

# SCIENTIFIC REPORTS



OPEN

## Characterization of the Optical Properties of Turbid Media by Supervised Learning of Scattering Patterns

Iman Hassaninia<sup>1</sup>, Ramin Bostanabad<sup>2</sup>, Wei Chen<sup>2</sup> & Hooman Mohseni<sup>1</sup>

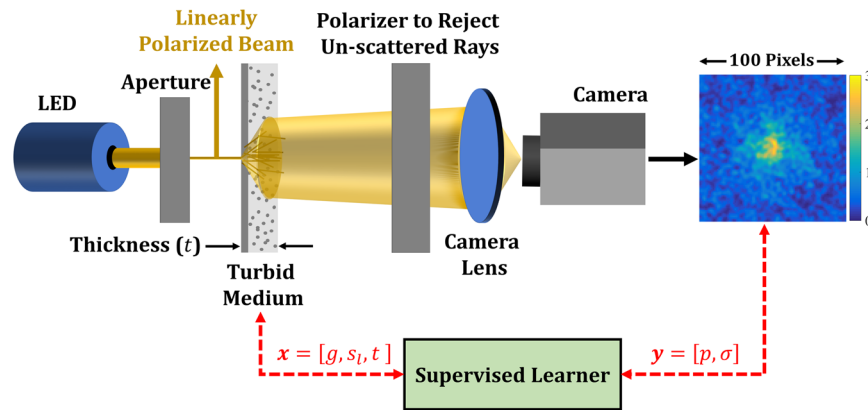
Fabricated tissue phantoms are instrumental in optical *in-vitro* investigations concerning cancer diagnosis, therapeutic applications, and drug efficacy tests. We present a simple non-invasive computational technique that, when coupled with experiments, has the potential for characterization of a wide range of biological tissues. The fundamental idea of our approach is to find a supervised learner that links the scattering pattern of a turbid sample to its thickness and scattering parameters. Once found, this supervised learner is employed in an inverse optimization problem for estimating the scattering parameters of a sample given its thickness and scattering pattern. Multi-response Gaussian processes are used for the supervised learning task and a simple setup is introduced to obtain the scattering pattern of a tissue sample. To increase the predictive power of the supervised learner, the scattering patterns are filtered, enriched by a regressor, and finally characterized with two parameters, namely, transmitted power and scaled Gaussian width. We computationally illustrate that our approach achieves errors of roughly 5% in predicting the scattering properties of many biological tissues. Our method has the potential to facilitate the characterization of tissues and fabrication of phantoms used for diagnostic and therapeutic purposes over a wide range of optical spectrum.

Recently, considerable effort has been devoted to improving the quality of fabricated tissue phantoms<sup>1–4</sup> as they are instrumental in the optical *in-vitro* investigations concerning cancer diagnosis<sup>5</sup>, therapeutic applications<sup>6,7</sup>, and drug efficacy tests<sup>8</sup>. In this regard, one avenue of research has pursued the use of accurate and cost-effective phantom characterization techniques to guide the fabrication process. The most widely recognized characterization techniques for this purpose are spatial frequency domain imaging (SFDI)<sup>9,10</sup>, frequency domain photon migration (FDPM)<sup>11,12</sup>, and inverse adding-doubling (IAD)<sup>13,14</sup>. The fundamental idea of these techniques is to computationally model the scattering phenomenon in tissue phantoms and subsequently estimate the scattering properties of such materials by calibrating the computational model against some experimental data. Below, we briefly describe these methods and then introduce our approach which enables the data-driven estimation of scattering properties of tissues by employing a supervised learner (which is fitted to a training dataset of tissues' characteristics) in an inverse optimization procedure. Our method is inexpensive, non-intrusive, efficient, and applicable to a wide range of materials.

In the case of a wavefront with a lateral sinusoidal intensity profile, the penetration depth and the diffuse reflectance depend on the lateral spatial frequency. The latter quantity, can be used to obtain the optical properties as well as the optical tomography of the sample<sup>10</sup>. The essence of the SFDI technique is to employ this relation by matching the measured and calculated diffuse reflectance for a set of wavefronts with different spatial frequencies. As for the FDPM technique, the analytical expressions of the phase lag, amplitude attenuation, and complex wave vector of a semi-infinite turbid medium are fitted to the corresponding measured values of the reflected beam to find the scattering parameters. In FDPM, the two ways of collecting the fitting data are: (i) for a fixed temporal modulation frequency, the distance between the LED source and the detector is changed, and (ii) for a fixed distance between the source and the detector, the temporal modulation frequency is varied over a wide range.

<sup>1</sup>Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, 60208, USA.

<sup>2</sup>Department of Mechanical Engineering, Northwestern University, Evanston, IL, 60208, USA. Iman Hassaninia and Ramin Bostanabad contributed equally to this work. Correspondence and requests for materials should be addressed to H.M. (email: [hmohseni@northwestern.edu](mailto:hmohseni@northwestern.edu))



**Figure 1.** The schematics of the simulation setup: The image of the transmitted scattered waves on the surface of the turbid medium is used to obtain its scattering properties. The input to the turbid medium is a linearly polarized pencil beam of an LED. A free space polarizer is used to reject the un-scattered beam based on the polarization of the input LED. The color bar besides the image indicates the range of gray intensity. See the text for more details on our simulation setup.

In both SFDI and FDP methods, the diffusion equation is used to approximate the Boltzmann transport equation. This results in the overestimation (underestimation) of the diffuse reflectance at low (high) spatial frequencies. In addition to the fitting error, enforcing the boundary conditions in the diffusion equation<sup>15</sup> introduces some error in arriving at the analytical formulas for realistic semi-infinite media. Moreover, the experimental setup in both SFDI and FDP methods are complex and costly. In SFDI, in addition to a spatial light modulator, two polarizers at the source and detector are needed to reject the specular reflection collected normal to the surface. As for FDP technique, a network analyzer is required to modulate the current of the LED and to detect the diffuse reflectance of the temporally modulated beam. These instruments render the setup complex and costly. Furthermore, these methods are incapable of measuring the *anisotropy coefficient* of the sample,  $g$ , which is an important parameter for characterizing turbid media<sup>16–18</sup>. In biological tissues, the probability of scattering a beam of light at an angle  $\theta$  (with respect to the incoming beam) can be described suitably by the Henyey-Greenstein phase function<sup>19,20</sup>:

$$P(\theta) = \frac{1}{4\pi} \frac{(1 - g^2)}{(1 + g^2 - 2g\cos(\theta))^{\frac{3}{2}}}, \quad (1)$$

where the optical properties of the turbid medium depend on both  $g$  (that characterizes the angular profile of scattering) as well as the scattering length,  $s_l$ , the average distance over which the scattering occurs.

Among these techniques, IAD is the most popular one due to its relatively higher accuracy and simpler experimental setup. Briefly, IAD is based on matching the measured and the calculated diffuse reflectance and transmittance by calibrating the scattering and absorption coefficients used in the simulations. When an accurate measurement of the un-scattered transmission can be made, it is possible to obtain  $g$  as well. In IAD, the errors are mostly attributed to the experimental data. For instance, when measuring the total transmission and reflectance, part of the light scattered from the edge of the sample can be lost, or when measuring the un-scattered transmission, the scattered rays may unavoidably influence the measurement<sup>13</sup>.

We propose an efficient method to address the above challenges and have a better compromise between accuracy and the cost of measuring the scattering parameters (i.e.,  $g$  and  $s_l$ ). Our method is based on a supervised learner that can predict the scattering pattern of a turbid medium given its thickness ( $t$ ) and scattering parameters. Once this supervised learner is found, the scattering parameters of any turbid sample can be calculated given its thickness and the image of the scattered rays' pattern either by inverting the supervised learner or performing an optimization task.

Our process for obtaining the scattering pattern, as illustrated in Fig. 1, starts by producing a pencil beam from an LED placed behind an aperture. The pencil beam has a well-defined but arbitrary polarization and is incident on the turbid medium with a *known* thickness. The surface of this medium is then imaged to a camera sensor through a lens, where the un-scattered beam with the well-defined polarization is rejected via a polarizer placed next to the turbid medium. We note that with such a non-coherent and phase-insensitive measurement, the size of the image as well as the components scale with the diameter of the laser. Because of this scaling rule, the length unit of the image shown in Fig. 1 equals the number of the *scaled* pixels of the camera. We also note that for a collimated illumination, the distance between the source and the sample is arbitrary. A similar argument holds for the distance between the polarizer and the sample because the un-scattered light is collimated.

We employed the same configuration as in Fig. 1 in our computational simulations. In particular, we placed the camera lens far from the sample (15 cm) such that the scattered light is almost parallel to the optical axis. We employed a lens with a focal length, radius, and maximum numerical aperture of, respectively, 4 cm, 6 mm, and

0.15. Additionally, the optical resolution of the system according to Rayleigh's criterion was  $4 \mu\text{m}$  at  $1550 \text{ nm}$ <sup>21</sup>, which is equal to the pixel pitch of the detector in our simulations. Since the pixel pitch was larger than half of the optical resolving limit and hence the Nyquist criterion was not satisfied, the scattering patterns are slightly blurred. As for the LED bandwidth,  $\Delta\lambda$ , we chose it wide enough to have a coherent length much smaller than the optical path length of the rays ( $L_{\text{coh}} < 0.1L_{\text{opt}}$ ). With this choice, the coherent effects do not distort the scattered images. In particular, we ensured that  $\Delta\lambda > \frac{2 \ln(2)}{\pi n} \frac{\lambda^2}{0.1L_{\text{opt}}}$  where  $n$  is the refractive index and  $\lambda$  is the wavelength of interest. The minimum required bandwidth is  $40 \text{ nm}$  when  $L_{\text{opt}} = 200 \mu\text{m}$ ,  $n = 1.33$ , and  $\lambda = 1550 \text{ nm}$ .

We note that, in our method the simulations are performed on thin slabs of phantom or tissue with known thickness. Although performing the same type of experiment using *reflection* is in principle possible, we expect much weaker reflection than transmission for such thin slabs. Additionally, the reduction of the signal strength translates into lower SNR and higher measurement errors in the case of reflection. We have also found some experimental works which are based on quantitative *phase of the transmission* images of thin samples for which  $t \ll s_j$ <sup>22,23</sup>. As opposed to this latter approach, our method is based on the intensity of the scattering patterns which is simpler and applicable to thicker samples.

To fit our supervised learner, a high-fidelity training dataset of input-output pairs is required. Here, the inputs (collectively denoted by  $\mathbf{x}$ ) are the characteristics of the turbid samples (i.e.,  $g$ ,  $s_j$ , and  $t$ ) while the outputs (collectively denoted by  $\mathbf{y}$ ) are some finite set of parameters that characterize the corresponding scattering patterns (i.e., the images similar to the one in Fig. 1). We elaborate on the choice of the latter parameters in Sec. 0 but note that they must be sufficiently robust to noise so that, given  $t$ , the scattering parameters of any turbid sample can be predicted with relatively high accuracy using the supervised learner.

## Results

To construct the computational training dataset, we used the Sobol sequence<sup>24,25</sup> to build a space filling design of experiments (DOE) with 400. points (i.e., simulation settings) over the hypercube  $0.7 \leq g \leq 0.93$ ,  $0.03 \leq s_j \leq 0.12 \text{ mm}$ , and  $200 \leq t \leq 600 \mu\text{m}$ . It is noted that the lower limit on the sample thickness is because of the considerable inaccuracies associated with the negligible probability of scattering in thin samples. In contrast, the upper limit on the sample thickness is bounded due to the computational costs associated with tracing the large number of ray scatterings. As for ranges of  $g$  and  $s_j$ , they cover the scattering properties of a wide range of biological tissues including but not limited to liver<sup>26</sup>, white brain matter, grey brain matter, cerebellum, and brainstem tissues (pons, thalamus)<sup>27</sup>.

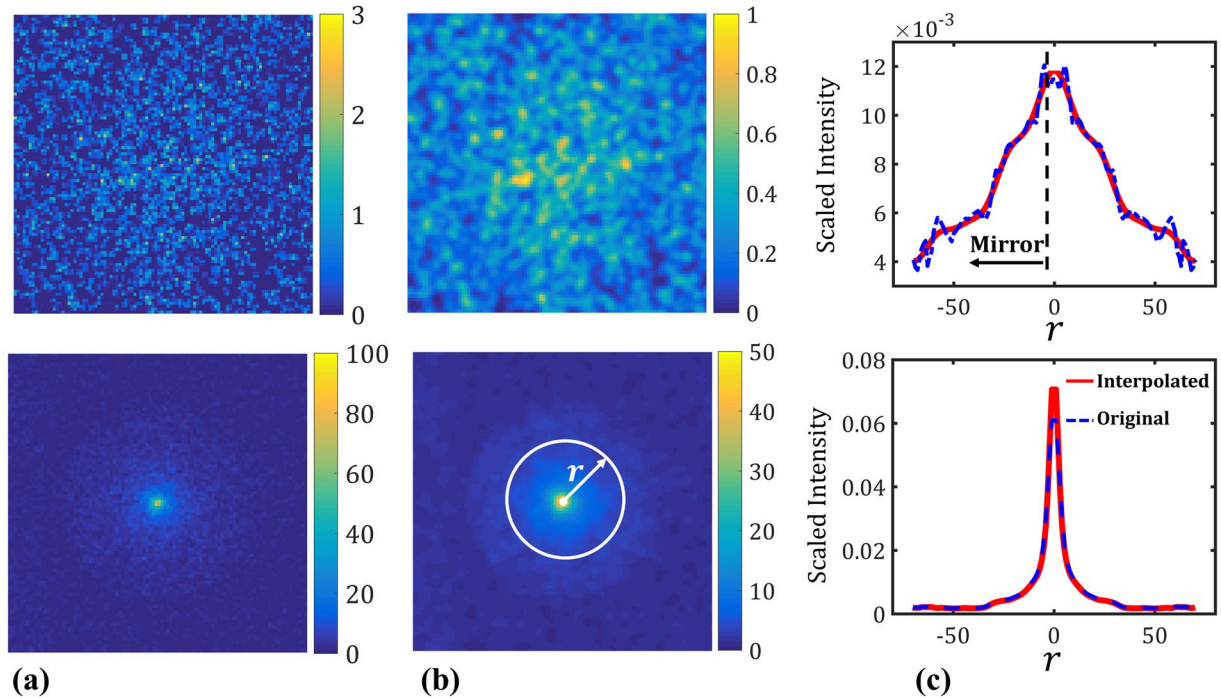
Once the simulation settings were determined, following the schematic in Fig. 1, the scattering pattern corresponding to each of them was obtained by the commercial raytracing software *Zemax OpticStudio*. Although there are many software programs applicable for this task (such as *Code v*, *Oslo*, and *FRED*), *Zemax* is perhaps the most widely used software for ray tracing. Unlike mode solvers, ray tracing is computationally fast. The significant scattering effects as well as the employed broadband light source (i.e., the LED) further justify the use of a ray-tracing software. In our simulations with *Zemax*, rigorous Monte-Carlo simulations were conducted for higher accuracy (instead of solving the simplified diffusion equation) and the turbid media were simulated with the built-in Henyey-Greenstein model<sup>28</sup>. To push the upper limit on the sample thickness to  $600 \mu\text{m}$ , we increased the number of Monte Carlo intersections and observed that the maximum capacity of *Zemax* (roughly two million segments per ray) must be employed for sufficient accuracy. Additionally, we found that a  $100 \times 100$  rectangular detector and five million launched rays provide a reasonable compromise between the accuracy and the simulation costs (about 3 minutes for each input setting).

As mentioned in Sec. 1, the scattering patterns corresponding to the simulation settings (i.e., the DOE points) must be characterized with a finite set of parameters (denoted by  $\mathbf{y}$  in Fig. 1) to reduce the problem dimensionality and enable the supervised learning process. To determine the sufficient number of parameters, we highlight that our end goal is to arrive at an inverse relation where the  $g$  and  $s_j$  of a tissue sample with a *specific* thickness can be predicted. Therefore, if the parameters are chosen such that both  $g$  and  $s_j$  are monotonic functions of them, two characterizing parameters are required for a one-to-one relation. It must be noted that, these parameters must be sufficiently robust to the inherent errors in the simulations mentioned above. We will elaborate on this latter point below and in Sec. 3.

We have conducted extensive studies and our results indicate that the *transmitted power*,  $p$ , and the *scaled Gaussian width*,  $\sigma$ , can sufficiently and robustly characterize the scattering patterns of a wide range of tissue samples. While  $p$  measures the amount of the LED beam power transmitted through the sample and collected at the image,  $\sigma$  measures the extent to which the sample scatters the LED beam. It is evident that these parameters are negatively correlated, i.e., increasing  $p$  would decrease  $\sigma$  and vice versa.

Measuring  $p$  for an image is straightforward as it only requires integrating the gray intensity over all the image pixels. Measuring  $\sigma$ , however, requires some pre-processing because the amount of scattering in an image is sensitive to noise and has a strong positive correlation with it (i.e., high scattering would involve a high degree of noise in the image and vice versa). As illustrated in Fig. 2, we take the following steps to measure  $\sigma$  for an image:

1. Filtering the image with a Gaussian kernel to eliminate the local noises (see panel b in Fig. 2). In general, the width of the Gaussian kernel depends on the resolution of the original image as well as the amount of noise. In our case, the filtering was conducted (in the frequency space) with a kernel width of 7 pixels.
2. Obtaining the radial distribution of the intensity by angularly averaging it over the image.
3. Mirroring the radial distribution to obtain a symmetric curve and then scaling it so that the area under the curve equals unity (see panel c in Fig. 2). At this point, the resulting symmetric curve would approximate a zero-mean Gaussian probability distribution function (PDF).



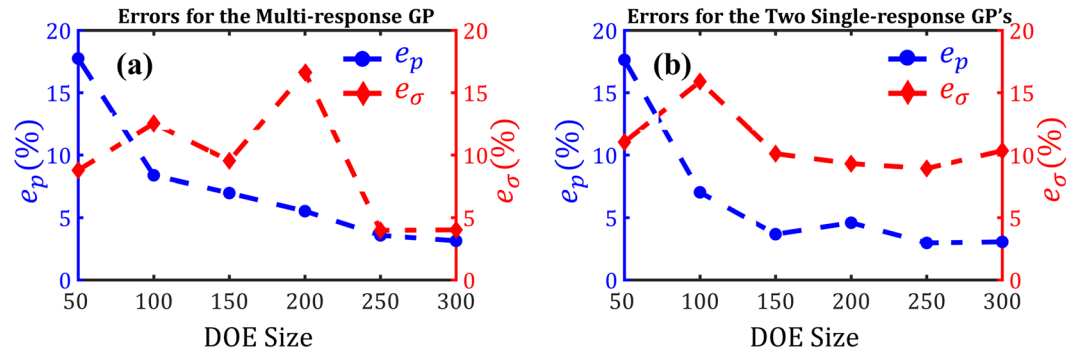
**Figure 2.** Measuring  $\sigma$  for two different images: (a) The original scattering patterns obtained by *Zemax* following the schematic in Fig. 1. Both images have a side length of 100 pixels. The simulation settings (i.e.,  $\mathbf{x} = [t, g, s_i]$ ) for the images on the top and bottom row correspond to  $[51.1, 0.75, 0.04]$  and  $[24.4, 0.83, 0.1]$ , respectively. The color bars indicate the range of gray intensity for each image. (b) The filtered images. The same Gaussian kernel is used to filter out the local noises in both images in panel (a). (c) The scaled radial distribution of gray intensity. The area under the dashed blue curve is approximately unity in both cases and the solid red curves are the regressors fitted to the original data. While in the top plot the noise is considerable (even after filtering with a Gaussian kernel), in the bottom plot the noise is negligible and hence the regressor almost interpolates. Gaussian processes are employed as regressors here with an estimated nugget of  $10^{-3}$  and  $10^{-5}$  for the top and bottom plots, respectively.

4. Fit a regressor to further reduce the noise and enrich the scattered data which resemble a Gaussian PDF (compare the solid and dashed lines in panel c).
5. Estimate the standard deviation of the Gaussian PDF via the enriched data. Divide this standard deviation by the power of the image (i.e.,  $p$ ) to obtain  $\sigma$ .

As for the regressor, we recommend employing a method that can address the *potential* high amount of noise in some of the images which, as mentioned earlier, happens when scattering is significant (e.g., when  $t$  is large while  $g$  and  $s_i$  are small). We have used Gaussian processes (GP's), neural networks, and polynomials for this purpose but recommend the use of GP's mainly because they, following the procedure outlined in ref.<sup>29</sup>, can automatically address high or small amounts of noise. Additional attractive features of GP's are discussed in Sec. 2.1 and Sec. 5.

The reason behind scaling the standard deviation in step 5 by  $p$  is to leverage the negative correlation between the transmitted power and the noise to arrive at a better measure for estimating *scattering*. To demonstrate this, consider two images where one of them is noisier than the other. It is obvious that the noisier image must be more scattered and hence have a larger scattering measure. To increase the difference between the scattering measures (and, subsequently, increase the predictive power of the supervised learner), one can divide them with a variable that is larger (smaller) for the smaller (larger) scattering measure. This variable, in our case, is the transmitted power which is rather robust to the noise.

Finally, we note that the images were not directly used in the supervised learning stage as outputs because: (i) Predicting the scattering pattern is not our only goal. Rather, we would like to have a limited set of parameters (i.e., outputs) that can sensibly characterize the image and hence provide guidance as to how the inputs (i.e.,  $[g, s_i, t]$ ) affect the outputs (and correspondingly the scattering patterns). Using the images directly as outputs is a more straightforward approach but renders monitoring the trends difficult. (ii) With  $100 \times 100$  outputs (the total number of pixels), fitting a multi-response supervised learner becomes computationally very expensive and, more importantly, may face severe numerical issues. One can also fit  $100 \times 100$  single-response supervised learners but this is rather cumbersome, expensive, and prone to errors due to high amounts of noise in some pixels. (iii) With  $100 \times 100$  outputs, the inverse optimization processes (for estimating  $g$  and  $s_i$  given  $t$  and an image) becomes expensive.



**Figure 3.** Reduction of prediction error (in percent) by increasing the number of training samples: (a) Prediction errors for the MRGP model. (b) Prediction errors for the SRGP models. The errors are all calculated based on 100 test points that were not used in fitting.

**Supervised learner: Linking Scattering Patterns with Tissue Sample Characteristics.** With the advances in computational capabilities, supervised and unsupervised learning methods have drawn considerable attention in a wide range of applications including computational materials science<sup>30–51</sup>, neuroscience<sup>52</sup>, clinical medicine<sup>53–56</sup>, biology<sup>57,58</sup>, protein analysis and genetics<sup>59</sup>, biotechnology<sup>60,61</sup>, robotics<sup>62</sup>, psychology<sup>63</sup>, climatology<sup>64</sup>, paleoseismology<sup>65</sup>, and economics<sup>66</sup>. These methods provide the means to predict the response of a system where no or limited data is available. Neural networks, support vector machines, decision trees, Gaussian processes (GP's), clustering, and random forests are amongst the most widely used methods. In case of biological tissues, supervised learning via neural networks has been previously employed, e.g., for classification of tissues using SFDI-based training datasets<sup>67–72</sup>.

We employ GP's to link the characterizing parameters of the scattering patterns (i.e.,  $p$  and  $\sigma$ ) with those of the tissue samples (i.e.,  $t$ ,  $g$ , and  $s_j$ ). Briefly, the essential idea behind using GP's as supervised learners is to model the input-output relation as a realization of a Gaussian process. GP's are well established in the statistics<sup>73</sup>, computational materials science<sup>33,40</sup>, and computer science<sup>74</sup> communities as they, e.g., readily quantify the prediction uncertainty<sup>75,76</sup> and enable tractable and efficient Bayesian analyses<sup>77,78</sup>. In addition, GP's are particularly suited to emulate highly nonlinear functions especially when insufficient training samples are available.

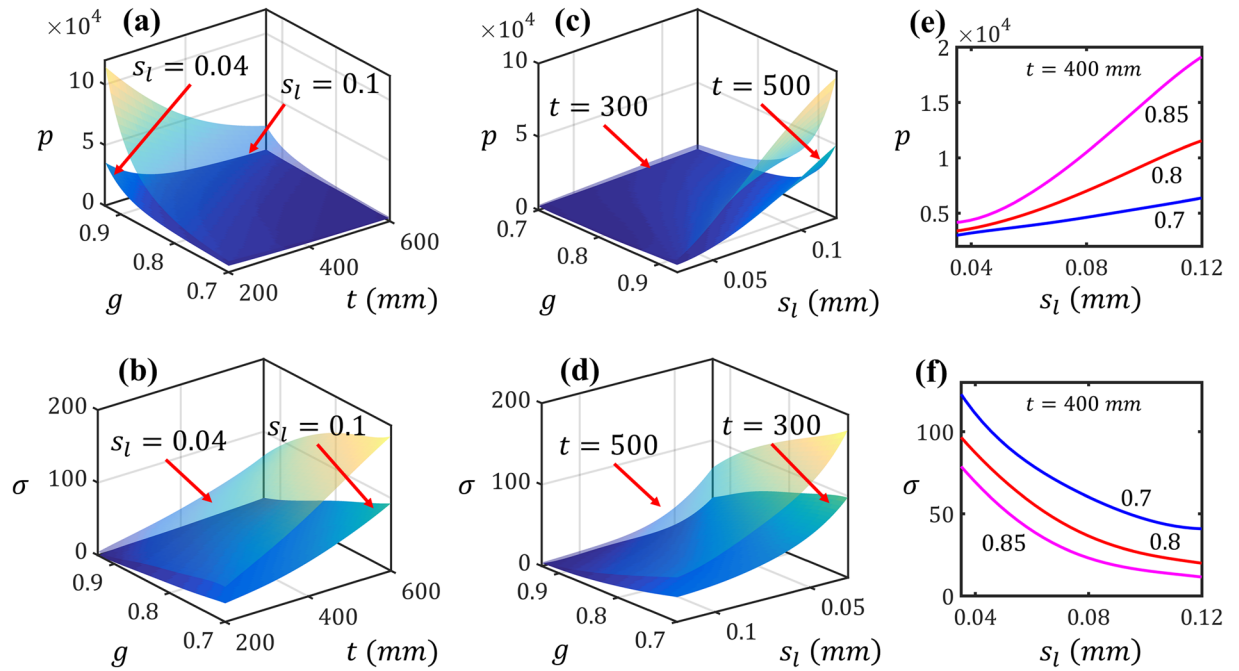
In our case, the inputs and outputs corresponds to  $\mathbf{x} = [t, g, s_j]$  and  $\mathbf{y} = [p, \sigma]$ , respectively. As there are two outputs, we can either fit a multi-response GP (MRGP) model or two independent single-response GP (SRGP) models. With the former approach, one GP model is fitted to map the three-dimensional (3D) space of  $\mathbf{x}$  to the two-dimensional (2D) space of  $\mathbf{y}$ . With the latter approach, however, two GP models are fitted: one for mapping  $\mathbf{x}$  to  $p$  and another for mapping  $\mathbf{x}$  to  $\sigma$ . The primary advantage of an MRGP model lies in capturing the correlation between the responses (if there is any) and, subsequently, requiring less data for a desired level of accuracy. An MRGP model might not provide more predictive power if the responses are independent, have vastly different behavior, or contain different levels of noise.

We conducted convergence studies to decide between the two modeling options and, additionally, determine the minimum DOE size required to fit a sufficiently accurate model. As mentioned earlier, the Sobol sequence was employed to build one DOE of size 400 over the hypercube  $0.7 \leq g \leq 0.93$ ,  $0.03 \leq s_j \leq 0.12$  mm, and  $200 \leq t \leq 600$   $\mu$ m. Sobol sequence was chosen over other design methods (e.g., Latin hypercube) because consecutive subsets of a Sobol sequence all constitute space-filling 50 designs. Following this, we partitioned the first 300 points in the original DOE of size 400 into six subsets with an increment of 50, i.e., the  $i^{\text{th}}$  DOE ( $i = 1, \dots, 6$ ) included points 1, ...,  $i \times 50$  from the original DOE. The last 100 points in the original DOE (which are space-filling and different from all the training points) were reserved for estimating the predictive power of the models. Next, three GP models were fitted to each DOE: (i) an MRGP model to map  $\mathbf{x}$  to  $\mathbf{y}$ , and (ii) two SRGP models; one to map  $\mathbf{x}$  to  $p$  and another to map  $\mathbf{x}$  to  $\sigma$ . Finally, the reserved 100 DOE points were used to estimate the scaled root-mean-squared error (RMSE) as:

$$e_q = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{q - \hat{q}}{q} \right)^2}, \quad (2)$$

where  $N$  is the number of prediction points ( $N = 100$  in our case),  $q$  is the quantity of interest (either  $p$  or  $\sigma$ ), and  $\hat{q}$  is the estimated quantity by the fitted model. Figure 3 summarizes the results of our convergence studies (see Sec. 5.1 for fitting costs) and indicates that:

1. As the sample size increases, the errors generally decrease. The sudden increases in the errors are either due to overfitting or the addition of some noisy data points.
2.  $e_\sigma$  of the MRGP model is almost always smaller than that of the SRGP (compare the red curves in Fig. 3a and b). The opposite statement holds for  $e_p$ . This is because  $p$ , as compared to  $\sigma$ , is much less noisy.



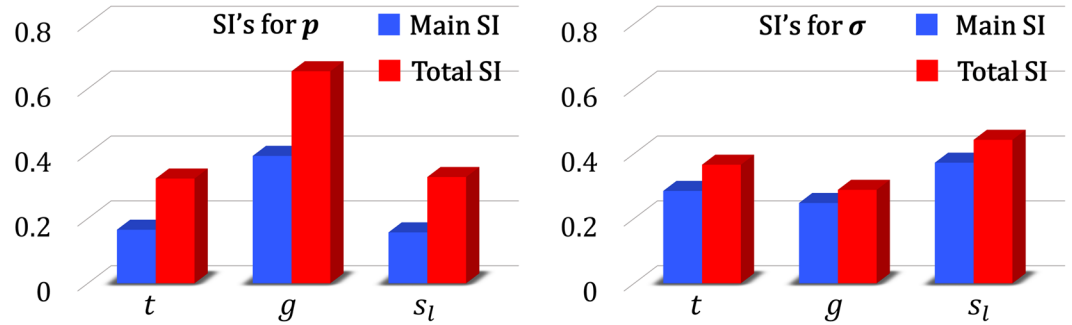
**Figure 4.** Effect of the characteristics of tissue samples (i.e.,  $[t, g, \mu]$ ) on the characteristics of scattering patterns (i.e.,  $[p, \sigma]$ ): The effect on the transmitted power and the scaled Gaussian width are demonstrated in the figures in, respectively, top and bottom rows. In (a) through (d), either  $t$  or  $\mu$  is fixed and the outputs are plotted versus the other two inputs. In (e) and (f),  $p$  and  $\sigma$  are plotted versus  $\mu$  for three values of  $g$  while having  $t$  fixed to 400 mm.

Based on the convergence studies, we can conclude that an MRGP model with at least 300 training data points can provide, on average, prediction errors smaller than 5%. Following this, we fitted an MRGP model in 28.6 seconds to the entire dataset (i.e., DOE of size 400) and employed it in the subsequent analyses in Sec. 2.2. Figure 4 illustrates how  $p$  and  $\sigma$  (and hence the scattering patterns) change as a function of tissue sample characteristics based on this MRGP model. The plots in top and bottom rows of Fig. 4 demonstrate the effect of inputs on, respectively, the transmitted power and the scaled Gaussian width. In Fig. 4(a) and (b),  $s_l$  is fixed to either 0.1 mm or 0.04 mm and the outputs are plotted versus  $t$  and  $g$ . In Fig. 4(c) and (d),  $t$  is fixed to either 300 mm, or 500 mm and the outputs are plotted versus  $g$  and  $s_l$ . In Fig. 4(e) and (f),  $p$  and  $\sigma$  are plotted versus  $s_l$  for three values of  $g$  while having  $t$  fixed to 400 mm. In summary, these plots demonstrate that decreasing a sample's  $g$  or  $s_l$ , or increasing its thickness, would decrease the transmitted power while increasing the scattering (i.e.,  $\sigma$ ). Moreover, both  $p$  and  $\sigma$  change monotonically as a function of the inputs. This latter feature enables us to uniquely estimate  $g$  and  $s_l$  given  $t, p$ , and  $\sigma$ .

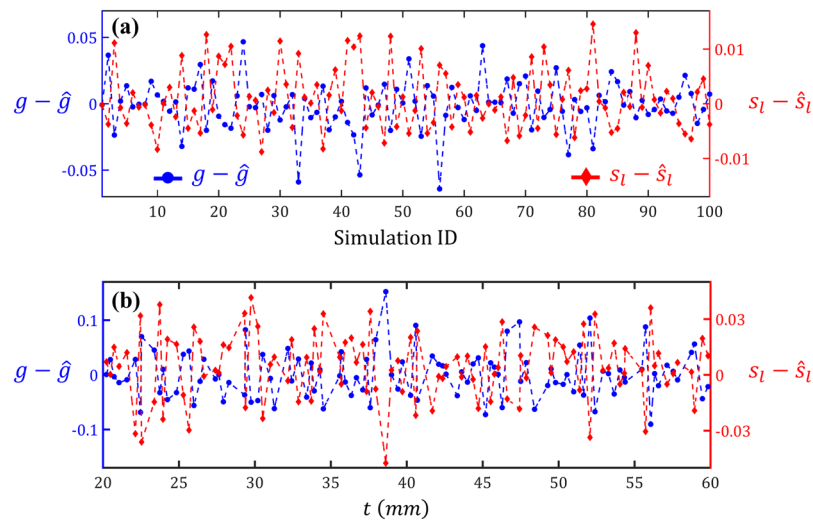
To quantify the relative importance of each input parameter on the two model outputs, we conducted global sensitivity analysis (SA) by calculating the Sobol indices (SI's)<sup>79,80</sup>. As opposed to local SA methods which are based on the gradient, SI's are variance-based quantities and provide a global measure for variable importance by decomposing the output variance as a sum of the contributions of each input parameter or combinations thereof. Generally, two indices are calculated for each input parameter of the model: *main* SI and *total* SI<sup>81</sup>. While a main SI measures the first order (i.e., additive) effect of an input on the output, the total SI measures both the first and higher order effects (i.e., including the interactions). SI's are normalized quantities and known to be efficient indicators of variable importance because they do not presume any specific form (e.g., linear, monotonic, etc.) for the input-output relation.

Using the MRGP model, we conducted quasi Monte Carlo simulations to calculate the main and total SI's of the three inputs for each of the outputs. The results are summarized in Fig. 5 and indicate that all the inputs affect both outputs. While  $p$  is noticeably sensitive to  $g$  (and equally sensitive to  $t$  and  $s_l$ ),  $\sigma$  is almost equally sensitive to all the inputs. It is also evident (as captured by the difference between the height of the two bars for each input) that there is more interaction between the inputs in the case of  $p$  than  $\sigma$ .

**Inverse Optimization: Estimating the Scattering Properties of a Tissue.** Noting that the sample thickness can be controlled in an experiment, tissue characterization is achieved by finding the scattering parameters of the sample given how it scatters a pencil beam in a setup similar to that in Fig. 1. More formally, in our case, tissue characterization requires estimating  $g$  and  $s_l$  given  $p, \sigma$ , and the sample thickness  $t$ . Although in principle we can inverse the MRGP model at any fixed  $t$  to map  $[p, \sigma]$  to  $[g, s_l]$ , this is rather cumbersome. Hence, we cast the problem as an optimization one by minimizing the cost function,  $F$ , defined as:



**Figure 5.** Results of the sensitivity analyses: Main and total sensitivity indices (SI's) for the transmitted power (left) and the scaled Gaussian width (right).



**Figure 6.** Errors in predicting the scattering properties of 100 tissue samples: The scattering properties of most samples have been estimated quite accurately. In (a) and (b) the simulations are ordered with respect to, respectively, simulation number and the sample thickness.

$$[\hat{g}, \hat{s}_l] = \underset{g, s_l}{\operatorname{argmin}} F = \underset{g, s_l}{\operatorname{argmin}} \log \left( \left( 1 - \frac{p_{GP}}{p_e} \right)^2 + \left( 1 - \frac{\sigma_{GP}}{\sigma_e} \right)^2 + 1 \right), \tag{3}$$

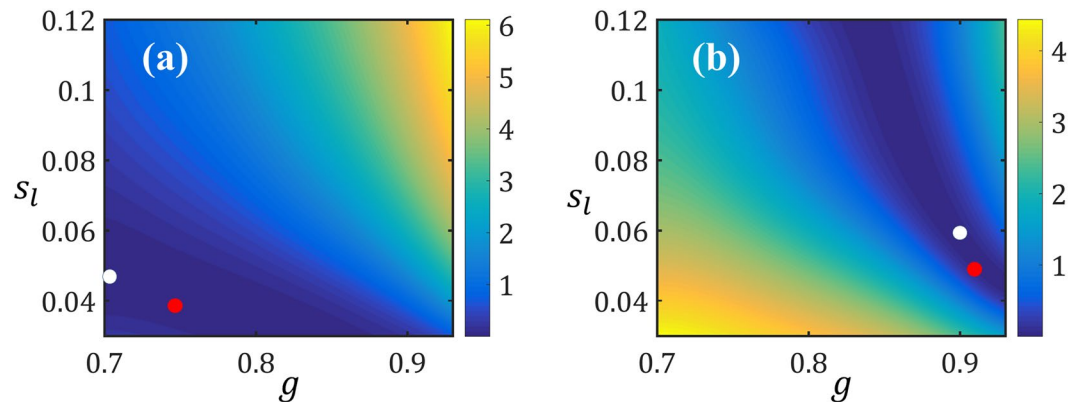
where  $F$  is the cost function which measures the difference between the experimental values (i.e.,  $p_e$  and  $\sigma_e$ ) and the predicted ones by the MRGP model (i.e.,  $p_{GP}(t, g, s_l)$  and  $\sigma_{GP}(t, g, s_l)$ ). We note that, the model predictions are subject to  $t = t_e$  where  $t_e$  is the sample thickness.

To test the accuracy of the fitted MRGP model in estimating the scattering parameters, we generated a space-filling test dataset of size 100 while ensuring that none of the test points were the same as the 400 training ones used in fitting the MRGP. For each test point, then, the outputs (i.e.,  $p$  and  $\sigma$ ) and the sample thickness (i.e.,  $t$ ) were used to estimate the inputs (i.e.,  $\hat{g}$  and  $\hat{s}_l$ ) by minimizing Eq. 3. To solve Eq. 3, we used the *Fmincon* command in the optimization toolbox of MATLAB®. Figure 6(a) illustrates the prediction errors of estimating  $g$  (on the left axis) and  $s_l$  (on the right axis) for the 100 test points. It is evident that the average errors are zero in estimating either  $g$  or  $s_l$ , indicating that the results are indeed unbiased. In Fig. 6(b) the errors are plotted with respect to the sample thickness to investigate whether they are correlated with  $t$ . As no obvious pattern can be observed, it can be concluded that our procedure for estimating  $g$  and  $s_l$  is quite robust over the range where  $t$  is sampled in the training stage.

For further investigations, we normalize the errors and provide the summary statistics in Table 1. As quantified by the scaled RMSE, on average, the prediction errors are relatively small, especially for the  $g$ . The maximum scaled errors are 9.1% (corresponding to simulation ID 56 in Fig. 6) and 18.1% (simulation ID 22) for  $g$  and  $s_l$ , respectively. Figure 7 demonstrates the contour plots of the cost function for these two simulations. As it can be observed, in each case there are regions in the search space of  $[g, s_l]$  where the cost function  $F$  is approximately constant. In fact, the true optimum and the estimated solution (indicated, respectively, with white and red dots in

	Scaled RMSE (%)	Scaled Max. Error (%)
$g$	2.41	9.1
$s_L$	7.87	18.1

**Table 1.** Summary of prediction errors: The scaled RMSE (see Eq. 2) and scaled maximum error are calculated for the 100 data points in Fig. 6.



**Figure 7.** Contours of  $F$  as a function of the optimization variables: The fixed variables are  $[t, p, \sigma] = [31.3, 4.2 \times 10^3, 70.1]$  in (a) and  $[t, p, \sigma] = [25.7, 2.7 \times 10^4, 8.3]$  in (b). The true and estimated optima are indicated with, respectively, white and red dots in each case.

Fig. 7) are on the loci where  $F$  is minimized. The existence of such loci can be explained by noting that  $g$  and  $s_L$  have similar effects on both responses, i.e., increasing (decreasing) either of them would decrease (increase)  $\sigma$  while increasing (decreasing)  $p$ .

Finally, we note that the inverse optimization cost is negligible (less than 10 seconds) in our case because we have employed a gradient-based optimization technique which converges fast because (i) it uses the predictions from the MRGP model for both the response and its gradient (which are done almost instantaneously), and (ii) we have reduced the dimensionality of the problem from  $100 \times 100$  (the number of pixels in each image) to two (i.e.,  $[p, \sigma]$ ).

## Discussions

In our computational approach, the accuracy in predicting the scattering parameters of a turbid medium mainly depends on (i) the errors in *Zemax* simulations, (ii) the predictive power of  $p$  and  $\sigma$  in characterizing the scattering patterns, and (iii) the effectiveness of the supervised learner and the optimization procedure.

The inherent numerical errors in *Zemax* inevitably introduce some error into the training dataset. In addition, the number of launched rays in our Monte Carlo simulations, though having utilized the maximum capacity of *Zemax*, might be insufficient and hence introduce some inaccuracies. This latter source of error particularly affects samples which scatter the incoming LED beam more (e.g., thick samples with small  $g$  and  $s_L$ ) because once the number of segments per a launched ray exceeds the software's limit, the ray is discarded.

To reduce the problem dimensionality and enable the supervised learning process, the images of the scattering patterns (see, e.g., Fig. 2a) were characterized with two negatively correlated parameters, namely,  $p$  and  $\sigma$ . Since the images are not entirely symmetric and may not completely resemble a Gaussian pattern, employing only  $\sigma$  to capture their patterns' spread will introduce some error. We have addressed this source of error, to some extent, by filtering the images with a Gaussian kernel (see, e.g., Fig. 2b) and enriching the radial distribution of the scattering patterns by a GP regressor (see Fig. 2c). Our choice of regressor, in particular, enabled automatic filtering of small to large amounts of noise through the so-called *nugget parameter*. Additionally, we leveraged the negative correlation between  $p$  and  $\sigma$  in the definition of  $\sigma$  to increase its sensitivity to the spreads. The supervised learning and inverse optimization procedures will, of course, benefit from reducing the simulation errors and finding parameters with more predictive power than  $\sigma$ .

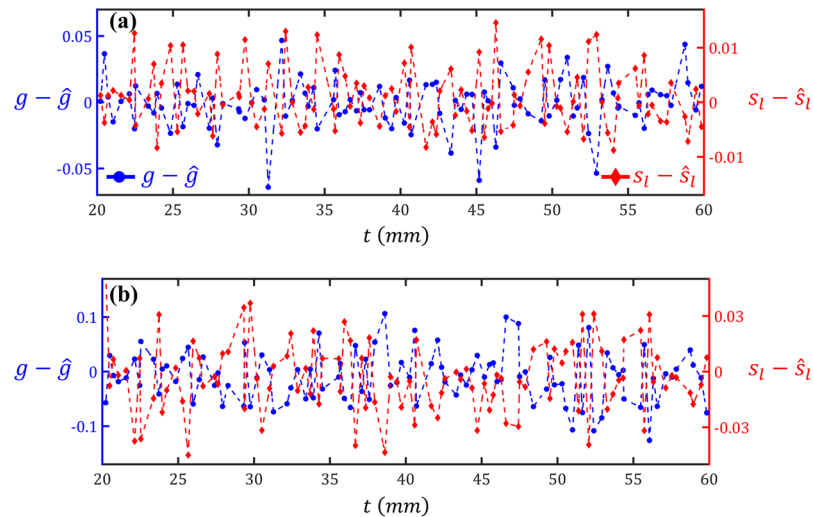
As for the supervised learner, we illustrated that a multi-response Gaussian process can provide sufficient accuracy with a relatively small training dataset (see Fig. 3a). Learning both responses (i.e.,  $p$  and  $\sigma$ ) simultaneously, in fact, helped to better address the noise due to the negative correlation between the responses. As demonstrated in Fig. 3a, the MRGP model with 300 training samples can achieve, on average, errors smaller than 5%. Increasing the size of the training dataset would decrease the error but, due to the simulation errors, an RMSE of zero cannot be achieved. Additionally, sensitivity analyses were conducted by calculating the Sobol indices of the inputs (i.e.,  $t, g$ , and) using the MRGP model. As illustrated in Fig. 5, all the inputs are effective and affect both outputs with  $p$  being noticeably sensitive to  $g$  and embodying more interactions between the inputs.

We casted the problem of determining the scattering parameters as an inverse optimization one where  $g$  and  $s_L$  of a tissue sample were estimated given its thickness  $t$ , and the corresponding scattering pattern (i.e.,  $p$  and  $\sigma$ ). In optimization parlance, the objective or cost function (defined in Eq. 3) achieves the target scattering pattern by



	Errors with 10% Overestimation of $t$	Errors with 10% Underestimation of $t$
$g$	5.90	5.32
$s_L$	33.02	28.11

**Table 2.** Summary of prediction errors while enforcing 10% error in thickness: The scaled RMSE's (see Eq. 2) are calculated for the 100 test points while enforcing 10% difference between the sample thickness used in the simulations and the thickness used in the inverse optimization.



**Figure 8.** Errors in predicting the scattering properties of 100 tissue samples while enforcing 10% error in thickness value: In (a) and (b) the thickness value used in the inverse optimization is, respectively, increased and decreased by 10% from the original value used in *Zemax* simulations. The simulations are ordered with respect to the sample thickness.

searching for the two unknown inputs while constraining the sample thickness. As illustrated in Fig. 6 for 100 test cases, our optimization procedure provides an unbiased estimate for the scattering parameters with an error of roughly 5%. The inaccurate estimations in our optimization studies are because  $g$  and  $s_L$  have similar effects on  $p$  and  $\sigma$ . This is demonstrated in Fig. 7 where the local optima of the objective function create a locus and hence overestimating  $g$  or  $s_L$  would result in underestimating the other and vice versa. To quantify this effect, we calculated Spearman's rank-order correlation between the errors  $g - \hat{g}$  and  $s_L - \hat{s}_L$  for the 100 data points reported in Fig. 6 and found it to be  $-0.90$ . Such a strong negative correlation value (Spearman's rank-order correlation is, in the absence of repeated data values, between  $-1$  and  $1$ ) indicates that when  $g$  is underestimated,  $s_L$  will be overestimated and vice versa.

When coupling our method with experimental data, there will be some measurement errors primarily due to the noise of the camera sensor, insufficient rejection of the un-scattered waves, and the inaccuracy in determining the sample thickness. To address the dark current noise of the camera, the camera integration time should be increased. It is also favorable to increase the power of LED to reduce the influence of the parasitic rays of the environment, but caution must be practiced to avoid saturating the camera or damaging the sample. To ensure proper rejection of the un-scattered rays, an image corresponding to the absolute pixel-by-pixel difference of the two images with and without the polarization filter should be obtained. Then, the maximum intensity on the difference image should be compared with that of the image obtained with the polarization filter. Lastly, to minimize the errors due to sample thickness, samples with uniform and carefully measured thickness must be prepared. Inaccuracies in measuring the sample thickness or using considerably non-uniform ones, will adversely affect the prediction results. To quantify the sensitivity of the predictions to inaccuracies associated with thickness, we repeated the inverse optimization process in Sec. 2.2 considering potential measurement errors of 10%. In particular, we redid the inverse optimization for the test dataset while employing thickness values with 10% difference from the true values (i.e., instead of  $t$ ,  $0.9t$  or  $1.1t$  were used). The estimated values (i.e.,  $\hat{g}$  and  $\hat{s}_L$ ) were then compared to the true ones. As summarized in Table 2, the scaled RMSE's have, especially in the case of  $s_L$ , increased. To see whether the sample thickness has a correlation with the errors, in Fig. 8 the errors are plotted versus  $t$ . Although the errors are quite large in some cases (which is expected because the true  $t$  value is not used in the inverse optimization), the overall results are unbiased (i.e., the average errors are close to zero). These results indicate that, as long as the sample thickness is measured sufficiently accurate, the model can provide an unbiased estimate for  $g$  and  $s_L$ .

Lastly, we compare the accuracy of our approach to other methods. The reported errors in Fig. 3, Table 1, and Fig. 6 do not consider the errors that will be introduced upon experimental data collection. As explained below, our error estimates are comparable to those of the FDP, SFDI, and IAD methods from a computational standpoint. It is noted that in all these methods (including ours) the dominating error will be associated with experimental data (once it is used in conjunction with simulations).

In FDP, the model bias<sup>77</sup> originates from the assumptions made for solving the diffusion equation (e.g., using a semi-infinite medium as opposed to a finite-size sample)<sup>82</sup>. Besides this, there are two other error sources in FDP (i) the preliminary error due to approximating light transport in tissues with diffusion equation is estimated to be 5~10%<sup>83</sup>, and (ii) the error due to the quantum shot noise limit of the instrument which depends on the configuration and components of the system. For a reasonable system comprising of two detectors and one source at the modulation frequency of 500 MHz, the limit of quantum shot noise results in about 2% error in estimating the scattering coefficient<sup>82</sup>. The scattering length,  $s_l$ , is the inverse of the scattering coefficient and will roughly have the same error. Similarly, the percentage errors of  $s_l$  roughly equals to that of the reduced scattering coefficient defined as  $(1 - g)/s_l$ . Considering only these two sources of noise, we can assume a total noise of around 10% for the FDP technique. As for the SFDI technique, the diffusion approximation results in an overall reported error of around 3% for the reduced scattering coefficient<sup>9</sup>. In the IAD method, the prediction errors are sensitive to the input data. For instance, it is reported that with a 1% perturbation in the inputted transmission and reflection amounts, the relative error in estimating the scattering coefficient and anisotropy factor increases 10 and 4 times, respectively<sup>13</sup>.

## Conclusion

We have introduced a non-invasive method for computational characterization of the scattering parameters (i.e., the anisotropy factor and the scattering length) of a medium. The essence of our approach lies in finding a supervised learner that can predict the scattering pattern of a turbid medium given its thickness and scattering parameters. Once this supervised learner is found, we solve an inverse optimization problem to estimate the scattering parameters of any turbid sample given its thickness and the image of the scattered rays' pattern. Additionally, our approach is computationally inexpensive because the majority of the cost lies in building the training dataset which is done once.

To the best of our knowledge, this is one of the simplest and most inexpensive methods of tissue characterization because, in practice, only a few basic and low-cost instruments such as an LED, an aperture, a polarizer, and a camera are required. Additionally, our analyses and results are independent of the wavelength of the LED and therefore the scattering parameters of many tissues can be estimated over a wide range of visible and infrared wavelengths. We note that, in our method it is assumed that the absorption is much weaker than the scattering and thus its effect on the output images is negligible. This assumption holds for some tissues including white brain matter, grey brain matter, cerebellum, and brainstem tissues where the scattering coefficient is more than 100 times larger than the absorption coefficient in most of the visible and in the reported near-infrared range<sup>27</sup> (see Table 3 in ref.<sup>84</sup> for more details). Measuring weak absorption of tissue with our method requires more intense data analysis and processing. However, we believe that this limit doesn't translate into impracticality of our method as there are methods<sup>22,23</sup> which can only estimate  $g$  and  $s_l$ .

We plan to experimentally validate our approach and quantify the effect of measurement errors (due to, e.g., the noise of the camera sensor and insufficient rejection of the un-scattered waves) on estimating the scattering parameters. We believe that this method has the potential to facilitate the fabrication of tissue phantoms used for diagnostic and therapeutic purposes over a wide range of optical spectrum.

## Methods

**Gaussian Process Modeling.** GP modeling has become the de-facto supervised learning technique for fitting a response surface to training datasets of either costly physical experiments or expensive computer simulations due to its simplicity, flexibility, and accuracy<sup>74,85-88</sup>. The fundamental idea of GP modeling, is to model the dependent variable,  $y$ , as a realization of a random process,  $Y$ , with inputs  $\mathbf{x} = [x_1, x_2, \dots, x_n]$  where  $n \in \mathbb{N}$  is the number of inputs. The underlying regression model can be formally stated as:

$$Y(\mathbf{x}) = \sum_{i=1}^k \beta_i f_i(\mathbf{x}) + Z(\mathbf{x}) \quad (4)$$

where  $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_k(\mathbf{x})]$  are a set of known basis functions,  $\beta = [\beta_1, \dots, \beta_k]$  are unknown coefficients, and  $Z(\mathbf{x})$  is the random process. In Eq. 4, following the statistical conventions, the random variables/processes and their realizations are denoted with, respectively, upper and lower cases. Assuming  $Z(\mathbf{x})$  is a zero-mean GP with covariance function  $c(\cdot, \cdot)$  of the form

$$c(\mathbf{x}, \mathbf{w}) = sR(\mathbf{x}, \mathbf{w}) \quad (5)$$

between  $Z(\mathbf{x})$  and  $Z(\mathbf{w})$ , GP modeling essentially consists of estimating the  $\beta$  coefficients, process variance  $s$ , and parameters of the correlation function  $R(\cdot, \cdot)$ . Often, the maximum likelihood estimation (MLE) method is used for this purpose<sup>89,90</sup>.

We implemented an in-house GP modeling code in Matlab<sup>®</sup> following the procedure outlined in ref.<sup>29</sup>. The so-called Gaussian correlation function was employed with an addition of a nugget parameter,  $\delta$ , to address the possible noises:

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Training Size	50	100	150	200	250	300
Fitting Cost (seconds)	1.83	3.68	5.89	8.00	14.00	16.89

**Table 3.** Computational cost of fitting MRGP models in our convergence study: As the number of the training samples increases, the fitting cost increases as well.

$$R(\mathbf{x}, \mathbf{w}) = \exp\left(-\sum_{i=1}^n \theta_i (x_i - w_i)^2\right) + \delta, \quad (6)$$

where  $\theta$  are the roughness parameters estimated via MLE. For noiseless datasets,  $\delta$  is generally set to either a very small number (e.g.,  $10^{-8}$ ) to avoid numerical issues, or zero. In our work, we have used GP's for two purposes: (i) to smooth out the radial distribution of the scattered rays and enrich the associated PDF for a better estimation of its standard deviation (see Fig. 2), and (ii) to fit a response surface for mapping  $\mathbf{x} = [t, g, \mu]$  to  $\mathbf{y} = [p, \sigma]$ . We emphasize that, the adaptive procedure of ref.<sup>29</sup> allows to adjust  $\delta$  in Eq. 6 to address negligible to large amounts of noise.

Multiple studies have extended GP modeling to multi-output datasets. Of particular interest, has been the work of Conti. *et al.*<sup>91</sup> where the essential idea is to concatenate the vector of responses (i.e.,  $\mathbf{y} = [y_1, \dots, y_u]^T$  for  $u$  outputs) and model the covariance function as  $c(\mathbf{x}, \mathbf{w}) = \mathbf{s} \otimes R(\mathbf{x}, \mathbf{w})$  where  $\mathbf{s}$  is the  $u \times u$  covariance matrix of the responses and  $\otimes$  is the Kronecker product. Finally, it is noted that since we did not know *a priori* how  $p$  and  $\sigma$  change as a function of  $\mathbf{x} = [t, g, \mu]$ , a constant basis function was used (i.e.,  $f(\mathbf{x}) = 1$ ) in all our simulations.

The computational cost of fitting each of the MRGP models used in the convergence study is summarized in Table 3. As it can be observed, the costs are all small and increase as the size of the training dataset increases.

**Sensitivity Analysis with Sobol Indices.** Sobol indices (SI's) are variance-based measures for quantifying the global sensitivity of a model output to its inputs. For a model of the form  $y(\mathbf{x}) = f(x_1, \dots, x_n)$ , the main SI for the  $i^{\text{th}}$  input  $x_i$  is calculated as:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))}{V(Y)}, \quad (7)$$

where  $V(Y)$  is the total variance of the output,  $\mathbf{X}_{-i}$  denotes all the inputs except  $X_i$ ,  $V_{X_i}$  is the variance with respect to  $x_i$ , and  $E_{\mathbf{X}_{-i}}(Y|x_i)$  is the expectation of  $Y$  for all the possible values of  $\mathbf{X}_{-i}$  while keeping  $X_i$  fixed. Using the law of total variance, one can show that  $S_i$ 's are normalized quantities and vary between zero and one. It is noted that, similar to the above, the random variables and their realizations are denoted with, respectively, upper and lower cases.

The total SI for the  $i^{\text{th}}$  input is calculated as:

$$S_{T_i} = \frac{E_{\mathbf{X}_{-i}}(V_{X_i}(Y|\mathbf{X}_{-i}))}{V(Y)} = 1 - \frac{V_{\mathbf{X}_{-i}}(E_{X_i}(Y|\mathbf{X}_{-i}))}{V(Y)} \quad (8)$$

which includes the contributions from all the terms in the variance decomposition that include  $X_i$ . Comparing Eqs 7 and 8, it is evident that  $0 \leq S_i \leq S_{T_i} \leq 1$ . Saltelli *et al.*<sup>81</sup> provided numerical methods based on quasi Monte Carlo simulations for efficiently calculating both the main and total SI's.

**Data availability.** The datasets and statistical models generated during this study are available from the corresponding author.

## References

- Chen, C. *et al.* Preparation of a skin equivalent phantom with interior micron-scale vessel structures for optical imaging experiments. *Bio. opt. exp* **5**, 3140–3149 (2014).
- Martelli, F. *et al.* Phantoms for diffuse optical imaging based on totally absorbing objects, part 2: experimental implementation. *J. bio.opt.* **19**, 076011–076011 (2014).
- Krauter, P. *et al.* Optical phantoms with adjustable subdiffusive scattering parameters. *J. bio.opt.* **20**, 105008–105008 (2015).
- Böcklin, C., Baumann, D., Stuker, F. & Fröhlich, J. Mixing formula for tissue-mimicking silicone phantoms in the near infrared. *J. Phy.D: Appl. Phys.* **48**, 105402 (2015).
- Manoharan, R. *et al.* Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging. *Photochem. Photobiol.* **67**, 15–22 (1998).
- Ng, D. C. *et al.* On-chip biofluorescence imaging inside a brain tissue phantom using a CMOS image sensor for *in vivo* brain imaging verification. *Sens. and Act. B: Chem.* **119**, 262–274 (2006).
- Ley, S., Stadthaler, M., Link, D., Laqua, D. & Husar, P. In *Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE*. 1432–1435 (IEEE) (2014).
- Tien, L. W. *et al.* Silk as a Multifunctional Biomaterial Substrate for Reduced Glial Scarring around Brain-Penetrating Electrodes. *Adv. Func. Mat.* **23**, 3185–3193 (2013).

9. Cuccia, D. J., Bevilacqua, F., Durkin, A. J., Ayers, F. R. & Tromberg, B. J. Quantitation and mapping of tissue optical properties using modulated imaging. *J. bio. opt.* **14**, 024012–024012–024013 (2009).
10. Cuccia, D. J., Bevilacqua, F., Durkin, A. J. & Tromberg, B. J. Modulated imaging: quantitative analysis and tomography of turbid media in the spatial-frequency domain. *Opt. lett.* **30**, 1354–1356 (2005).
11. Tromberg, B. J. *et al.* Non-invasive measurements of breast tissue optical properties using frequency-domain photon migration. *Phil. Trans. Roy. Soc. London B: Bio. Sci.* **352**, 661–668 (1997).
12. Pham, T. H., Coquoz, O., Fishkin, J. B., Anderson, E. & Tromberg, B. J. Broad bandwidth frequency domain instrument for quantitative tissue optical spectroscopy. *Rev. Sci. Instr.* **71**, 2500–2513 (2000).
13. Prael, S. A., van Gemert, M. J. & Welch, A. J. Determining the optical properties of turbid media by using the adding-doubling method. *Appl. opt.* **32**, 559–568 (1993).
14. Prael, S. Everything I think you should know about Inverse Adding-Doubling. *Oregon Medical Laser Center, St. Vincent Hospital*, 1–74 (2011).
15. Haskell, R. C. *et al.* Boundary conditions for the diffusion equation in radiative transfer. *JOSA. A* **11**, 2727–2741 (1994).
16. Peters, V., Wyman, D., Patterson, M. & Frank, G. Optical properties of normal and diseased human breast tissues in the visible and near infrared. *Phys. in med. and bio.* **35**, 1317 (1990).
17. Van der Zee, P., Essenpreis, M. & Delpy, D. T. In *Proc. SPIE*. 454–465.
18. Patterson, M. S., Wilson, B. C. & Graff, R. *In vivo* tests of the concept of photodynamic threshold dose in normal rat liver photosensitized by aluminum chlorosulphonated phthalocyanine. *Photochem. Photobiol.* **51**, 343–349 (1990).
19. Cornette, W. M. & Shanks, J. G. Physically reasonable analytic expression for the single-scattering phase function. *Appl. opt.* **31**, 3152–3160 (1992).
20. Henyey, L. G. & Greenstein, J. L. Diffuse radiation in the galaxy. *The Astro. J.* **93**, 70–83 (1941).
21. Goodman, J. W. *Introduction to Fourier optics*. (Roberts and Company Publishers, 2005).
22. Lee, M. *et al.* Label-free optical quantification of structural alterations in Alzheimer's disease. *Sci. Rep.* **6**, 31034 (2016).
23. Ding, H. *et al.* Measuring the scattering parameters of tissues from quantitative phase imaging of thin slices. *Opt. lett.* **36**, 2281–2283 (2011).
24. Sobol, I. M. On quasi-Monte Carlo integrations. *Math. Comput. Simul.* **47**, 103–112, [https://doi.org/10.1016/S0378-4754\(98\)00096-2](https://doi.org/10.1016/S0378-4754(98)00096-2) (1998).
25. Sobol, I. Y. M. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **7**, 784–802 (1967).
26. De Jode, M. Monte Carlo simulations of light distributions in an embedded tumour model: studies of selectivity in photodynamic therapy. *Lasers. Med. Sci.* **15**, 49–56 (2000).
27. Yaroslavsky, A. *et al.* Optical properties of selected native and coagulated human brain tissues *in vitro* in the visible and near infrared spectral range. *Phys. Med. Biol.* **47**, 2059 (2002).
28. Kim, M. *et al.* Optical lens-microneedle array for percutaneous light delivery. *Biomedical opt. expr.* **7**, 4220–4227 (2016).
29. Bostanabad, R., Kearney, T., Tao, S., Apley, D. W. & Chen, W. Leveraging the Nugget Parameter for Efficient Gaussian Process Modeling. *Intl. J. Num. Meth in Eng.* (2017).
30. Bostanabad, R., Bui, A. T., Xie, W., Apley, D. W. & Chen, W. Stochastic microstructure characterization and reconstruction via supervised learning. *Acta. Mat.* **103**, 89–102 (2016).
31. Breneman, C. M. *et al.* Stalking the Materials Genome: A Data-Driven Approach to the Virtual Design of Nanostructured Polymers. *Adv. Funct. Mater.* **23**, 5746–5752, <https://doi.org/10.1002/adfm.201301744> (2013).
32. Kalidindi, S. R. & De Graef, M. Materials Data Science: Current Status and Future Outlook. *Ann. Rev. of Mat. Res.* **45**, 171–193, <https://doi.org/10.1146/annurev-matsci-070214-020844> (2015).
33. Bessa, M. A. *et al.* A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality. *Comp. Meth. in Appl. Mech. and Eng.* **320**, 633–667, <https://doi.org/10.1016/j.cma.2017.03.037> (2017).
34. Liu, R., Yabansu, Y. C., Agrawal, A., Kalidindi, S. R. & Choudhary, A. N. Machine learning approaches for elastic localization linkages in high-contrast composite materials. *Integ. Mat. and Manufac. Innov.* **4**, 13, <https://doi.org/10.1186/s40192-015-0042-z> (2015).
35. Le, B., Yvonnet, J. & He, Q. C. Computational homogenization of nonlinear elastic materials using neural networks. *Int. J. Numeric. Meth. in Eng.* **104**, 1061–1084 (2015).
36. Bostanabad, R., Chen, W. & Apley, D. W. Characterization and reconstruction of 3D stochastic microstructures via supervised learning. *J. Microsc.* **264**, 282–297, <https://doi.org/10.1111/jmi.12441> (2016).
37. Matouš, K., Geers, M. G., Kouznetsova, V. G. & Gillman, A. A review of predictive nonlinear theories for multiscale modeling of heterogeneous materials. *J. of Comp. Phys.* **330**, 192–220 (2017).
38. Xu, H. Y., Liu, R. Q., Choudhary, A. & Chen, W. A Machine Learning-Based Design Representation Method for Designing Heterogeneous Microstructures. *J. Mech. Design* **137**, 051403, <https://doi.org/10.1115/1.4029768> (2015).
39. Geers, M. & Yvonnet, J. Multiscale modeling of microstructure-property relations. *MRS. Bullet.* **41**, 610–616 (2016).
40. Xue, D. *et al.* Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **7**, 11241, <https://doi.org/10.1038/ncomms11241> (2016).
41. Curtarolo, S. *et al.* The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201, <https://doi.org/10.1038/nmat3568> (2013).
42. Kalinin, S. V., Sumpter, B. G. & Archibald, R. K. Big-deep-smart data in imaging for guiding materials design. *Nat. Mater.* **14**, 973–980, <https://doi.org/10.1038/nmat4395> (2015).
43. Balachandran, P. V., Young, J., Lookman, T. & Rondinelli, J. M. Learning from data to design functional materials without inversion symmetry. *Nat. Comm.* **8**, 14282 (2017).
44. Seko, A., Maekawa, T., Tsuda, K. & Tanaka, I. Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids. *Phys. Rev. B.* **89**, <https://doi.org/10.1103/PhysRevB.89.054303> (2014).
45. Cang, R. J. & Ren, M. Y. Deep Network-Based Feature Extraction and Reconstruction of Complex Material Microstructures. *Proc. ASME Int. Design Eng. Tech. Conf. and Comp. and Infor. in Eng. Conference, 2016, Vol 2b*, 95–104 (2016).
46. Balachandran, P. V., Theiler, J., Rondinelli, J. M. & Lookman, T. Materials Prediction via Classification Learning. *Sci. Rep.* **5**, 13285, <https://doi.org/10.1038/srep13285> (2015).
47. Liu, R. *et al.* A predictive machine learning approach for microstructure optimization and materials design. *Sci. Rep.* **5**, 11551, <https://doi.org/10.1038/srep11551> (2015).
48. Pilania, G., Wang, C., Jiang, X., Rajasekaran, S. & Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **3**, 2810, <https://doi.org/10.1038/srep02810> (2013).
49. Balachandran, P. V., Xue, D., Theiler, J., Hogden, J. & Lookman, T. Adaptive Strategies for Materials Design using Uncertainties. *Sci. Rep.* **6** (2016).
50. Kolb, B., Lentz, L. C. & Kolpak, A. M. Discovering charge density functionals and structure-property relationships with PROPhet: A general framework for coupling machine learning and first-principles methods. *Sci. Rep.* **7**, 1192 (2017).
51. Wang, C., Yu, S., Chen, W. & Sun, C. Highly efficient light-trapping structure design inspired by natural evolution. *Sci. Rep.* **3** (2013).

52. Akil, H., Martone, M. E. & Van Essen, D. C. Challenges and Opportunities in Mining Neuroscience Data. *Science* **331**, 708–712, <https://doi.org/10.1126/science.1199305> (2011).
53. Bellazzi, R. & Zupan, B. Predictive data mining in clinical medicine: Current issues and guidelines. *Int. J. Med. Infor.* **77**, 81–97, <https://doi.org/10.1016/j.ijmedinf.2006.11.006> (2008).
54. Buonamici, S. *et al.* CCR7 signalling as an essential regulator of CNS infiltration in T-cell leukaemia. *Nat.* **459**, 1000–1004, [http://www.nature.com/nature/journal/v459/n7249/supinfo/nature08020\\_S1.html](http://www.nature.com/nature/journal/v459/n7249/supinfo/nature08020_S1.html) (2009).
55. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nat.* **452**, 571–579 (2008).
56. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nat.* **506**, 328–333, <https://doi.org/10.1038/nature13038> (2014).
57. Hehemann, J.-H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nat.* **464**, 908–912, [http://www.nature.com/nature/journal/v464/n7290/supinfo/nature08937\\_S1.html](http://www.nature.com/nature/journal/v464/n7290/supinfo/nature08937_S1.html) (2010).
58. Warnecke, F. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nat.* **450**, 560–565, [http://www.nature.com/nature/journal/v450/n7169/supinfo/nature06269\\_S1.html](http://www.nature.com/nature/journal/v450/n7169/supinfo/nature06269_S1.html) (2007).
59. Vinayagam, A. *et al.* Protein Complex-Based Analysis Framework for High-Throughput Data Sets. *Sci. Sign.* **6**, rs5–rs5, <https://doi.org/10.1126/scisignal.2003629> (2013).
60. Besnard, J. *et al.* Automated design of ligands to polypharmacological profiles. *Nat.* **492**, 215–220, <http://www.nature.com/nature/journal/v492/n7428/abs/nature11691.html#supplementary-information> (2012).
61. Tatonetti, N. P., Ye, P. P., Daneshjou, R. & Altman, R. B. Data-Driven Prediction of Drug Effects and Interactions. *Sci. Trans. Med.* **4**, 125ra131–125ra131, <https://doi.org/10.1126/scitranslmed.3003377> (2012).
62. Cully, A., Clune, J., Tarapore, D. & Mouret, J.-B. Robots that can adapt like animals. *Nat.* **521**, 503–507, <https://doi.org/10.1038/nature14422> <http://www.nature.com/nature/journal/v521/n7553/abs/nature14422.html#supplementary-information> (2015).
63. Whelan, R. *et al.* Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nat.* **512**, 185–189, <https://doi.org/10.1038/nature13402> <http://www.nature.com/nature/journal/v512/n7513/abs/nature13402.html#supplementary-information> (2014).
64. Chavez, E., Conway, G., Ghil, M. & Sadler, M. An end-to-end assessment of extreme weather impacts on food security. *Nat. Clim. Change.* **5**, 997–1001, <https://doi.org/10.1038/nclimate2747> <http://www.nature.com/nclimate/journal/v5/n11/abs/nclimate2747.html#supplementary-information> (2015).
65. Yoon, C. E., O'Reilly, O., Bergen, K. J. & Berzoza, G. C. Earthquake detection through computationally efficient similarity search. *Sci. Adv.* **1**, <https://doi.org/10.1126/sciadv.1501057> (2015).
66. Einav, L. & Levin, J. Economics in the age of big data. *Science* **346**, <https://doi.org/10.1126/science.1243089> (2014).
67. Warnecke, D., Lewis, E., Lochmann, S. & Leahy, M. In *J. of Phys.: Conf. Ser.* 012047 (IOP Publishing).
68. Farrell, T. J., Patterson, M. S., Hayward, J. E., Wilson, B. C. & Beck, E. R. In *OE/LASE'94*. 117–128 (International Society for Optics and Photonics).
69. Pfefer, T. J. *et al.* Reflectance-based determination of optical properties in highly attenuating tissue. *J. Bio. Opt.* **8**, 206–215 (2003).
70. Sharma, D., Agrawal, A., Matchette, L. S. & Pfefer, T. J. Evaluation of a fiberoptic-based system for measurement of optical properties in highly attenuating turbid media. *Biomed. Eng. Online.* **5**, 49 (2006).
71. Bruulsema, J. *et al.* Correlation between blood glucose concentration in diabetics and noninvasively measured tissue optical scattering coefficient. *Opt. Lett.* **22**, 190–192 (1997).
72. Zhang, L., Wang, Z. & Zhou, M. Determination of the optical coefficients of biological tissue by neural network. *J. of Mod. Opt.* **57**, 1163–1170 (2010).
73. Plumlee, M. & Apley, D. W. Lifted Brownian kriging models. *Technometrics* (2016).
74. Rasmussen, C. E. Gaussian processes for machine learning (2006).
75. Jin, R., Du, X. & Chen, W. The use of metamodelling techniques for optimization under uncertainty. *Struc. and Multi. Opt.* **25**, 99–116 (2003).
76. Worley, B. Deterministic uncertainty analysis. (Oak Ridge National Lab., 1987).
77. Kennedy, M. C. & O'Hagan, A. Bayesian calibration of computer models. *J. Roy. Stat. Soc. Ser. B. (Stat. Meth.)* **63**, 425–464 (2001).
78. Farhang-Mehr, A. & Azarm, S. Bayesian meta-modelling of engineering design simulations: a sequential approach with adaptation to irregularities in the response behaviour. *Int. J. Num. Meth. Eng.* **62**, 2104–2126, <https://doi.org/10.1002/nme.1261> (2005).
79. Sobol, I. Y. M. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie* **2**, 112–118 (1990).
80. Sudret, B. Global sensitivity analysis using polynomial chaos expansions. *Rel. Eng. & Sys. Safe.* **93**, 964–979 (2008).
81. Saltelli, A. *et al.* Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comp. Phys. Comm.* **181**, 259–270 (2010).
82. Pogue, B. W. & Patterson, M. S. Error assessment of a wavelength tunable frequency domain system for noninvasive tissue spectroscopy. *J. Biomed. Opt.* **1**, 311–323 (1996).
83. Fantini, S., Franceschini, M. A. & Gratton, E. Semi-infinite-geometry boundary problem for light migration in highly scattering media: a frequency-domain study in the diffusion approximation. *JOSA. B.* **11**, 2128–2138 (1994).
84. Cheong, W.-F., Prael, S. A. & Welch, A. J. A review of the optical properties of biological tissues. *IEEE. J. Quant. Elect.* **26**, 2166–2185 (1990).
85. Sacks, J., Welch, W. J., Mitchell, T. J. & Wynn, H. P. Design and analysis of computer experiments. *Stat. Sci.* 409–423 (1989).
86. MacDonald, B., Ranjan, P. & Chipman, H. GPfit: An R package for fitting a gaussian process model to deterministic simulator outputs. *J. Stat. Soft.* **64** (2015).
87. Ba, S. & Joseph, V. R. Composite Gaussian process models for emulating expensive functions. *Ann. Appl. Stat.* 1838–1860 (2012).
88. Zhang, L., Wang, K. & Chen, N. Monitoring wafer geometric quality using additive gaussian process model. *IIE Transactions* (2015).
89. Martin, J. D. & Simpson, T. W. Use of kriging models to approximate deterministic computer models. *AIAA. J.* **43**, 853–863 (2005).
90. Jin, R., Chen, W. & Simpson, T. W. Comparative studies of metamodelling techniques under multiple modelling criteria. *Struc. and Multi. Opt.* **23**, 1–13 (2001).
91. Conti, S., Gosling, J. P., Oakley, J. E. & O'hagan, A. Gaussian process emulation of dynamic computer codes. *Biometrika* asp208 (2009).

## Acknowledgements

The authors appreciate the anonymous reviewers for their insightful comments. Grant support from National Science Foundation (NSF EEC-1530734) is appreciated. We would like to acknowledge partial support from ARO award #W911NF-11-1-0390. In addition, the authors would like to thank the Digital Manufacturing and Design Innovation Institute (DMDII), a UI LABS collaboration, for its funding support to Ramin Bostanabad through award number 15-07-07.

### Author Contributions

I.H. initiated the project and performed the *Zemax* simulations. R.B. conducted the design of experiments, supervised learning, sensitivity analyses, and optimization tasks. All authors discussed the results and contributed to the manuscript.

### Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017