Research Article Vol. 31, No. 20 / 25 Sep 2023 / Optics Express 32737
Optics EXPRESS

Neural network approach for correction of multiple scattering errors in the LISST-VSF instrument

HÅVARD S. UGULEN,* DANIEL KOESTNER, HÅKON SANDVEN, BØRGE HAMRE, ARNE S. KRISTOFFERSEN, AND CAMILLA SAETRE

Department of Physics and Technology, University of Bergen, Allegaten 55, 5007 Bergen, Norway
havard.ugulen@uib.no

Abstract: The LISST-VSF is a commercially developed instrument used to measure the volume scattering function (VSF) and attenuation coefficient in natural waters, which are important for remote sensing, environmental monitoring and underwater optical wireless communication. While the instrument has been shown to work well at relatively low particle concentration, previous studies have shown that the VSF obtained from the LISST-VSF instrument is heavily influenced by multiple scattering in turbid waters. High particle concentrations result in errors in the measured VSF, as well as the derived properties, such as the scattering coefficient and phase function, limiting the range at which the instrument can be used reliably. Here, we present a feedforward neural network approach for correcting this error, using only the measured VSF as input. The neural network is trained with a large dataset generated using Monte Carlo simulations of the LISST-VSF with scattering coefficients $b = 0.05 - 50\,\mathrm{m}^{-1}$, and tested on VSFs from measurements with natural water samples. The results show that the neural network estimated VSF is very similar to the expected VSF without multiple scattering errors, both in angular shape and magnitude. One example showed that the error in the scattering coefficient was reduced from 103% to 5% for a benchtop measurement of natural water sample with expected $b = 10.6\,\mathrm{m}^{-1}$. Hence, the neural network drastically reduces uncertainties in the VSF and derived properties resulting from measurements with the LISST-VSF in turbid waters.

1. Introduction

Accurate characterization of light-particle interactions in marine environments have become increasingly important in recent years due to growing interest in applications of optical measurements for environmental monitoring, remote sensing, and underwater wireless communication [1–6]. The volume scattering function (VSF) is an important inherent optical property (IOP), describing the angular distribution and intensity of scattered light. Measurements of the VSF β can be used to derive the scattering coefficient b through integration of the VSF over all scattering directions and the phase function $\tilde{\beta}$ through normalization of the VSF by b. Accurate measurements of the VSF of natural water samples have been difficult to obtain, requiring custom-built and complex instrumentation which are often challenging to deploy for *in situ* measurements. Consequently, the VSF has been sparingly measured, and common simplified models have not been extensively validated [7], despite its importance in radiative transfer modelling [2,5,8,9]. Variations in the scattering properties of natural waters are primarily governed by suspended particles on the micro- and nano-scale. The VSF may be utilized to extract considerable information on the particulate composition in the water mass [10–13].

The LISST-VSF (Sequoia Scientific) is a commercially available instrument, designed to perform *in situ* measurements of the VSF at a single light wavelength of 515 nm over angles

#495523 https://doi.org/10.1364/OE.495523
Journal © 2023 Received 16 May 2023; revised 20 Jul 2023; accepted 31 Aug 2023; published 18 Sep 2023

Optics EXPRESS

0.1 – 150° with two detection systems; a series of 32 logarithmically-spaced ring detectors covering the range 0.1 – 15° and a rotating eyeball detector covering the range 15 – 150° in 1° increments. Additionally, the LISST-VSF can simultaneously measure the attenuation coefficient c, enabling the extraction of the absorption coefficient a through the relation $c = a + b$. Although the LISST-VSF has been validated to perform well with relatively low concentrations of known particles [14–16], previous studies have also shown that large errors in the measured VSF arise in turbid waters [13,17,18]. In particular, Monte Carlo simulations of the LISST-VSF revealed elevated and distorted VSFs originating from multiple scattering. Consequently, the phase function and scattering coefficient derived from the VSF measurement are also subject to multiple scattering errors. The Monte Carlo simulation has been verified by laboratory measurements [17], where only small deviations were observed. Simulations show that, for an expected scattering coefficient of $b \sim 1.5\,\mathrm{m}^{-1}$, an overestimation of 10% can be observed in the measured scattering coefficient, reaching 100% at $b \sim 10\,\mathrm{m}^{-1}$. On the other hand, the attenuation coefficient c is not significantly affected by multiple scattering errors [13,15,19]. Hence, errors in the scattering coefficient also affect the derived absorption coefficient ($a = c - b$), which can become negative for sufficient errors in the measured scattering coefficient.

Multiple scattering is a known challenge for optical measurement systems seeking to characterize turbid media [10,11,20,21]. Multiple scattering of light within the LISST-VSF instrument impacts the VSF measurement results in two ways: overestimation of VSF magnitude and incorrect angular shape (i.e., phase function). These issues relate to multiple scattered light which is presumed lost along the path between the light source and detector. Both scattering and absorption losses are accounted for along this path using an attenuation correction determined as e^{-cz} where z is pathlength. This correction utilizes a measurement of attenuation from the LISST-VSF with a very narrow acceptance angle (<0.04°). However, a portion of the scattered light assumed lost is not truly lost as it continues through the water and is ultimately detected. For example, light which in a single scattering scenario would be detected at 20° is scattered a second or third time, resulting in detected light at either the same, or a different detection angle. This light is accounted for by the attenuation correction for the 20° detection angle, but is also observed as an actual signal in the detector. Hence, this light is effectively double-counted. Furthermore, multiple scattering generally results in a flattening of the typical forward-peaked phase function of marine particles (e.g., [22]). Combined, these effects result in an overestimation of VSF magnitude at all scattering angles and a flatter VSF than would be observed in a single scattering scenario.

The correction of these undesirable multiple scattering effects on light scattering measurements is nontrivial and depends on many factors including particle properties (e.g., concentration and phase function) and instrument geometry (e.g., pathlength and acceptance angle). A common solution has been to develop light scattering meters with relatively short pathlengths; however, this approach is not always feasible considering desired angular-resolution and retrieval of adequate signal for detection of small scattering signals common in oceanic environments. The HydroScat-6 instrument (HS-6; HOBI Labs, Inc.) has a 15 cm pathlengh, approximately the same as the LISST-VSF instrument, but only measures light scattered at 141° from incident direction. HOBI Labs recommend adjusting the scattering coefficient by a factor of 0.4 when deriving an attenuation coefficient for the attenuation correction (i.e., $c^* = a + 0.4b$, where c^* refers to an adjusted c for attenuation correction), under the assumption that only a portion of scattered light is truly lost along the path [23]. Recently, Doxaran et al. [24] used light simulations of HydroScat instruments to show that adjustments to the scattering coefficient of factors of about 0.05–0.4 were actually warranted for different scenarios of optical properties. This type of approach may be sufficient for some cases with a single-angle scattering meter; however, more advanced approaches are necessary for the complexity of multiple scattering effects in the LISST-VSF instrument which includes two detector systems covering a wide angular range.

Optics EXPRESS

The Monte Carlo model developed in [17] is able to reliably reproduce a LISST-VSF measurement given a phase function and scattering coefficient as inputs. By simulating a large variety of different VSFs, we have shown that the magnitude of the multiple scattering error is dependent on both the scattering coefficient and phase function, i.e., magnitude and shape of the VSF [18]. Furthermore, the error varies with angle of detection, such that the percentage error in the measured VSF at 15° may be significantly different from the error at 150°, making it challenging to apply corrections to the measured VSF. Koestner et al. [13] developed a correction function based on empirical data, where two mineral-dominated samples were diluted to obtain benchtop measurements with attenuation coefficients ranging from 1.8 m^{-1} to 14 m^{-1}. Assuming negligible multiple scattering errors for the smallest attenuation coefficient, an expected VSF could be calculated for the higher concentration samples affected by multiple scattering. Hence, by comparing expected VSF with measured VSF, a correction was found as a function of scattering angle and scattering coefficient. For applications, the scattering coefficient was estimated by $b = c_{\mathrm{LISST}} - a_{\mathrm{ac\text{-}s}}$, where $a_{\mathrm{ac\text{-}s}}$ is the absorption coefficient measured with an ac-s instrument (Sea-Bird Scientific) and c_{LISST} is the attenuation coefficient measured with the LISST VSF instrument. While this method accounts for the angle dependent error, it does not account for the effect of varying phase functions and requires an unbiased estimate of the absorption coefficient.

Recently, artificial neural networks (ANNs) have seen a rise in popularity due to their wide range of applications and generalization capabilities [25]. ANNs are being increasingly utilized in aquatic sciences, providing useful models for water quality monitoring [26–33], and the prediction of wave heights [34–37]. One of the primary attributes of ANNs are their ability to provide solutions to complex multivariate problems. In the current study, the desired outcome is to take a VSF measurement which is influenced by multiple scattering errors as input, and find the "true" VSF unaffected by multiple scattering errors representing what would be measured in a single scattering regime. To accomplish this, we recognized the multivariate capabilities of ANNs and developed a feedforward neural network to correct multiple scattering errors unique to LISST-VSF measurements. To train the ANN, a large set of varied data containing both VSFs with multiple scattering errors and the corresponding VSF unaffected by multiple scattering errors was needed. As such experimental datasets are not readily available on a large scale, a large training dataset was generated using 154 different phase functions representative of natural assemblages of aquatic particles and the previously developed Monte Carlo model of the LISST-VSF instrument [17]. In this study, we present comprehensive analysis of the ANN model development and testing to reliably correct for multiple scattering errors in VSF measurements made by the LISST-VSF instrument. We present results of ANN testing with two contrasting natural samples prepared using dilutions to produce various particle concentrations in a laboratory environment and we examine the application of the ANN to a large dataset of LISST-VSF measurements from natural seawater samples spanning highly turbid coastal waters to clear oceanic waters.

2. Methods

A description of the ANN is given in Section 2.1, which includes how the ANN works and how it is structured. The dataset used for training the ANN is presented in section 2.2, followed by a description of how data was treated for use in the ANN (Section 2.3). Details of the training process, such as training algorithm and loss function, are given in Section 2.4. Finally, Section 2.5 describes how the ANN was evaluated and the datasets used for this purpose.

Note that all optical properties presented in this study refer to the properties of the particles in the water, i.e., scattering and absorption by pure water is not included. Also, the primary LISST-VSF used in this study refers to a single wavelength light source of 515 nm, meaning that all quantified properties refer to the value at this particular wavelength. Some data presented also refer to measurements from a different LISST-VSF instrument which utilizes a 532 nm light

Optics EXPRESS

source, however no distinction is made for the proposes of this study and expected differences between optical measurements with 515 nm or 532 nm light source are assumed minor.

2.1. ANN structure

The ANN employed in this study was a multi-layer feedforward neural network containing five layers in total (see Fig. 1) [38]. The model was built in Python using Keras [39], which runs on top of TensorFlow [40]. The first layer is called the input layer and simply contains the input data which is a measured VSF. The input layer is followed by three hidden layers, and the final layer is the output layer containing the VSF corrected for multiple scattering errors. All nodes in one layer are connected to all nodes in the neighboring layers, but the connections only work in the forward direction. After the input layer, each node in a following layer is calculated as a function of every node in the previous layer. For instance, the first node in the first hidden layer in Fig. 1 is calculated as

$$h_1^1 = F(W_{11}^1 x_1 + W_{21}^1 x_2 + \cdots + W_{n1}^1 x_n + b_1^1). \tag{1}$$

Here, x_i is the input VSF at measurement angle i, the weights W_{ij}^1 describes how much the node j in the hidden layer depends on node i in the input layer, i.e., the strength of the connection. The superscript in W_{ij}^1 refers to which layers the weights connect to (first hidden layer is 1, output layer is 4), and the superscript in h_1^1 refers to which hidden layer the node belongs to (see Fig. 1). The term b_1^1 is the bias for node h_1^1, which is a constant that is added to the product of inputs and weights. The function F is called an activation function, of which there are several alternatives to choose from. The hyperbolic tangent function is commonly used and was found to produce good results for the ANN developed in this study. Thus, Eq. (1) becomes

$$h_1^1 = \tanh(W_{11}^1 x_1 + W_{21}^1 x_2 + \cdots + W_{n1}^1 x_n + b_1^1). \tag{2}$$

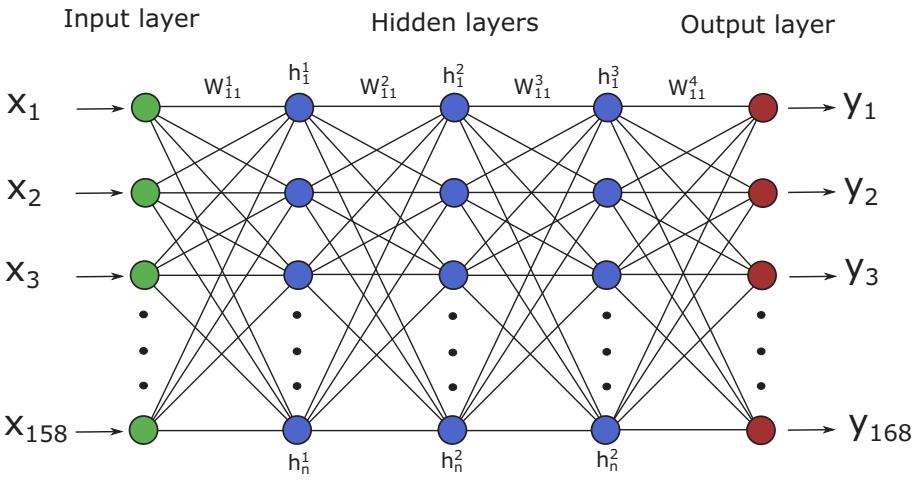

Fig. 1. Illustration of the ANN with a size of three hidden layers with n nodes per layer. The input values X are the original VSF values for angles 0.09 – 140°, nodes h in the hidden layers are calculated as function of W and b (weights and biases), and the output values Y are the estimated VSF values for angles 0.09 – 150°.

The nodes in the first hidden layer are fed forward as inputs to the second hidden layer, and so on until reaching the output layer. The output layer is calculated without the activation function,

Optics EXPRESS

so that the output corresponding to the first angle in the corrected VSF is given as

$$y_1 = W_{11}^4 h_1^3 + W_{21}^4 h_2^3 + \cdots + W_{r1}^4 h_r^3 + b_1^4. \tag{3}$$

The ANN developed in this study takes 158 inputs, corresponding to the VSF measured for angles $0.09 - 140^\circ$. The exclusion of data points above 140° is due to deviations between measured and simulated VSFs in this angular range for large scattering coefficients ($b > 20\,\mathrm{m}^{-1}$) [17]. The output layer contains 168 nodes, corresponding to the corrected VSF over the angles $0.09 - 150^\circ$ (same as the LISST-VSF). Thus, the ANN estimated VSF for angles $140 - 150^\circ$ is predicted based on the measured VSF at angles $0.09 - 140^\circ$.

The size of the network was determined based on a validation process, where the available data was split into a training dataset and a validation dataset. Due to limited data, only 20% of the data was used for validation, while the remaining 80% was kept for training. The phase functions used for validation were randomly selected, but were the same for all configurations. A total of 15 configurations of the ANN structure were tested, varying in both the number of layers and the number of nodes per layer. Based on the results, a network size of three hidden layers with 168 nodes per layer was chosen (same number of nodes as the output layer). The results from the validation process is presented in Section 3.1.

2.2. Training dataset

As mentioned, large experimental VSF datasets with high quality reference measurements are not readily available. However, by utilizing the previously developed Monte Carlo simulation of the LISST-VSF instrument, a large training dataset can be generated [17,18]. The bulk of the training data consists of VSFs simulated with the one-term Henyey-Greenstein (HG), Fournier-Forand (FF), and two-term Reynolds-McCormick (TTRM) phase functions, which covers a wide range of phase functions with resemblance to those found in seawater [2,7,11]. In addition, some VSF were simulated with phase functions obtained from measurements with natural sample, so that more natural VSFs are represented in the training dataset. A total of 154 different phase functions were used, of which 18 were from natural samples. Each was simulated with scattering coefficients ranging from $b = 0.05 - 50\,\mathrm{m}^{-1}$, resulting in a total of 1386 VSFs. The particulate phase functions $\tilde{\beta}$ have backscattering ratios in the range $b_b/b = 0.0003$-0.3 with a median of 0.034, and are shown in Fig. 2.

By using several types of phase functions, one obtains a training set of large variety. Typically, the TTRM and HG phase functions plateau when approaching 0°, while the FF phase functions continue to rise with decreasing angle (see Fig. 2). The TTRM phase function also offers more variety in terms of shape in the backwards direction, where the FF and HG phase functions are much more limited. The phase functions obtained from measurements of natural seawater samples generally provide more realistic training data, accounting for variations that are not adequately represented by the HG, FF, and TTRM phase functions.

2.3. Data transformation and standardization

Since VSF data can vary with several orders of magnitude, it was preprocessed to make the data suitable for application with the ANN using a log transformation and statistical standardization [41,42]. The first step is to transform the values to the $\log_e$ domain, so that

$$x'_{i,j} = \log_e(x_{i,j}), \tag{4}$$

where $x_{i,j}$ is the original VSF data at input node i for sample j. Then the data is standardized by subtracting the mean and dividing by the standard deviation for each individual node in the input

Optics EXPRESS

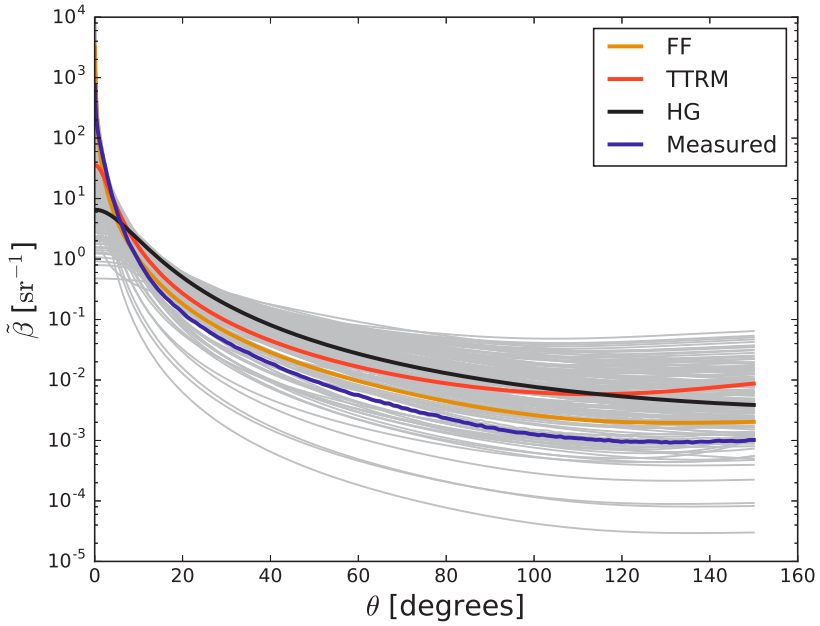

Fig. 2. All phase functions $\tilde{\beta}$ used to generate training data are shown in grey. A representative phase function for each type are shown for Fournier-Forand (yellow), two-term Reynolds-McCormick (Orange), Henyey-Greenstein (black), and from measurement (purple).

layer, giving

$$x''_{i,j} = \frac{x'_{i,j} - \overline{x'_i}}{s_{x'_i}}. \tag{5}$$

Here, $x''_{i,j}$ is the transformed and standardized value for input node i in training sample j, $\overline{x'_i}$ is the mean value of input node i over all training samples, and $s_{x'_i}$ is the standard deviation. This transformation gives a distribution of $x''_{i,j}$ with a mean of zero and a standard deviation of one. The true values $\hat{y}$ are transformed following the equivalent formula, and the output values y from the ANN can be transformed back to regular values by applying the reverse transformation.

2.4. Training the ANN

The aim of the training process is to compute weights W_{ij}^k that minimizes the deviation between the output of the ANN and the true values, and is expressed by a loss function. Here, the mean squared error (MSE) function, which is a commonly used loss function, was found to work well,

$$\text{loss} = \text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{6}$$

Here, y_i is the predicted values, $\hat{y}_i$ is the true values and N is the total number of outputs in the training set, i.e. the number of VSFs times the number of measurement points per VSF. The MSE was calculated for the transformed data, and is likely insensitive to specific angular regions. The ANN was trained using the Adam algorithm, which is a built-in optimizer in Keras. The algorithm minimizes the loss with respect to its parameters W and b (weights and biases) through stochastic gradient descent [43].

Optics EXPRESS

2.5. Evaluating the ANN

It is not possible to directly evaluate the ANN's performance on LISST-VSF measurements of turbid samples influenced by multiple scattering errors, since we do not know the "true" VSF (i.e., the VSF in the single scattering regime unaffected by multiple scattering errors). The preferred way to evaluate the ANN's performance is to obtain LISST-VSF measurements with both a turbid water sample and the same sample diluted to particle concentrations unlikely to induce multiple scattering errors. At a sufficient dilution, the measured VSF will have negligible multiple scattering error and the expected VSF for higher concentration samples can be calculated by multiplying with an appropriate dilution factor. The dilution-corrected VSF can thus be used as "true", or expected, VSF for evaluating the ANN performance. This has been done for two different samples, resulting in two sample sets that contain the original VSF measurements at different dilutions, in addition to the dilution-corrected VSFs. These sample sets are used for evaluation of the ANN in Section 3.2.

Further evaluation of the ANN model was carried out with 98 measurements on natural samples for which we do not have dilution corrections. The dataset presented in Section 3.3 contain a large variation of samples, with measured attenuation coefficients ranging from $c = 0.64 - 21.82\,\mathrm{m}^{-1}$ and backscattering ratios ranging from $\tilde{b}_b = 0.006 - 0.027$. The dataset consist of LISST-VSF measurements collected from various places: Coastal waters of Southern California and Northern Alaska [13]; Coastal waters around Svalberd, including Riåfjorden, Hinlopen and Isfjorden [15]; Norwegian fjords, including Gaupnefjorden [15] and an Emiliana Huxleyi algae bloom in Hardangerfjorden; Coastal measurements in Storfjorden, Svalbard [44].

Assuming that the LISST-VSF attenuation measurement is accurate, the performance of the ANN can be evaluated through the relation between b and measured c. Ideally, one would also have absorption measurements obtained simultaneously with a different instrument, so that a comparable scattering coefficient could be obtained indirectly through the relation $c = a + b$. However, such measurements were not available for the entire dataset presented here.

Percentage deviation was used to quantify differences between values obtained from original uncorrected measurements or values obtained from application of the ANN, and reference values obtained from dilution correction. The percentage deviation is defined as

$$\frac{X_m - X_r}{X_r} \times 100\%, \tag{7}$$

and describes how much the value X_m deviates from the reference value X_r. The values X_m and X_r are specified in the text where Eq. (7) is applied.

3. Results and discussion

This section is divided into three parts. First, the results from the validation process used to determine the size of the network is presented. Then, evaluation of the ANN is performed with dilution-corrected VSFs obtained for two different sample sets. Finally, further evaluation is performed with VSF measurements obtained from natural samples, for which dilution-corrected values are not available. For this section, all scattering and attenuation measurements are performed with the LISST-VSF instrument, unless otherwise is stated.

3.1. Validation and determination of network size

An appropriate ANN size was determined by testing 15 different ANN configurations with hidden layers ranging from 1 to 5, and with either 84, 168 or 252 nodes per hidden layer. Each model was trained for 100,000 iterations and the results are presented in Fig. 3.

From Fig. 3(a), it can be seen that the MSE of the training set (dashed) generally decreases with both increasing number of hidden layers and number of nodes per layer. For the validation

Optics EXPRESS

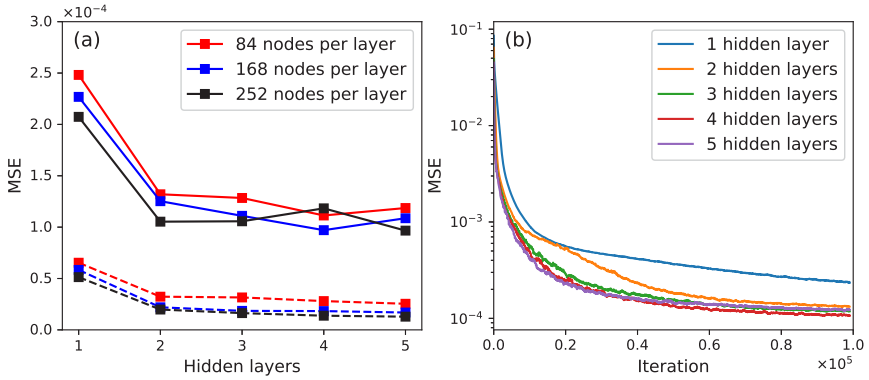

Fig. 3. (a) Mean squared error (MSE) after 100,000 iterations plotted as a function of hidden layers for an ANN with 84, 168 and 252 nodes per layer. The MSE is shown for both the validation set (solid lines) and the training set (dashed lines). (b) MSE of the validation set plotted as a function of iterations for ANNs with a number of hidden layers ranging from 1 to 5, and with 168 nodes per hidden layer.

set (solid), the largest improvement in performance is observed when going from 1 to 2 hidden layers. For both 84 and 168 nodes per layer, the MSE has a small downwards trend going from 2 to 4 layers, with a slight increase at 5 hidden layers. This could be a sign of overfitting, where the amount of parameters (weights and biases) becomes so large that the solution for the training data may not be a good general solution for data not included in the training. Additional evidence of overfitting is observed in Fig. 3(b), where it can be seen that the 5 hidden layer ANN is only the top performer in the early stages of the training. The 5 hidden layer ANN displays rapid improvement in MSE during the first 30,000 iterations at which point the 4 hidden layer network begins to outperform, followed by the 3 hidden layer network at approximately 55,000 iterations (Fig. 3(b)). Similar plots were made for the 84 and 252 nodes per layer configurations (not presented here), displaying similar trends as those seen in Fig. 3(b). For the 252 nodes per layer configurations, similar signs of overfitting were observed with 4 hidden layers. Based on these observations, an ANN with 3 layers and 168 nodes per layer was chosen as our model. The chosen ANN size results in low risk of overfitting, while providing sufficient complexity for generalization.

3.2. *Testing with diluted samples*

The results from the dilution-corrected sample sets are presented in Fig. 4, showing both the measured, dilution-corrected (expected), and ANN estimated VSF. Here, only the lowest concentration and the highest concentration samples are shown. The first sample set (Fig. 4(a)) was collected from Los Peñasquitos Lagoon, San Diego County, California, and will be referred to as the lagoon samples. The second sample set (Fig. 4(b)) was collected from glacial meltwater near King George Island, Antarctica [13], and will be referred to as the meltwater samples.

The ANN estimation of the true VSFs agrees well with the dilution corrected VSFs (Fig. 4). The two sample sets have noticeable differences in both the forward (<10°) and backward scattering angles, showcasing the generalization capabilities of the ANN. The ANN estimates deviate slightly from the expected VSFs, most noticeable for the lowest concentration sample in the meltwater samples where the ANN estimate is somewhat lower than the expected VSF. Here, the measured scattering coefficient is $b = 1.77\,\mathrm{m}^{-1}$, while the ANN's estimate is $b = 1.62\,\mathrm{m}^{-1}$. According to previous investigations into the effect of multiple scattering, the percentage error in the scattering coefficient at $b = 1.62\,\mathrm{m}^{-1}$ is likely around 10% [18]. In this case, the ANN

Optics EXPRESS

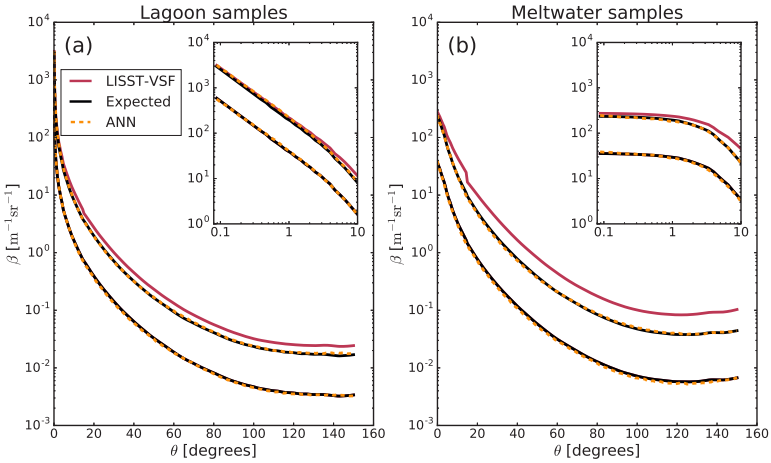

Fig. 4. VSFs obtained from LISST-VSF measurement (red), expected VSFs from dilution correction (black), and ANN estimated VSFs (dashed yellow). The insets show the log-log plot of the VSFs for angles $\theta \leq 10$. Highest (top black curve) and lowest concentration (lower black curve) samples shown here. (a) Lagoon samples. Dilution factor of 10. (b) Meltwater samples. Dilution factor of 6.6.

estimate of b is about 9% lower than b measured by the LISST-VSF, which agrees with expectation. The multiple scattering errors observed in the LISST-VSF measurement at the lowest particle concentration will naturally propagate into the estimations of expected VSF at higher particle concentrations, partly explaining why the ANN estimated VSF is lower than the dilution corrected VSF for the meltwaters. For the lagoon samples, this problem is less severe, as the ANN estimated scattering coefficient is only $b = 0.99\,\mathrm{m}^{-1}$, which is likely to contain a multiple scattering error of about $\sim 3\%$ [18].

The performance of the ANN can be quantified by comparing the scattering coefficients b and backscattering coefficients b_b obtained from the original (uncorrected) LISST-VSF measurement, dilution-corrected (expected) LISST-VSF measurement, and VSF estimated from application of the ANN to the original LISST-VSF measurement. In Fig. 5, the scattering and backscattering coefficients are plotted against the expected values from the dilution correction for both sample sets. Here, the expected values are plotted with an adjustment for expected multiple scattering error in the LISST-VSF measurements with lowest particle concentration, accounting for $\sim 3\%$ and $\sim 10\%$ error in the lagoon and meltwater samples, respectively.

For increasing scattering coefficients, the measured scattering and backscattering coefficients deviate increasingly from the expected values for both sample sets, as seen in Fig. 5. This is most noticeable for the meltwater samples (see Fig. 5(c)), where the highest concentration sample has a measured scattering coefficient of $b = 21.5\,\mathrm{m}^{-1}$, which is more than twice the expected value of $b = 10.6\,\mathrm{m}^{-1}$. On the other hand, the ANN estimates a scattering coefficient of $b = 11.1\,\mathrm{m}^{-1}$, corresponding to a reduction in percent deviation from 102.8% to 4.7%. Here, the percentage deviation is calculated according to Eq. (7), where the reference value X_r is b expected from dilution correction, while X_m is b obtained from LISST-VSF measurement and from application of ANN, respectively. In general, the ANN estimates are very similar to the expected values, with both sample sets showing an increase in deviation at the highest concentration for both b and b_b. The largest deviation is observed in the backscattering coefficient for the highest concentration sample in the meltwater samples. Here, the percentage deviation is 13%, which is still a large improvement from the deviation of 138.8% in the measured b_b.

Optics EXPRESS

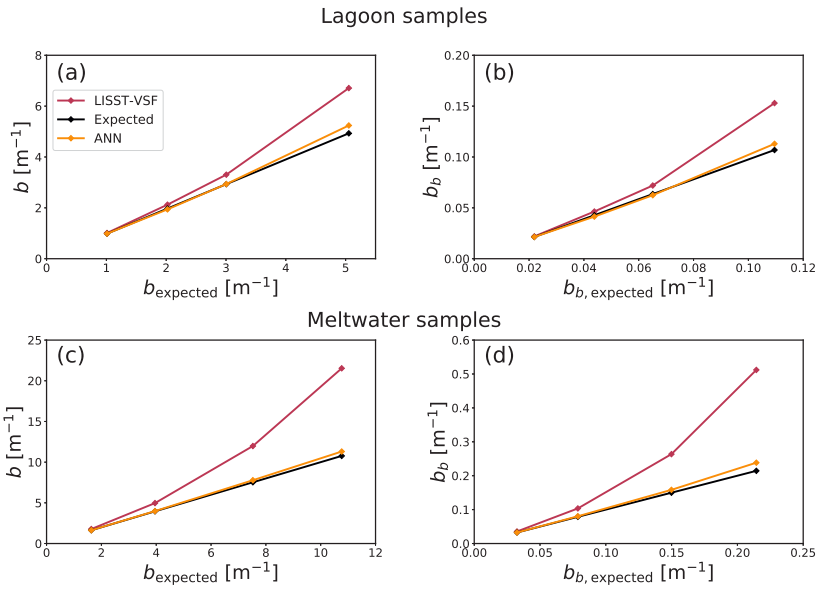

Fig. 5. Scattering coefficients and backscattering coefficients plotted against the expected values for the lagoon samples (a and b), and the meltwater samples (c and d). Values measured by LISST-VSF (red), expected from dilution correction (black, 1:1 line), and estimated by the ANN (yellow).

There are several possible reasons for the observed trend in deviation between expected and ANN estimated VSF. For these samples specifically, the dilution factor is considered to be a significant source of uncertainty. For the meltwater samples, the dilution was done subsequently, meaning that any error in the first dilution would propagate to the next dilution. On the other hand, the samples in the lagoon samples were diluted individually. These differences in dilution methodology could explain why the ANN estimate for the lagoon samples is observed to be both below and above the expected value (random error), while the ANN estimate for the meltwater samples is always above the expected value (systematic error). Also, due to increased measurement uncertainty in the very near forward direction ($< 1°$), measurement results in this angular region were treated using extrapolation and smoothing routines. In a previous study, we have shown that the multiple scattering error is dependent on the shape of the VSF [18]. Specifically, the extreme forward part seems to be of special importance. Hence, the VSF shape resulting from extrapolation and smoothing at $\theta < 1°$ may be slightly inconsistent with the multiple scattering error in these measurements, resulting in deviations between the expected and ANN estimated VSFs. Furthermore, the results presented here are particularly sensitive to errors in the lowest concentration samples, as the expected VSF is directly calculated from these.

Other possible reasons to the observed deviation are of a more general form. The ANN is trained on simulated data, meaning that the ANN can only be as good as the simulation. While the simulation has been shown to be generally accurate, there are noticeable deviations at large scattering coefficients [17]. Another likely reason is that these samples are not adequately represented in the training data. Only 18 of the 154 phase functions used for training are obtained from measurements, meaning that the majority of the training set is obtained from artificial phase functions, e.g. Henyey-Greenstein and Fournier-Forand. By including more VSFs obtained from measurements of natural samples, the performance of the ANN is very likely to improve. However, due to limited available VSF measurements, data for training and data for testing had to be balanced. As more *in situ* data becomes available, additional samples can be added to

Optics EXPRESS

the training dataset. The final factor to consider is uncertainty in the measured data, where all individual sample measurements contain errors of varying degree, in addition to multiple scattering errors [11,15].

3.3. Testing with measurements of natural seawater samples

A large dataset consisting of 98 VSFs obtained from measurements on natural samples are used for further evaluation of the ANN. In Fig. 6(a), both the measured and ANN estimated b is plotted against measured c. The same data is also presented in Fig. 6(b), but here plotted against the ANN estimated b.

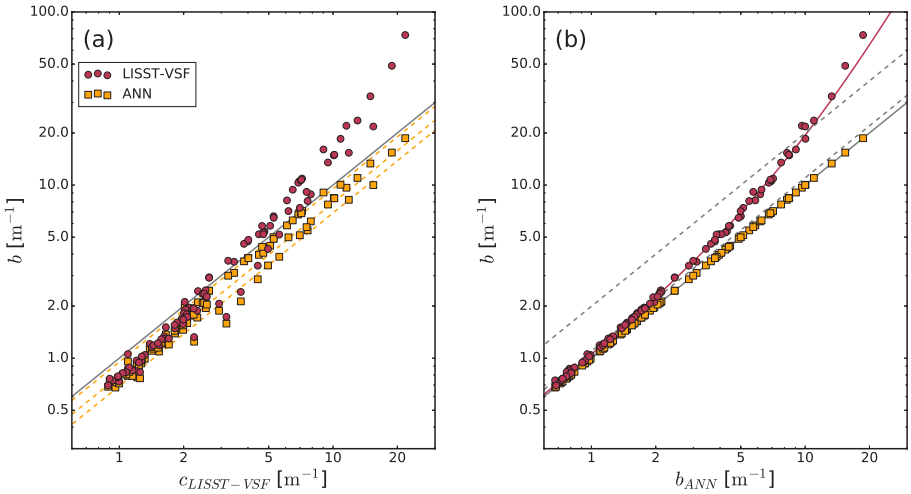

Fig. 6. (a) Measured (red) and ANN estimated (yellow) scattering coefficient b plotted against the measured attenuation coefficient c. The 1:1 line is plotted in solid grey, and the yellow dashed lines represent the median ratio of $b/c = 0.79$, together with the 10th percentile ($b/c = 0.69$) and 90th percentile ($b/c = 0.97$). (b) Measured (red) and ANN estimated (yellow) scattering coefficient b plotted against the ANN estimated scattering coefficient. The dashed grey lines show the 10% and 100% deviation between measured and ANN estimated b. The 1:1 line is plotted in solid grey.

From the relation $c = a + b$, we must always have $b < c$ to have a positive absorption coefficient. Thus, all measurement points plotted in Fig. 6(a) should be below the 1:1 line (grey). However, this is not the case for many of the LISST-VSF measurements, where $b > c$ can be observed for attenuation coefficients $c > 2\,\mathrm{m}^{-1}$. On the other hand, the ANN estimates of b are always below the 1:1 line. The ANN estimated b follows a linear trend with a median single scattering albedo b/c of 0.79. The observed variations around the median are expected due to natural variability in the optical properties of natural assemblages of marine particles. In Fig. 6(b), the same data are plotted as a function of the ANN estimated b. Here, a clear relationship is observed between measured and ANN estimated b, where the percentage deviation follows a trend that can be approximated by a power law fit (see Fig. 7(a)).

The estimated adjustment after applying the ANN can be assessed by looking at the deviation between the original LISST-VSF measurement and ANN estimates. In Fig. 7, the percentage deviation is plotted for the scattering coefficient b, backscattering coefficient b_b and backscattering ratio $\tilde{b}_b = b_b/b$. The latter may be of special interest, as $\tilde{b}_b$ describes the VSF shape, rather than magnitude. Here, the percentage deviation is calculated according to Eq. (7), where X_r is the ANN estimate, and X_m is the measured value.

Optics EXPRESS

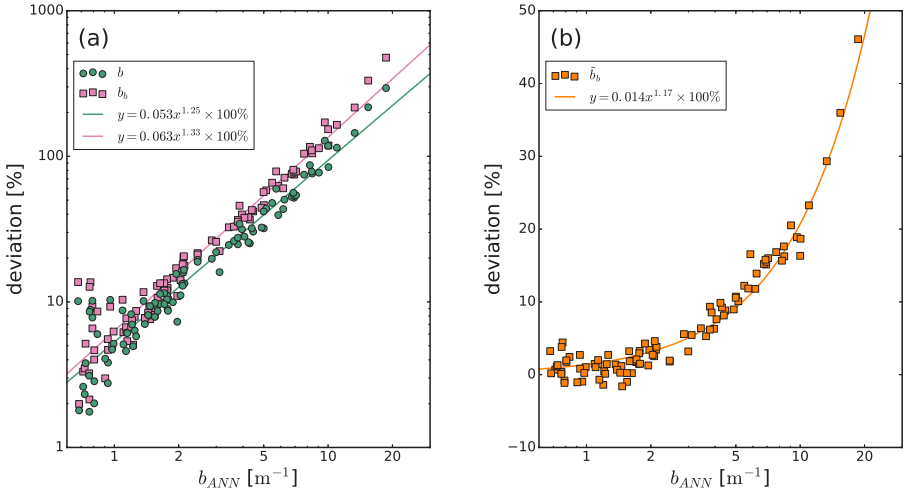

Fig. 7. Percentage deviation plotted against ANN estimated b for (a) the scattering coefficient b (green) and backscattering coefficient b_b (pink), and (b) the backscattering ratio (orange). Percentage deviations are calculated according to Eq. (7), where X_r is the ANN estimate and X_m is the measured value. Note that panel (a) is a log-log plot, while panel (b) is a semi-log plot.

As seen in Fig. 7(a), the deviations in both b and b_b follow a trend that can be approximated by a power law fit, where the deviation in b_b is larger than for b. From the trend lines, the percentage deviation in b reaches 10% at approximately $b = 1.7\,\mathrm{m}^{-1}$ and 100% at $b = 10.5\,\mathrm{m}^{-1}$, while for b_b it is 10% at $b = 1.4\,\mathrm{m}^{-1}$ and 100% at $b = 8.0\,\mathrm{m}^{-1}$. These results are consistent with our previous study on multiple scattering errors in LISST-VSF measurements [18]. The results presented in Fig. 7(a) also illustrate the dramatic increase in measurement error when the scattering coefficient becomes large ($b > 10\,\mathrm{m}^{-1}$). The highest concentration sample in this dataset has an ANN estimated scattering coefficient of $b = 18.7\,\mathrm{m}^{-1}$, and the percent deviation is 294% for b and 476% for b_b.

From Fig. 7(b), the deviation in the backscattering ratio $\tilde{b}_b$ is less severe than that observed for b and b_b. Here, the deviation is 10% at $b \approx 5\,\mathrm{m}^{-1}$, reaching 46% for the highest concentration sample at $b = 18.7\,\mathrm{m}^{-1}$. The data points generally follow an exponential fit, but as with b and b_b, the fit is slightly too high for $b < 2\,\mathrm{m}^{-1}$. Here, a positive deviation in $\tilde{b}_b$ means that the increase in the measured VSF β due to multiple scattering is relatively larger in the backwards direction than in the forward direction. This is also true for the phase function, which is calculated as $\tilde{\beta} = \beta/b$.

4. Summary and conclusion

In this study, we have developed an artificial neural network (ANN) with the aim of correcting multiple scattering errors in VSFs measured by the LISST-VSF instrument. It is a traditional feed forward network, containing 3 hidden layers with 168 nodes layer. The ANN estimates the true VSF taking only the measured VSF as input, eliminating the necessity for additional measurements. Training data was generated using a previously developed Monte Carlo simulation of the LISST-VSF instrument, verified by laboratory measurements [17]. A large and varied training set was constructed, consisting of VSFs simulated using Henyey-Greenstein, Fournier-Forand, and two-term Reynolds-McCormick phase functions, in addition to some phase functions obtained from *in situ* measurements. A total of 154 unique phase functions were simulated with

Optics EXPRESS

9 different scattering coefficients ranging from $b = 0.05 - 50\,\mathrm{m}^{-1}$, resulting in a training set consisting of 1386 VSFs.

A benchmark test of the ANN was performed using measurements with dilution corrected samples. The results showed that the VSFs estimated by the ANN were very similar to the expected VSFs in both shape and magnitude. The performance of the ANN was evaluated in terms of the scattering and backscattering coefficients. Both sample sets showed similar trends, where the deviation between LISST-VSF measurement and expected coefficients increased significantly with increasing particle concentration. On the other hand, the ANN estimates were very similar to the expected values, with a slight increase in deviation with increasing concentration. For the highest concentration sample, the expected scattering coefficient was $b = 10.6\,\mathrm{m}^{-1}$. In this case, applying the ANN correction reduced the scattering coefficient from a measured value of $b = 21.5\,\mathrm{m}^{-1}$ to the ANN estimated value of $b = 11.1\,\mathrm{m}^{-1}$, equivalent to a reduction in error from 103% to 5%.

Further testing was performed with a large database of natural seawater samples from various oceanic environments, for which measurements were made with original samples without any dilutions. The sample set contained 98 individual measurements with measured attenuation coefficients ranging from $c = 0.64 - 21.82\,\mathrm{m}^{-1}$. Here, the ANN was evaluated by calculating the scattering coefficients for the measured and ANN estimated VSF, and comparing them to the attenuation coefficients measured by the LISST-VSF. From the relation $c = a + b$, the expected b must be lower than the measured c. However, for samples with an attenuation coefficient greater than $c = 4\,\mathrm{m}^{-1}$, the measured b was consistently larger than the measured c, while instances of $b > c$ were observed at attenuation coefficients as low as $c = 2\,\mathrm{m}^{-1}$. The ANN was effective in reducing this discrepancy, with all estimates satisfying the criterion $b < c$. Furthermore, reductions in VSF following application of the ANN were more significant for the backscattering coefficient as compared with the scattering coefficient, typically resulting in a reduction of the backscattering ratio of about 5–20% for $b = 2 - 10\,\mathrm{m}^{-1}$.

The ANN developed in this study has been shown to correct multiple scattering errors in VSF measurements collected by the LISST-VSF for a wide range of phase functions and scattering coefficients. While undesirable multiple scattering errors arise for scattering coefficients $b > 2\,\mathrm{m}^{-1}$, the ANN provides reduced uncertainty in derived optical properties far above $b = 2\,\mathrm{m}^{-1}$. This significantly increases the turbidity range the LISST-VSF can reliably operate in, allowing for more accurate measurements in particle rich waters such as phytoplankton blooms, and coastal environments including river outlets, harbors and glacial meltwaters.

While the ANN has been evaluated on a large dataset of VSFs measured in natural waters, further evaluation through comparison with other instruments and additional serial dilutions is desirable.

Funding. Universitetet i Bergen.

Acknowledgements. Data from Storfjorden, Svalbard were collected and processed in the project Useful Arctic Knowledge (project no. 274891), and made freely available by NERSC under projects INTAROS (GA No. 727890) and Norwegian Scientific Data Network (project no. 245967). The second author (DK) received support from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Marie Skłodowska- Curie grant agreement No. 101034309. Measurements from Hardangerfjorden and Gaupnefjorden were collected as part of the EcoSens project (project no. 303190).

Disclosures. The authors declare no conflicts of interest.

Data availability. The artificial neural network (ANN) can be downloaded in both Python and MATLAB format from GitHub [45]. Other data, such as the training and validation data are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. P. J. Werdell, L. I. W. McKinna, E. Boss, S. G. Ackleson, S. E. Craig, W. W. Gregg, Z. Lee, S. Maritorena, C. S. Roesler, C. S. Rousseaux, D. Stramski, J. M. Sullivan, M. S. Twardowski, M. Tzortziou, and Z. Xiaodong, "An

Optics EXPRESS

overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing," Prog. Oceanogr. **160**, 186–212 (2018).
2. J. Chowdhary, P.-W. Zhai, E. Boss, H. Dierssen, R. Frouin, A. Ibrahim, Z. Lee, L. A. Remer, M. Twardowski, F. Xu, X. Zhang, M. Ottaviani, W. R. Espinosa, and D. Ramon, "Modeling atmosphere-ocean radiative transfer: A pace mission perspective," Front. Earth Sci. **7**, 100 (2019).
3. Y. C. Agrawal, A. Whitmire, O. A. Mikkelsen, and H. C. Pottsmith, "Light scattering by random shaped particles and consequences on measuring suspended sediments by laser diffraction," J. Geophys. Res.: Oceans **113**(C4), C04023 (2008).
4. B. Cochenour, K. Dunn, A. Laux, and L. Mullen, "Experimental measurements of the magnitude and phase response of high-frequency modulated light underwater," Appl. Opt. **56**(14), 4019–4024 (2017).
5. Z. Zeng, S. Fu, H. Zhang, Y. Dong, and J. Cheng, "A survey of underwater optical wireless communications," IEEE Commun. Surv. Tutorials **19**(1), 204–238 (2016).
6. C. D. Mobley, *Light and water: radiative transfer in natural waters* (Academic press, 1994).
7. T. Harmel, J. Agagliate, M. Hieronymi, and P. Gernez, "Two-term reynolds–mccormick phase function parameterization better describes light scattering by microalgae and mineral hydrosols," Opt. Lett. **46**(8), 1860–1863 (2021).
8. C. Supplis, J. Dauchet, V. Gattepaille, F. Gros, T. Vourc'h, and J.-F. Cornet, "Radiative analysis of luminescence in photoreactive systems: Application to photosensitizers for solar fuel production," PLoS One **16**(7), e0255002 (2021).
9. C. D. Mobley, L. K. Sundman, and E. Boss, "Phase function effects on oceanic light fields," Appl. Opt. **41**(6), 1035–1050 (2002).
10. T. J. Petzold, "Volume scattering functions for selected ocean waters," (1972).
11. M. Jonasz and G. Fournier, *Light scattering by particles in water: theoretical and experimental foundations* (Elsevier, 2007).
12. X. Zhang, R. H. Stavn, A. U. Falster, D. Gray, and R. W. Gould Jr, "New insight into particulate mineral and organic matter in coastal ocean waters through optical inversion," Estuarine, Coastal Shelf Sci. **149**, 1–12 (2014).
13. D. Koestner, D. Stramski, and R. A. Reynolds, "Characterization of suspended particulate matter in contrasting coastal marine environments with angle-resolved polarized light scattering measurements," Appl. Opt. **60**(36), 11161–11179 (2021).
14. D. Koestner, D. Stramski, and R. A. Reynolds, "Measurements of the volume scattering function and the degree of linear polarization of light scattered by contrasting natural assemblages of marine particles," Appl. Sci. **8**(12), 2690 (2018).
15. H. Sandven, A. S. Kristoffersen, Y.-C. Chen, and B. Hamre, "In situ measurements of the volume scattering function with LISST-VSF and LISST-200X in extreme environments: Evaluation of instrument calibration and validity," Opt. Express **28**(25), 37373–37396 (2020).
16. L. Hu, X. Zhang, Y. Xiong, and M.-X. He, "Calibration of the LISST-VSF to derive the volume scattering functions in clear waters," Opt. Express **27**(16), A1188–A1206 (2019).
17. H. S. Ugulen, H. Sandven, B. Hamre, A. S. Kristoffersen, and C. Saetre, "Analysis of multiple scattering errors in LISST-VSF volume scattering function measurements using Monte Carlo simulations and experimental data," Opt. Express **29**(8), 12413–12428 (2021).
18. H. S. Ugulen, H. Sandven, B. Hamre, A. S. Kristoffersen, and C. Saetre, "Efficient monte carlo simulation reveals significant multiple scattering errors in underwater angular scattering measurements," Opt. Express **30**(7), 10802–10817 (2022).
19. J. Piskozub, D. Stramski, E. Terrill, and W. K. Melville, "Influence of forward and multiple light scatter on the measurement of beam attenuation in highly scattering marine environments," Appl. Opt. **43**(24), 4723–4731 (2004).
20. H. C. van de Hulst, "Multiple light scattering," (1980).
21. E. Dan Hirleman, "Modeling of multiple scattering effects in fraunhofer diffraction particle size analysis," Part. Part. Syst. Charact. **5**(2), 57–65 (1988).
22. J. Piskozub and D. McKee, "Effective scattering phase functions for the multiple scattering regime," Opt. Express **19**(5), 4786–4794 (2011).
23. R. A. Maffione and D. R. Dana, "Instruments and methods for measuring the backward-scattering coefficient of ocean waters," Appl. Opt. **36**(24), 6057–6067 (1997).
24. D. Doxaran, E. Leymarie, B. Nechad, A. Dogliotti, K. Ruddick, P. Gernez, and E. Knaeps, "Improved correction methods for field measurements of particulate light backscattering in turbid waters," Opt. Express **24**(4), 3615–3637 (2016).
25. O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," Heliyon **4**(11), e00938 (2018).
26. K.-S. Jeong, G.-J. Joo, H.-W. Kim, K. Ha, and F. Recknagel, "Prediction and elucidation of phytoplankton dynamics in the nakdong river (korea) by means of a recurrent artificial neural network," Ecol. Modell. **146**(1-3), 115–129 (2001).
27. Y. Chebud, G. M. Naja, R. G. Rivero, and A. M. Melesse, "Water quality monitoring using remote sensing and an artificial neural network," Water, Air, Soil Pollut. **223**(8), 4875–4887 (2012).

Optics EXPRESS

28. H. G. Kim, S. Hong, K.-S. Jeong, D.-K. Kim, and G.-J. Joo, "Determination of sensitive variables regardless of hydrological alteration in artificial neural network model of chlorophyll a: case study of nakdong river," Ecol. Modell. **398**, 67–76 (2019).
29. J. Anmala and T. Venkateshwarlu, "Statistical assessment and neural network modeling of stream water quality observations of green river watershed, ky, usa," Water Supply **19**(6), 1831–1840 (2019).
30. H.-S. Yi, S. Park, K.-G. An, and K.-C. Kwak, "Algal bloom prediction using extreme learning machine models at artificial weirs in the nakdong river, korea," Int. J. Environ. Res. Public Health **15**(10), 2078 (2018).
31. E. M. Alves, R. J. Rodrigues, C. dos Santos Corrêa, T. Fidemann, J. C. Rocha, J. L. L. Buzzo, P. de Oliva Neto, and E. G. F. Nú nez, "Use of ultraviolet–visible spectrophotometry associated with artificial neural networks as an alternative for determining the water quality index," Environ. Monit. Assess. **190**(6), 319 (2018).
32. D. Gebler, G. Wiegleb, and K. Szoszkiewicz, "Integrating river hydromorphology and water quality into ecological status modelling by artificial neural networks," Water Res. **139**, 395–405 (2018).
33. T. M. Tung Tiyasha and Z. M. Yaseen, "A survey on river water quality modelling using artificial intelligence models: 2000–2020," J. Hydrol. **585**, 124670 (2020).
34. M. Deo and C. S. Naidu, "Real time wave forecasting using neural networks," Ocean Eng. **26**(3), 191–203 (1998).
35. O. Makarynskyy, "Improving wave predictions with artificial neural networks," Ocean Eng. **31**(5-6), 709–724 (2004).
36. R. Kalra, M. Deo, R. Kumar, and V. K. Agarwal, "Rbf network for spatial mapping of wave heights," Mar. Struct. **18**(3), 289–300 (2005).
37. N. P. Juan and V. N. Valdecantos, "Review of the application of artificial neural networks in ocean engineering," Ocean Eng. **259**, 111947 (2022).
38. D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," Chemom. Intell. Lab. Syst. **39**(1), 43–62 (1997).
39. F. Chollet, "Keras," https://keras.io (2015).
40. M. Abadi, A. Agarwal, and P. Barham, *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," (2015). Software available from tensorflow.org.
41. G. J. Bowden, G. C. Dandy, and H. R. Maier, "Data transformation for neural network models in water resources applications," J. Hydroinf. **5**(4), 245–258 (2003).
42. I. Ioannou, A. Gilerson, B. Gross, F. Moshary, and S. Ahmed, "Deriving ocean color products using neural networks," Remote. Sens. Environ. **134**, 78–91 (2013).
43. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv, arXiv:1412.6980 (2014).
44. H. Sandven, T. Petit, A. Stallemo, E. B. V. den Bergh, F. Klockmann, H. R. Økland, J. Knutsen, K. T. Galtung, M. H. Malin Lunde, N. Eryilmaz, and T. S. Sagen, "Optical property measurements collected in storfjorden, svalbard, during the uak 2020 cruise," https://doi.org/doi:10.21335/NMDC-NERSC-1741962272 (2020).
45. H. Ugulen, "Correction of multiple scattering errors in lisst-vsf measurements," https://github.com/HavardUgulen/LISST-VSF-neural-network.git (2023).