $z^*_{\text{probed}} = z^*_{\text{peak}} + 2\sigma^*$ (where σ^* is the nondimensionalized standard deviation of the fit distribution), a proxy for the upper limit of the distribution commonly used for Gaussian curves; only a minuscule (small single-digit) percentage of collected Raman photons originate from this depth or deeper, though the exact number is variable due to the left-truncation of the distributions at $z^* = 0$.[50–52] These relationships can be characterized by closed-form equations of the form $y = c - ae^{-bx}$ (where x represents the nondimensionalized spatial offset and y represents a Gaussian curve parameter) *via* a second least-squares regression, with $R^2 \geq 0.96$ (Table 2). These unified characterizations at different δ^* demonstrate that spatial offset determines the general shape of the intensity *versus* depth distribution, while turbidity determines the scaling, resulting in multiple SORS setups behaving identically
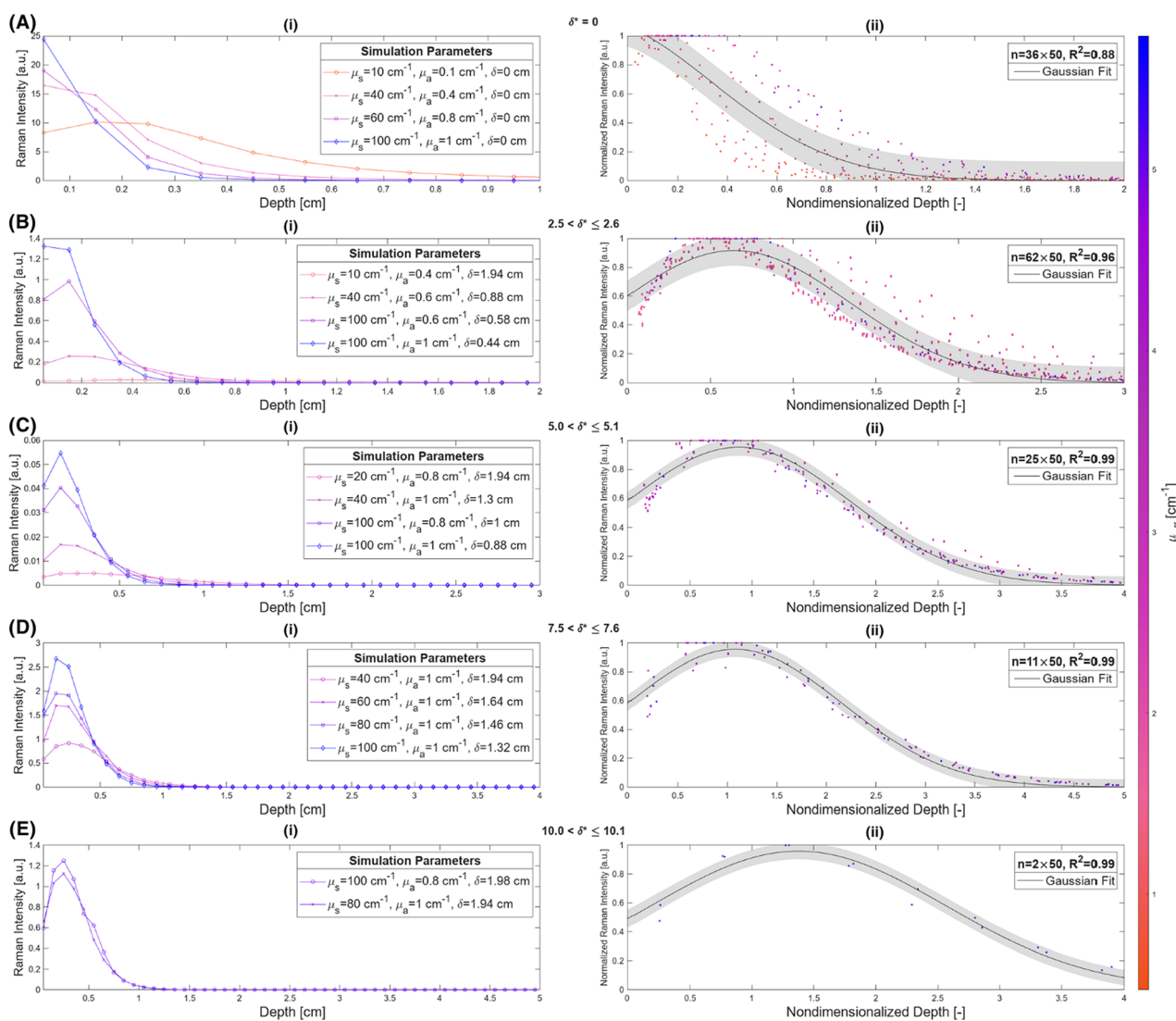

Fig. 3 Nondimensionalized spatial offset uniquely determines key parameters for nondimensionalized depth distributions of collected Raman intensity. (A–E) Representative raw data (i) and Gaussian fitting to aggregate normalized Raman intensity *versus* nondimensionalized depth data (ii) for various nondimensionalized spatial offset values. All shaded areas represent 95% confidence prediction intervals.
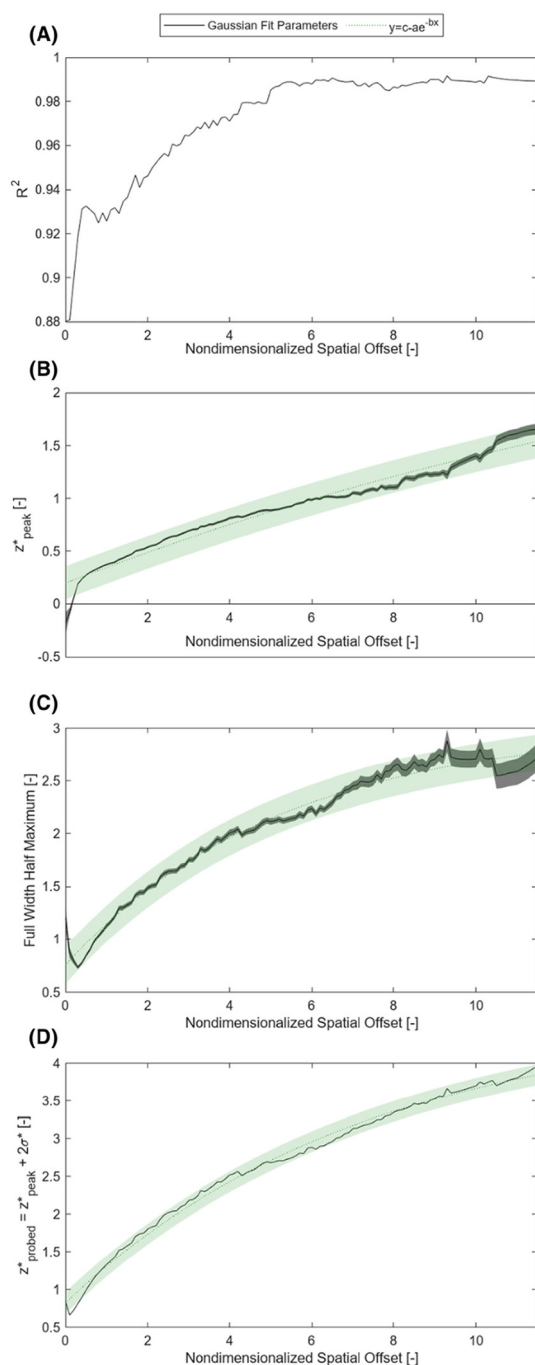
Fig. 4 Gaussian fit parameters (black) and overlaid $y = c - ae^{-bx}$ regression curves (green); values for a, b, and c presented in Table 2. All shaded areas represent 95% confidence prediction intervals.

Table 2 Gaussian fit parameters as functions of nondimensionalized spatial offset. Functions are of the form $y = c - ae^{-bx}$, where $x = \delta^*$. Valid domains were obtained from visual convergence of data to the functions in Fig. 4B and C

$y(x)$	a	b	c	R^2	Valid domain
z^*_{peak}	3.087	0.04943	3.287	0.96	$0.5 \leq \delta^* \leq 10.5$
FWHM*	2.259	0.1902	3.017	0.98	$1.0 \leq \delta^* \leq 10.0$
z^*_{probed}	3.814	0.1366	4.639	0.99	$0.5 \leq \delta^* \leq 11.5$

lower (Fig. 4A), and larger divergence between the individual nondimensionalized and normalized data points and the aggregate fit is observed (Fig. 3A.ii). Also at small δ^*, $z^*_{\text{peak}} \leq 0$, resulting in distributions with peak intensity at the surface and monotonic decay with depth; more negative z^*_{peak} values result in steeper negative slopes at $z^* = 0$ (Fig. 4B). FWHM* (which is a scalar multiple of σ^*) initially decreases before increasing in the form of $y = c - ae^{-bx}$ (Fig. 4C); the relative minimum appears to occur at the same δ^* where the z^*_{peak} versus δ^* slope starts to decrease (Fig. 4B). A similar trend is observed in the nondimensionalized probed depth, though the dip results from counterplay between the increasing z^*_{peak} and decreasing FWHM* at small δ^* (Fig. 4D). The growth equations for all three parameters imply the existence of asymptotic limits (Table 2). The significance of such upper bounds is particularly profound for z^*_{probed}, as this implies that, for a fixed μ_{eff}, there is a hard limit on the deepest probed depth attainable by arbitrarily increasing spatial offset. Plus, reaching such a high value for probed depth requires such a large spatial offset that the overall collected Raman signal may be extraordinarily low, further supporting this conclusion that SORS cannot probe infinitely deep layers due to both theoretical and practical limitations.

Optimal source-detector placement assuming *a priori* knowledge of subsurface sample size and depth

The use of spatial nondimensionalization to enable a unified analysis of the spatial distribution of collected Raman photons was extended from stratified media to spherical samples. Spherical samples of diameter $D = 0.1, 0.25, 0.5, 0.75, 1.0$ cm were placed at depths $z = 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ cm. The source and detector probes were placed at spatial offsets $\delta = 0, 0.02,..., 2.0$ cm such that the plane connecting the two probes and the center of the sphere contained the basis vector in the z direction, with various displacements $r = 0, 0.02, \cdots, \delta$ cm between the laser and the center of the sphere (Fig. 5). $\text{SNR}_{\text{photon}}$ was plotted against $\zeta = r/\delta$ for four different materials at $z = 0.5, 1.5$ cm, at $\delta = 0.02, 0.5, 1.0$ cm (Fig. 6A–B.i-ii and Fig. S1A–B.i-ii in the ESI†). These plots show that for each material and fixed depth, different spatial offsets yield a different optimal ζ that maximizes $\text{SNR}_{\text{photon}}$. For example, in the majority of plotted cases, $\text{SNR}_{\text{photon}}$ appears to be maximum when $\zeta \approx 0.5$, *i.e.*, the source and detector probes are placed such that the spherical sample is somewhere in the middle between them.
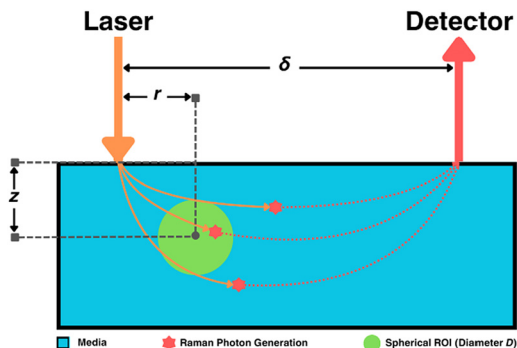
(with respect to the depth dependence on collected Raman intensity) as long as $\delta \cdot \mu_{\text{eff}}$ are equal. The utility of such equations is their ability to guide SORS setups for a given μ_{eff}, as the nondimensionalized parameters can be divided by μ_{eff} and a specific spatial offset can be chosen to obtain a specific peak layer (or other distribution parameter).

These formulas show good correspondence with the plotted values, except for small δ^* (Fig. 4B–D). At smaller δ^*, R^2 is

Fig. 5 Visualization of Raman signal acquisition from subsurface spherical samples.

The natural collected photon pathways from the laser to detector extend to deeper layers in the middle, as represented in Fig. 5 and corroborated by results in the previous subsection. However, in the cases of $\mu_s = 10$ cm^{-1}, $z = 0.5$ cm, $\delta = 1.0$ cm, SNR$_{photon}$ is actually minimum around $\zeta \approx 0.5$, $i.e.$, it is more optimal to place either the source or the detector nearly on top of the sample (Fig. 6A.i and Fig. S1A.i†). In these cases, the natural photon pathways extend to layers deeper than the depth of the spherical sample. The peak around $\zeta \approx 0$ is due to higher ϕ closer to the laser, though this is counteracted by a lower E further from the detector; conversely, the secondary peak around $\zeta \approx 1$ is due to higher E closer to the detector, though ϕ is lower further from the laser. This phenomenon is illustrated in the ESI (Fig. S2†).

Ultimately, different choices for δ result in a different maximum attainable SNR$_{photon}$ value and choice of ζ that maximizes SNR$_{photon}$. This dependence of max$_\zeta$(SNR$_{photon}$) on δ for the different materials and sample depths are plotted in Fig. 5A–D.iii–iv, along with the value of δ and corresponding ζ that obtains the absolute maximum attainable SNR$_{photon}$ for a given material and sample depth:

$$\delta_{opt}, \zeta_{opt} = \text{argmax}_{\delta,\zeta}(\text{SNR}_{photon})$$

For each material, δ_{opt} is smaller for $z = 0.5$ cm than for $z = 1.5$ cm. The dependencies of the absolute maximum attainable SNR$_{photon}$ and values of δ_{opt} and ζ_{opt} on the diameter and depth of the spherical samples are plotted in Fig. 6A, B.v–vii and Fig. S1A, B.v–vii.† Note that the "ridge" in the max$_{\delta,\zeta}$(SNR$_{photon}$) plots at high D and low z is due to the fact that when $z < D$, part of the spherical ROI is above the surface, effectively decreasing the volume of the subsurface sample (Fig. 6A, B.v and Fig. S1A, B.v†). Otherwise, these plots show a relatively log-linear relationship between max$_{\delta,\zeta}$(SNR$_{photon}$) and sample size and depth for all materials. Additionally, δ_{opt} primarily exhibits exponential or parabolic growth in z for all materials with minimal variation in D, though these plots are truncated at $\delta_{opt} = 2$ cm, the maximum spatial offset for which η was calculated (Fig. 6A, B.vi and Fig. S1A, B.vi†). ζ_{opt} appears to grow primarily in z as well, settling around $\zeta_{opt} \approx 0.5$ (Fig. 6A, B.vii and Fig. S1A, B.vii†). Interestingly, these plots do not exhibit any values of $\zeta_{opt} \approx 1$, indicating that it is seemingly never optimal to place the detector probe above the subsurface sample instead of the laser. This implies that the
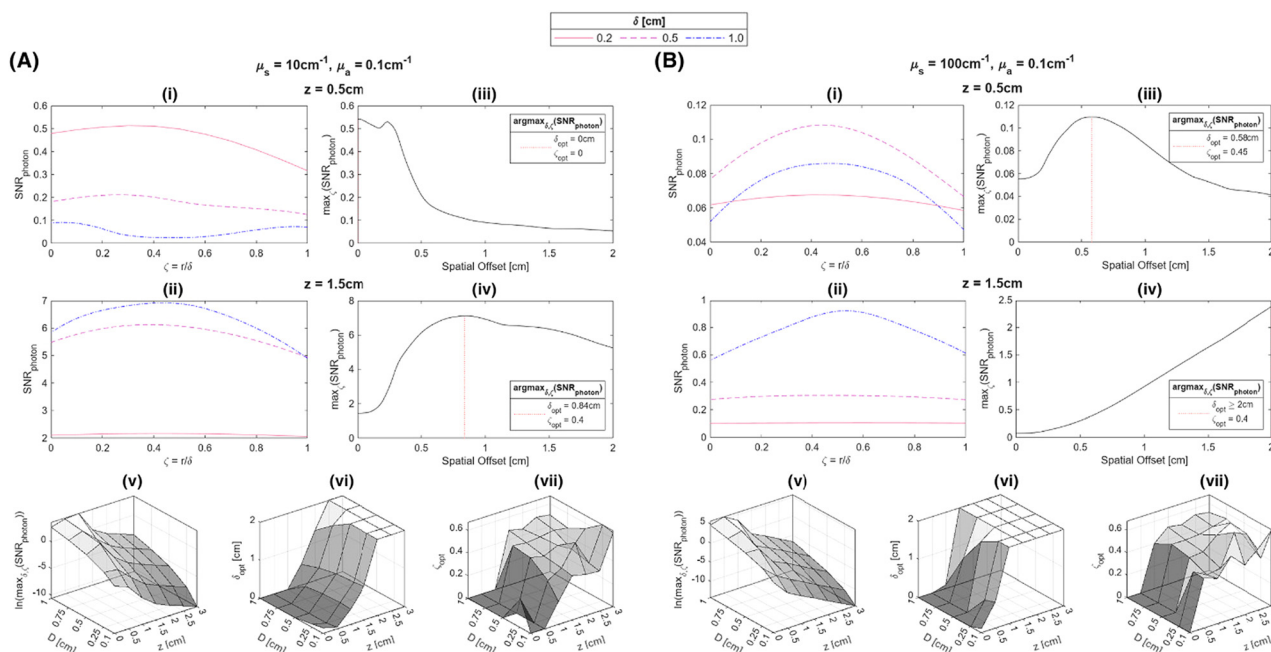

Fig. 6 Maximum attainable SNR$_{photon}$ values and corresponding δ_{opt} and ζ_{opt} values follow similar dependencies on D and z in different materials. Representative SNR$_{photon}$ $versus$ ζ plots at $z = 0.5$ cm (i) and 1.5 cm (ii) and $\delta = 0.2, 0.5, 1.0$ cm, max$_\zeta$(SNR$_{photon}$) $versus$ δ plots at $z = 0.5$ cm (iii) and 1.5 cm (iv), max$_{\delta,\zeta}$(SNR$_{photon}$) $versus$ (D, z) log-scale plots (v), δ_{opt} $versus$ (D, z) plots (vi), and ζ_{opt} $versus$ (D, z) plots (vii) for $\mu_s = 10$ cm^{-1}, $\mu_a = 0.1$ cm^{-1} (A) and $\mu_s = 100$ cm^{-1}, $\mu_a = 0.1$ cm^{-1} (B).

Raman photon detection efficiency near the detector is overshadowed by the laser beam depletion during photon migration.

In order to obtain generalized characterizations of optimal source-detector placement across varying degrees of turbidity, maximum attainable SNR_{photon} values and corresponding δ_{opt} and ζ_{opt} values were obtained for μ_s = 10, 20, 40, 60, 80, 100 cm^{-1}, μ_a = 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 cm^{-1}, D = 0.1, 1.0 cm, and z = 0.5, 1.0, 2.0 cm (n = 216 data points). The data were nondimensionalized in the same manner as the previous subsection. The dependencies of $\max_{\delta^*,\zeta}(SNR_{photon})$ and $\left(\delta_{opt}^*, \zeta_{opt}\right) = \arg\max_{\delta^*,\zeta}(SNR_{photon})$ on nondimensionalized sample diameter and nondimensionalized sample depth are plotted in Fig. 7; note that data points where δ_{opt} = 2 cm were excluded because the true value of δ_{opt} may be higher than 2 cm (n = 124 data points remain). Explicit formulas for ln $(\max_{\delta^*,\zeta}(SNR_{photon}))$ and δ_{opt}^* as functions of D^* and z^* were obtained via least-absolute-residuals regression and are plotted in Fig. 7A and B and summarized in Table 3. The R^2 values and residual plots showcase good correspondence with the aggregate data, save for some outliers at extremely high μ_{eff} values (Fig. 7D and E); the domain for which these functions are assumed to be valid was obtained as a convex hull of the data projected onto the (D^*, z^*) plane (shaded region in Fig. 7F). To the left of this region, the sample may protrude from the surface, and to the right of this region, δ_{opt} may exceed 2 cm. The ζ_{opt} values were unable to be fit as an explicit

function of D^* and z^*, but was fit to an explicit function of δ_{opt}^* via least-absolute-residuals regression and summarized in Table 3 (Fig. 7C and G). These correlations can be used to inform optimal SORS geometry if the sample diameter and depth are known a priori by multiplying these parameters by μ_{eff}, obtaining δ_{opt}^* and the corresponding ζ_{opt} from Table 3, and then dividing δ_{opt}^* by μ_{eff}. For example, computerized tomography (CT) or magnetic resonance imaging (MRI) scans can identify suspicious subsurface masses such as potential tumors, and the proposed SORS methodology can be subsequently applied for optimal chemical interrogation.[14,29,36-39,53,54]

Estimation of subsurface sample location from multi-detector probe measurements

The potential utility for these generalized characterizations for SORS applications was demonstrated via attempts to estimate the location of spherical subsurface samples using an array of detector fibers and a single source laser. Training was performed on an aggregate dataset spanning multiple materials, and testing was performed on materials not used for training; nondimensionalization enables the training of a widely applicable model on a heterogeneous dataset. One example use case is calibrating SORS measurements for in vivo clinical applications, as optical properties are variable between different patients.[41,55-57] Eighteen detector probes were placed surrounding a laser as illustrated in Fig. 8A and B. Spherical
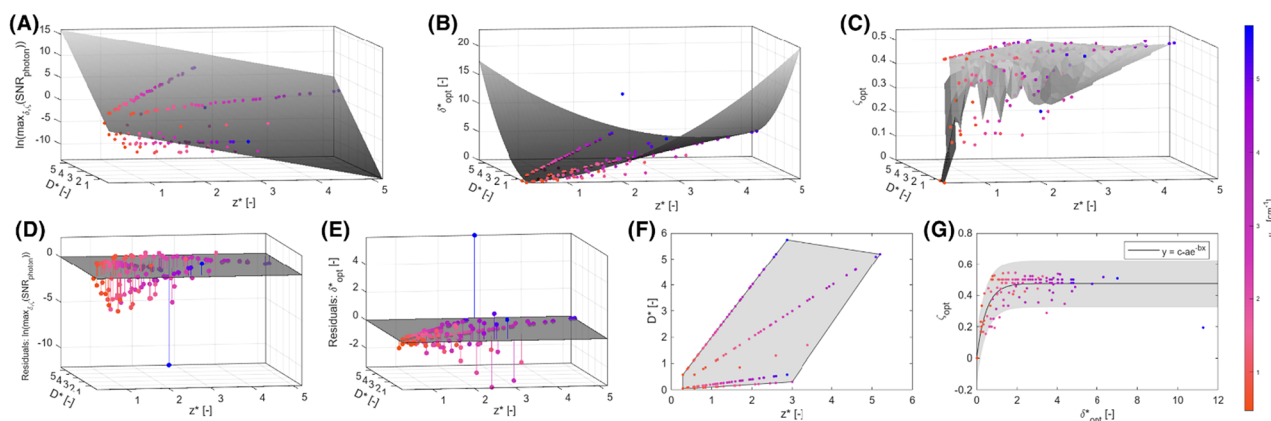

Fig. 7 Aggregate optimization data closely follow unified regression models. Raw data (scatter points) of $\max_{\delta^*,\zeta}(SNR_{photon})$ (A), δ_{opt}^* (B), and ζ_{opt} (C) versus (D^*, z^*) overlaid with regression models (A and B) or bilinear interpolation (C) in gray. Residuals plots (D and E) and valid domain (F) for the regression models. (G) Raw data of ζ_{opt} versus δ_{opt}^* overlaid with regression of the form $y = c - ae^{-bx}$ and 95% confidence prediction interval. Details on all regression models presented in Table 3.

Table 3 Functions for $\ln(\max_{\delta^*,\zeta}(SNR_{photon}))$ and $\left(\delta_{opt}^*, \zeta_{opt}\right) = \arg\max_{\delta^*,\zeta}(SNR_{photon})$ obtained via least-absolute-residuals regression (n = 124)

Function	R^2	Valid domain
$\ln(\max_{\delta^*,\zeta}(SNR_{photon})) = 1.563 + 3.169D^* - 2.367z^*$	0.99	See Fig. 7F
$\delta_{opt}^* = 6.112 \times 10^{-11} + 0.92D^* + \left(4.424 \times 10^{-11}\right)z^*$ $+ 0.4321(D^*)^2 - 1.296D^*z^* + 0.8642(z^*)^2$	0.95	See Fig. 7F
$\zeta_{opt} = 0.5001 - 0.5186e^{-2.169\delta_{opt}^*}$	0.97	$0 < \delta_{opt}^* \leq 7$

samples of diameter $D = 1.0$ cm were placed at random locations $0.1 \le \sqrt{x^2 + y^2} \le 1.0$ cm and $0.25 \le z \le 3.0$ cm, and the 18 signals I_{ROI} from each detector were recorded for each sample; 150 samples were recorded for each of the 36 materials. The 600 data points from the ($\mu_\mathrm{s} = 20$ cm^{-1}, $\mu_\mathrm{a} = 0.2$ cm^{-1}), ($\mu_\mathrm{s} = 20$ cm^{-1}, $\mu_\mathrm{a} = 0.8$ cm^{-1}), ($\mu_\mathrm{s} = 80$ cm^{-1}, $\mu_\mathrm{a} = 0.2$ cm^{-1}), and ($\mu_\mathrm{s} = 80$ cm^{-1}, $\mu_\mathrm{a} = 0.8$ cm^{-1}) materials were reserved as a hold-out dataset for validation. The other 4800 data points were used for the following partial-least-squares (PLS) regression:

$$X_{\mathrm{train}} = \begin{bmatrix} I_{\mathrm{ROI},1,1} & \cdots & I_{\mathrm{ROI},1,18} & \ln I_{\mathrm{ROI},1,1} & \cdots & \ln I_{\mathrm{ROI},1,18} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ I_{\mathrm{ROI},4800,1} & \cdots & I_{\mathrm{ROI},4800,18} & \ln I_{\mathrm{ROI},4800,1} & \cdots & \ln I_{\mathrm{ROI},4800,18} \end{bmatrix} \in \mathbb{R}^{4800 \times 36}$$

$$Y_{\mathrm{train,obs}} = \begin{bmatrix} x_1 \cdot \mu_{\mathrm{eff},1} & y_1 \cdot \mu_{\mathrm{eff},1} & z_1 \cdot \mu_{\mathrm{eff},1} \\ \vdots & \vdots & \vdots \\ x_{4800} \cdot \mu_{\mathrm{eff},4800} & y_{4800} \cdot \mu_{\mathrm{eff},4800} & z_{4800} \cdot \mu_{\mathrm{eff},4800} \end{bmatrix} \in \mathbb{R}^{4800 \times 3}$$

$$Y_{\mathrm{train,pred}} = [1 X_{\mathrm{train}}]\beta; 1 \in \mathbb{R}^{4800}, \beta \in \mathbb{R}^{37 \times 3}$$

The matrices X and Y are designed to capture potential log-linear relationships similar to those observed in Fig. 6A. PLS was performed with 37 components to estimate β. The resultant model was used to predict subsurface sample locations from the Raman signals in the 600 validation data points:

$$X_{\mathrm{val}} = \begin{bmatrix} I_{\mathrm{ROI},1,1} & \cdots & I_{\mathrm{ROI},1,18} & \ln I_{\mathrm{ROI},1,1} & \cdots & \ln I_{\mathrm{ROI},1,18} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ I_{\mathrm{ROI},600,1} & \cdots & I_{\mathrm{ROI},600,18} & \ln I_{\mathrm{ROI},600,1} & \cdots & \ln I_{\mathrm{ROI},600,18} \end{bmatrix} \in \mathbb{R}^{600 \times 36}$$

$$Y_{\mathrm{val,obs}} = \begin{bmatrix} x_1 \cdot \mu_{\mathrm{eff},1} & y_1 \cdot \mu_{\mathrm{eff},1} & z_1 \cdot \mu_{\mathrm{eff},1} \\ \vdots & \vdots & \vdots \\ x_{600} \cdot \mu_{\mathrm{eff},600} & y_{600} \cdot \mu_{\mathrm{eff},600} & z_{600} \cdot \mu_{\mathrm{eff},600} \end{bmatrix} \in \mathbb{R}^{600 \times 3}$$

$$Y_{\mathrm{val,pred}} \approx [1 X_{\mathrm{val}}]\beta; 1 \in \mathbb{R}^{600}$$

Fig. 8C–I summarizes the predictive accuracy of this regression model. The predicted *versus* observed values closely follow the identity line for both the training (Fig. 8C–E) and validation (Fig. 8F–H) datasets, though the root-mean-squared-error (RMSE) is consistently lower for the validation dataset. This is likely simply due to the smaller number of materials in the validation dataset compared to the number used for training, as a single unified model will struggle to account for all edge cases. In particular, a number of low μ_{eff} value data points appear far above the identity line in shallow depths in the training dataset, and follows a shape reminiscent of a curved checkmark (Fig. 8E). A similar (but less prominent) trend is observed for the validation dataset (Fig. 8H). Most localization attempts in both datasets have a mean error (Euclidean distance) of 0.2–0.4 cm between the predicted and observed sample locations. The error is highest for low values of scattering and absorption coefficients, which corroborates the observations in Fig. 8E and H. One potential explanation is that the linear model captures monotonic relationships, but not reversals of this trends. Such reversals have been observed,

for example, in Fig. 5A–D.v, where SNR$_{\mathrm{photon}}$ first increases before decreasing with depth, and in Fig. 2B, C, E and F, where Raman intensity first increases before decreasing with scattering. As a result, the model, trained on aggregate data from a heterogeneous population of materials, successfully captures the dominant inverse relationship between Raman intensity and nondimensionalized depth, but erroneously predicts that lower Raman intensity for lower scattering values must be due to deeper buried samples. Nevertheless, while the predicted *versus* observed plots correlate nicely with the identity line, the distance error in the validation dataset being around 20% of the sample diameter demonstrates significant room for improvement. Optimization of regression parameters or the use of nonlinear models such as neural networks may prove more suitable for these and more complex localization tasks, and further exploration is needed to evaluate the potential translation to *in vivo* analysis.

Limitations

While a number of simplifying assumptions were made to enable generalization across a wide range of materials, these same assumptions may limit the applicability of the reported results. Firstly, the assumption that the bulk media can be modeled with a single set of optical properties, even when subsurface samples are present, may not accurately reflect the heterogeneity of samples such as human skin.[33,41,42] Additionally, the presented results only apply for the tested range of optical properties, and are further limited in their valid domains (Tables 2 and 3). The assumption that Raman scattering is directly proportional to fluence rate allows for the optimization methods in this paper to result in the same optimal experimental parameters to maximize true Raman signal and contrast-to-noise ratios, but can be problematic if the Raman cross-section is exceedingly low. Neglecting specific Raman cross-sections and spectral shapes also breaks the applicability
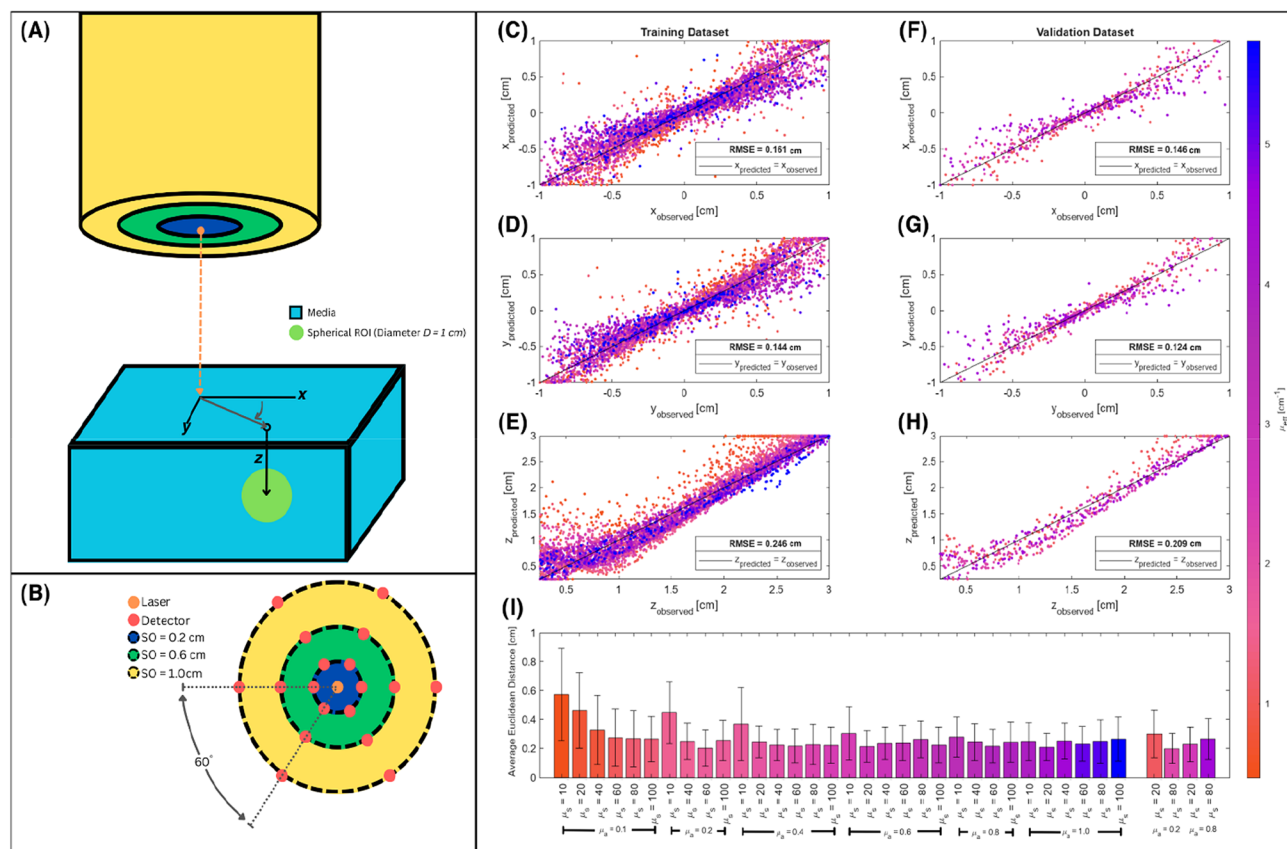
Fig. 8 Spatial nondimensionalization enables successful training of a regression model to locate subsurface samples from collected Raman signals. (A and B) Visualization of simulated probe and setup. Predicted *versus* observed Cartesian coordinates of subsurface samples for training (D and E) and validation (F−H) data. (I) Mean and standard deviation of Euclidean distance between predicted and observed sample locations for training (left) and validation (right) data.

of spatial nondimensionalization for training models on SORS measurements on a heterogeneous population if the Raman signatures vary between samples. This irrespectivity also limited the characterizations of the depth distributions to descriptors of peak depth and distribution spread, as descriptors of the Raman intensity values in arbitrary units lack physical utility without accounting for Raman cross-section and spectra. Also, these calculations for η assume a semi-infinite medium; η was only calculated for a finite domain, leading to potential inaccuracies arising from an inability to capture nonzero values at large values of x, y, or z. Investigation of such edge effects revealed that voxels touching the border of the $5 \times 5 \times 5$ cm (excluding the top surface) accounted for less than 0.25% of collected Raman photons in each calculated distribution (and accounted for 0% in nearly half of all cases). Hence, these effects were safely neglected. Finally, while the use of Monte Carlo simulations enables the acquisition of a large and standardized dataset for systematic analysis, experimental validation and empirical modeling should be performed in future works, noting whether significant discrepancies arise between the characterizations derived from simulated *versus* physical data.

Conclusions

This thorough and systematic Monte Carlo simulation study advances progress towards a quantitative and universal protocol for optimizing and interpreting SORS experiments. In particular, closed-form expressions for the measures of location, dispersion, and upper limits for the depth distribution of collected Raman intensity, as well as for optimal source and detector probe placement, are reported as functions of spatial parameters multiplied by the bulk medium's effective attenuation coefficient. Such results facilitate the selective interrogation of subsurface regions of interest, as well as estimation of the location of subsurface Raman scatterers. The power of spatial nondimensionalization to enable unified analyses was demonstrated by the high R^2 values of regression models used to derive the closed-form expressions, as well as successful localization of subsurface samples in materials outside the training dataset. These advancements contribute to a deeper understanding of SORS phenomena and facilitate its development for use in biomedical research, materials characterization, cultural heritage fields, and more.

Author contributions

Z.E.J., P.R., and I.B. designed the research. Z.E.J and P.R. performed the Monte Carlo simulations. Z.E.J. analyzed the data. Z.E.J. wrote the initial draft of the manuscript. All authors contributed to editing the manuscript. I.B. supervised the study and secured funding for the project.

Data availability

The MATLAB code and raw data used to generate the simulation results for this paper are freely available on GitHub (**https://github.com/zurieljoven/SORS-Simulations/**) and Figshare (SORS Excitation and Escape Function Distributions, 10.6084/m9.figshare.26354497).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 P. Matousek, I. P. Clark, E. R. C. Draper, M. D. Morris, A. E. Goodship, N. Everall, M. Towrie, W. F. Finney and A. W. Parker, Subsurface Probing in Diffusely Scattering Media Using Spatially Offset Raman Spectroscopy, *Appl. Spectrosc.*, 2005, **59**(4), 393–400.

2 N. Stone, K. Faulds, D. Graham and P. Matousek, Prospects of Deep Raman Spectroscopy for Noninvasive Detection of Conjugated Surface Enhanced Resonance Raman Scattering Nanoparticles Buried within 25 Mm of Mammalian Tissue, *Anal. Chem.*, 2010, **82**(10), 3969–3973.

3 S. Mosca, C. Conti, N. Stone and P. Matousek, Spatially Offset Raman Spectroscopy, *Nat. Rev. Methods Primers*, 2021, **1**(1), 1–16.

4 *Principles and Techniques of Diffuse–Reflectance Spectroscopy - Kortüm - 1963 - Angewandte Chemie International Edition in English - Wiley Online Library*.

5 C. Conti, M. Realini, C. Colombo and P. Matousek, Comparison of Key Modalities of Micro-Scale Spatially Offset Raman Spectroscopy, *Analyst*, 2015, **140**(24), 8127–8133.

6 M. Park, A. Somborn, D. Schlehuber, V. Keuter and G. Deerberg, Raman Spectroscopy in Crop Quality Assessment: Focusing on Sensing Secondary Metabolites: A Review, *Hortic. Res.*, 2023, **10**(5), uhad074.

7 U. J. Kim, S. Lee, H. Kim, Y. Roh, S. Han, H. Kim, Y. Park, S. Kim, M. J. Chung, H. Son and H. Choo, Drug Classification with a Spectral Barcode Obtained with a Smartphone Raman Spectrometer, *Nat. Commun.*, 2023, **14**(1), 5262.

8 C. Eliasson, N. A. Macleod and P. Matousek, Noninvasive Detection of Concealed Liquid Explosives Using Raman Spectroscopy, *Anal. Chem.*, 2007, **79**(21), 8185–8189.

9 R. J. Hopkins, S. H. Pelfrey and N. C. Shand, Short-Wave Infrared Excited Spatially Offset Raman Spectroscopy (SORS) for through-Barrier Detection, *Analyst*, 2012, **137**(19), 4408–4410.

10 H. Xie, R. Stevenson, N. Stone, A. Hernandez-Santana, K. Faulds and D. Graham, Tracking Bisphosphonates through a 20 Mm Thick Porcine Tissue by Using Surface-Enhanced Spatially Offset Raman Spectroscopy, *Angew. Chem., Int. Ed.*, 2012, **51**(34), 8509–8511.

11 B. Sharma, K. Ma, M. R. Glucksberg and R. P. Van Duyne, Seeing through Bone with Surface-Enhanced Spatially Offset Raman Spectroscopy, *J. Am. Chem. Soc.*, 2013, **135**(46), 17290–17293.

12 F. Nicolson, L. E. Jamieson, S. Mabbott, K. Plakas, N. C. Shand, M. R. Detty, D. Graham and K. Faulds, Through Tissue Imaging of a Live Breast Cancer Tumour Model Using Handheld Surface Enhanced Spatially Offset Resonance Raman Spectroscopy (SESORRS), *Chem. Sci.*, 2018, **9**(15), 3788–3792.

13 Z. Liu, M. Huang, Q. Zhu, J. Qin and M. S. Kim, A Packaged Food Internal Raman Signal Separation Method Based on Spatially Offset Raman Spectroscopy Combined with FastICA, *Spectrochim. Acta, Part A*, 2022, **275**, 121154.

14 F. Nicolson, M. F. Kircher, N. Stone and P. Matousek, Spatially Offset Raman Spectroscopy for Biomedical Applications, *Chem. Soc. Rev.*, 2021, **50**(1), 556–568.

15 P. Matousek, Inverse Spatially Offset Raman Spectroscopy for Deep Noninvasive Probing of Turbid Media, *Appl. Spectrosc.*, 2006, **60**(11), 1341–1347.

16 S. M. Asiala, N. C. Shand, K. Faulds and D. Graham, Surface-Enhanced, Spatially Offset Raman Spectroscopy (SESORS) in Tissue Analogues, *ACS Appl. Mater. Interfaces*, 2017, **9**(30), 25488–25494.

17 R. R. Jones, D. C. Hooper, L. Zhang, D. Wolverson and V. K. Valev, Raman Techniques: Fundamentals and Frontiers, *Nanoscale Res. Lett.*, 2019, **14**, 231.

18 J. R. Durig, Practical Raman Spectroscopy, in *TrAC Trends in Analytical Chemistry*, 1990, vol. 9, p. IX.

19 S. Mosca, P. Dey, M. Salimi, B. Gardner, F. Palombo, N. Stone and P. Matousek, Estimating the Reduced Scattering Coefficient of Turbid Media Using Spatially Offset Raman Spectroscopy, *Anal. Chem.*, 2021, **93**(7), 3386–3392.

20 F. Angelini and F. Colao, Optimization of Laser Wavelength, Power and Pulse Duration for Eye-Safe Raman Spectroscopy, *JEOS-RP*, 2019, **15**(1), 2.

21 D. Pestov, R. K. Murawski, G. O. Ariunbold, X. Wang, M. Zhi, A. V. Sokolov, V. A. Sautenkov, Y. V. Rostovtsev,

A. Dogariu, Y. Huang and M. O. Scully, Optimizing the Laser-Pulse Configuration for Coherent Raman Spectroscopy, *Science*, 2007, **316**(5822), 265–268.

22 Y. Wang and R. L. McCreery, Evaluation of a Diode Laser/Charge Coupled Device Spectrometer for near-Infrared Raman Spectroscopy, *Anal. Chem.*, 1989, **61**(23), 2647–2651.

23 G. Insero, F. Fusi and G. Romano, The Safe Use of Lasers in Biomedicine: Principles of Laser-Matter Interaction, *J. Public Health Res.*, 2023, **12**(3), 22799036231187077.

24 Y. Zhang, R. Chen, F. Liu, P. Miao, L. Lin and J. Ye, In Vivo Surface-Enhanced Transmission Raman Spectroscopy under Maximum Permissible Exposure: Toward Photosafe Detection of Deep-Seated Tumors, *Small Methods*, 2023, **7**(2), 2201334.

25 S. Mosca, P. Dey, M. Salimi, B. Gardner, F. Palombo, N. Stone and P. Matousek, Spatially Offset Raman Spectroscopy—How Deep?, *Anal. Chem.*, 2021, **93**(17), 6755–6762.

26 M. D. Keller, R. H. Wilson, M.-A. Mycek and A. Mahadevan-Jansen, Monte Carlo Model of Spatially Offset Raman Spectroscopy for Breast Tumor Margin Analysis, *Appl. Spectrosc.*, 2010, **64**(6), 607–614.

27 M. E. Berry, S. M. McCabe, N. C. Shand, D. Graham and K. Faulds, Depth Prediction of Nanotags in Tissue Using Surface Enhanced Spatially Offset Raman Scattering (SESORS), *Chem. Commun.*, 2022, **58**(11), 1756–1759.

28 P. Raj, L. Wu, C. Almeida, L. Conway, S. Tanwar, J. Middendorf and I. Barman, Shining Light on Osteoarthritis: Spatially Offset Raman Spectroscopy as a Window into Cartilage Health, *ACS Sens.*, 2024, **9**(7), 3794–3804.

29 M. D. Keller, E. Vargis, N. de Matos Granja, R. H. Wilson, M.-A. Mycek, M. C. Kelley and A. Mahadevan-Jansen, Development of a Spatially Offset Raman Spectroscopy Probe for Breast Tumor Surgical Margin Evaluation, *J. Biomed. Opt.*, 2011, **16**(7), 077006.

30 G. Thomas, T.-Q. Nguyen, I. J. Pence, B. Caldwell, M. E. O'Connor, J. Giltnane, M. E. Sanders, A. Grau, I. Meszoely, M. Hooks, M. C. Kelley and A. Mahadevan-Jansen, Evaluating Feasibility of an Automated 3-Dimensional Scanner Using Raman Spectroscopy for Intraoperative Breast Margin Assessment, *Sci. Rep.*, 2017, **7**(1), 13548.

31 C. Conti, C. Colombo, M. Realini and P. Matousek, Subsurface Analysis of Painted Sculptures and Plasters Using Micrometre-Scale Spatially Offset Raman Spectroscopy (Micro-SORS), *J. Raman Spectrosc.*, 2015, **46**(5), 476–482.

32 M. D. Keller, S. K. Majumder and A. Mahadevan-Jansen, Spatially Offset Raman Spectroscopy of Layered Soft Tissues, *Opt. Lett.*, 2009, **34**(7), 926–928.

33 Y.-H. Zhang, H.-Z. Zhu, Y.-J. Dong, J. Zeng, X.-P. Han, I. A. Bratchenko, F.-R. Zhang, S.-Y. Xu and S. Wang, Reconstructing in Vivo Spatially Offset Raman Spectroscopy of Human Skin Tissue Using a GPU-Accelerated Monte Carlo Platform, *Chin. Phys. B*, 2023, **32**(11), 118702.

34 P. Matousek, M. D. Morris, N. Everall, I. P. Clark, M. Towrie, E. Draper, A. Goodship and A. W. Parker, Numerical Simulations of Subsurface Probing in Diffusely Scattering Media Using Spatially Offset Raman Spectroscopy, *Appl. Spectrosc.*, 2005, **59**(12), 1485–1492.

35 K. Buckley, J. G. Kerns, A. W. Parker, A. E. Goodship and P. Matousek, Decomposition of in Vivo Spatially Offset Raman Spectroscopy Data Using Multivariate Analysis Techniques, *J. Raman Spectrosc.*, 2014, **45**(2), 188–192.

36 D. Kotturi, S. Paterson and M. McShane, Surface-Enhanced Spatially Offset Raman Spectroscopy in Tissue, *Biosensors*, 2024, **14**(2), 81.

37 P. Matousek and N. Stone, Development of Deep Subsurface Raman Spectroscopy for Medical Diagnosis and Disease Monitoring, *Chem. Soc. Rev.*, 2016, **45**(7), 1794–1802.

38 R. Vanna, C. Morasso, B. Marcinnò, F. Piccotti, E. Torti, D. Altamura, S. Albasini, M. Agozzino, L. Villani, L. Sorrentino, O. Bunk, F. Leporati, C. Giannini and F. Corsi, Raman Spectroscopy Reveals That Biochemical Composition of Breast Microcalcifications Correlates with Histopathologic Features, *Cancer Res.*, 2020, **80**(8), 1762–1772.

39 I. Barman, N. C. Dingari, A. Saha, S. McGee, L. H. Galindo, W. Liu, D. Plecha, N. Klein, R. R. Dasari and M. Fitzmaurice, Application of Raman Spectroscopy to Identify Microcalcifications and Underlying Breast Lesions at Stereotactic Core Needle Biopsy, *Cancer Res.*, 2013, **73**(11), 3206–3215.

40 D. Marti, R. N. N. Aasbjerg, P. E. E. Andersen and A. K. K. Hansen, MCmatlab: An Open-Source, User-Friendly, MATLAB-Integrated Three-Dimensional Monte Carlo Light Transport Solver with Heat Diffusion and Tissue Damage, *J. Biomed. Opt.*, 2018, **23**(12), 121622.

41 J. L. Sandell and T. C. Zhu, A Review of In-Vivo Optical Properties of Human Tissues and Its Impact on PDT, *J. Biophotonics*, 2011, **4**(11–12), 773–787.

42 S. Wang, J. Zhao, H. Lui, Q. He, J. Bai and H. Zeng, Monte Carlo Simulation of in Vivo Raman Spectral Measurements of Human Skin with a Multi-Layered Tissue Optical Model, *J. Biophotonics*, 2014, **7**(9), 703–712.

43 S. Sathyendranath and T. Platt, Analytic Model of Ocean Color, *Appl. Opt.*, 1997, **36**(12), 2620–2629.

44 W.-C. Shih, K. L. Bechtel and M. S. Feld, Intrinsic Raman Spectroscopy for Quantitative Biological Spectroscopy Part I: Theory and Simulations, *Opt. Express*, 2008, **16**(17), 12726.

45 A. M. K. Enejder, T.-W. Koo, J. Oh, M. Hunter, S. Sasic, M. S. Feld and G. L. Horowitz, Blood Analysis by Raman Spectroscopy, *Opt. Lett.*, 2002, **27**(22), 2004–2006.

46 A. M. Chiarelli, K. A. Low, E. L. Maclin, M. A. Fletcher, T. S. Kong, B. Zimmerman, C. H. Tan, B. P. Sutton, M. Fabiani and G. Gratton, The Optical Effective Attenuation Coefficient as an Informative Measure of Brain Health in Aging, *Photonics*, 2019, **6**(3), 79.

47 A. M. Chiarelli, E. L. Maclin, K. A. Low, S. Fantini, M. Fabiani and G. Gratton, Low-Resolution Mapping of the

Effective Attenuation Coefficient of the Human Head: A Multidistance Approach Applied to High-Density Optical Recordings, *Neurophotonics*, 2017, **4**(2), 021103.

48 S.-Å. Stark and S. Carlsson, A Phantom for the Measurement of Transfer Functions in Nuclear Medicine Image Processing, *Radiat. Prot. Dosim.*, 1993, **49**(1–3), 303–305.

49 S.-Å. Starck and S. Carlsson, The Determination of the Effective Attenuation Coefficient from Effective Organ Depth and Modulation Transfer Function in Gamma Camera Imaging, *Phys. Med. Biol.*, 1997, **42**(10), 1957–1964.

50 E. K. Harris and D. L. DeMets, Estimation of Normal Ranges and Cumulative Proportions by Transforming Observed Distributions to Gaussian Form, *Clin. Chem.*, 1972, **18**(7), 605–612.

51 *Business Statistics: Based on Schaum's Outline of Theory and Problems of Business Statistics, Third Edition*, ed. L. J. Kazmier, McGraw-Hill, New York, 2003.

52 D. Dowson and A. Wragg, Maximum-Entropy Distributions Having Prescribed First and Second Moments (Corresp.), *IEEE Trans. Inf. Theory*, 1973, **19**(5), 689–693.

53 M. F. Kircher and J. K. Willmann, Molecular Body Imaging: MR Imaging, CT, and US. Part I. Principles, *Radiology*, 2012, **263**(3), 633–643.

54 M. A. Pysz, S. S. Gambhir and J. K. Willmann, Molecular Imaging: Current Status and Emerging Strategies, *Clin. Radiol.*, 2010, **65**(7), 500–516.

55 M. N. Hannan, A. K. Sharma and T. M. Baran, Preliminary Measurements of Optical Properties in Human Abscess Cavities Prior to Methylene Blue Photodynamic Therapy, *Proc. SPIE-Int. Soc. Opt. Eng.*, 2023, **12359**, 123590A.

56 M. Späth, M. Rohde, D. Ni, F. Knieling, F. Stelzle, M. Schmidt, F. Klämpfl and M. Hohmann, The Influence of the Optical Properties on the Determination of Capillary Diameters, *Sci. Rep.*, 2022, **12**(1), 270.

57 R. Al-Halawani, M. Qassem and P. A. Kyriacou, Monte Carlo Simulation of the Effect of Melanin Concentration on Light-Tissue Interactions in Transmittance and Reflectance Finger Photoplethysmography, *Sci. Rep.*, 2024, **14**(1), 8145.

58 S. L. Jacques and L. Wang, Monte Carlo Modeling of Light Transport in Tissue, in *Optical-Thermal Response of Laser-Irradiated Tissue. Lasers, Photonics, and Electro-Optics*, Springer, Boston, MA, 1995, pp 73–100.