

# Methods to retrieve the complex refractive index of aquatic suspended particles: going beyond simple shapes

A.-M. Sánchez<sup>1</sup> and J. Piera<sup>1</sup>

<sup>1</sup>Institute of Marine Sciences (ICM-CSIC), Physical and Technological Oceanography Department, Pg. Marítim de la Barceloneta, 37-49, 08003, Barcelona, Spain

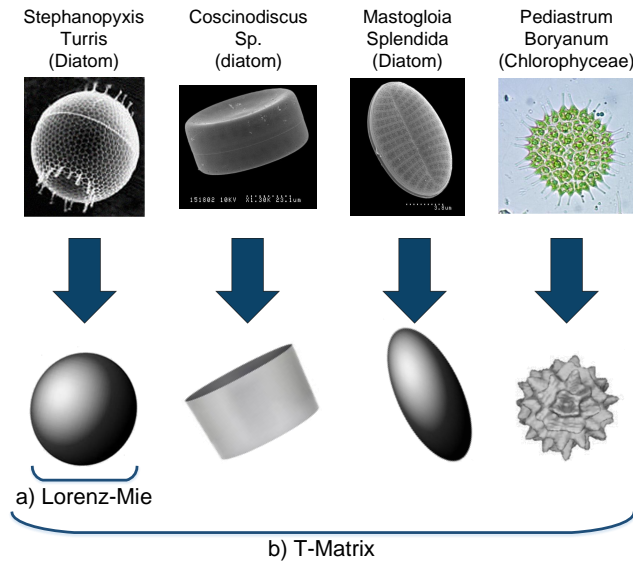
Correspondence to: A.-M. Sánchez (amsanchez@icm.csic.es)

**Abstract.** The scattering properties of aquatic suspended particles have many optical applications. Several data inversion methods have been proposed to estimate important features of particles, such as their size distribution or their refractive index. Most of the proposed methods are based on the Lorenz-Mie theory to solve Maxwell's equations, where particles are considered homogeneous spheres. A generalization that allows considering more complex shaped particles is the *T*-matrix method. Although this approach imposes some geometrical restrictions -particles must be rotationally symmetric- it is applicable to many life forms of phytoplankton. In this paper, three different scenarios are considered in order to compare the performance of several inversion methods for retrieving refractive indices. The error associated with each method is discussed and analyzed. The results suggest that inverse methods using the *T*-matrix approach are useful to accurately retrieve the refractive indices of particles with complex shapes, such as for many phytoplankton organisms.

## 1 Introduction

Light-particle interactions, usually wavelength dependent, cause observable optical phenomena (e.g., changes on the ocean color or light extinction with depth) that allow assessing the composition of small particles (e.g., phytoplankton, sediment or microplastics) in the water column (Mobley, 1994; Kirk, 1994). Understanding the interaction of light with particles is the central topic of many bio-optical studies where the water-particle composition is inferred from *in-situ* or remote-sensing optical observations (Gordon and Morel, 2012, and references therein).

Maxwell's equations are the basis of theoretical and computational methods describing light interaction with particles. However, exact solutions to Maxwell's equations are only known for selected geometries. Scattering from any homogeneous spherical particle of arbitrary size is explained analytically by the Lorenz-Mie (also known as Mie) theory (Lorenz, 1898; Mie, 1908). Although bio-optical models usually assume that particles are spheres, most particles that contribute significantly to light interactions are indeed non-spherical, with aspect ratios (ratio of the principal axes

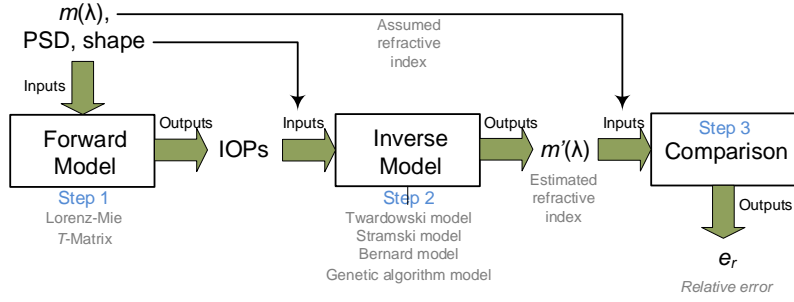


**Figure 1.** (a) The Lorenz-Mie method only describes the scattering of an electromagnetic plane wave by a homogeneous sphere, while (b) the *T*-Matrix method also characterizes the scattering by non-spherical particles such as spheroids, cylinders or Chebyshev particles. Both can be used on actual phytoplankton particles by using a suitable shape.

of a particle) spanning between 0.4 and 72 (Clavano et al., 2007, and references therein). For more complex shaped particles, scattering can be computed using the *T*-Matrix theory (Waterman, 1965), currently the fastest exact technique for the computation of non-spherical scattering based on a direct solution of Maxwell's equations (Mischenko et al., 1996). The *T*-matrix method has some geometrical restrictions, such as axial symmetry, but it is applicable to many life forms of phytoplankton and suspended mineral particles (Quirantes and Bernard, 2004; Sun et al., 2016), as shown in Fig. 1.

Both the Lorenz-Mie and *T*-Matrix strategies to solve Maxwell's equations share one important requirement: the complex refractive index of the particles must be known. There are classical methods to measure the refractive index, as the emergence of particles in oils of different refractive index to find the one in which minimal scattering occurs, or new techniques such as Tomographic Phase Microscopy (Choi et al., 2007). However, measuring absorption and doing modeling is one of the most used methods to estimate the imaginary part of the refractive index (Aas, 1996).

Several inverse models to retrieve the refractive index from optical measurements can be found in the literature. For instance, a single equation based on the Lorenz-Mie theory was used by Twardowski et al. (2001) to estimate the real part of the refractive index of a bulk oceanic distribution. It is indeed a fast method if optical backscattering measurements are available. Stramski et al. (1988) presented an extension of a model from Bricaud and Morel (1986), designed for isolated phytoplankton cultures (or dominated by one particular phytoplankton species). It is based on the anomalous-diffraction approximation (ADA), which allows computing the real and imaginary parts



**Figure 2.** Procedure used to analyze the accuracy of the inverse models.

of the complex refractive index as separate variables, using only the absorption and attenuation efficiency factors and the concurrent measurement of the particle-size distribution (PSD). Bernard et al. (2001) simplified this model by replacing the Lorentzian oscillators with a simple Hilbert transform. All these methods share one thing in common: they approximate the shape of the particles by homogeneous spheres. Meyer (1979) first and Bernard et al. (2009) later suggested that two-layered spherical-geometry models reproduce more accurately the measured algal angular-scattering properties. Finally, a combination of a genetic algorithm with the Lorenz-Mie and *T*-Matrix approaches was used by Sánchez et al. (2014), thereby allowing the study of more complex structures than simple homogeneous or coated spheres. A genetic algorithm is a method to solve problems simulating the process of natural selection using inheritance, mutation, selection and crossover between different possible solutions. This method only requires the measured attenuation and scattering coefficients, together with the PSD, to find the complex refractive index. It is much slower than the previous ones (in particular, for non-spherical particles), but it can provide very accurate estimations.

In this paper, the above refractive-index retrieval models are reviewed and tested against simulated data in order to analyze their accuracy when modeling real (and usually complex-shaped) particles suspended in water, such as phytoplankton. The comparison has been done following the three steps presented in Fig. 2. First, the forward models (basically, Lorenz-Mie and *T*-Matrix) are used to obtain the inherent optical properties (IOPs) of a selected configuration using as inputs the postulated wavelength-dependent refractive index,  $m(\lambda)$ , the PSD and the particle shape. Second, the above inverse models are used to estimate the refractive index from the IOPs along with the PSD and the particle shape. Finally, the estimated refractive index is compared with the postulated one in order to assess the accuracy of the inverse model.

The simulated examples are implemented using complex refractive indices and PSDs similar to those found in nature for phytoplankton species. Since phytoplankton particles exhibit a wide variety of shapes, each example has been provided with a different outline, accounting for homogeneous sphere, a coated sphere and homogeneous cylinder. None of these idealized shapes is an exact representation of real algae presenting cell walls, chloroplasts, vacuoles, nuclei and other internal

organelles, each one with its own optical properties. However, they can be considered as a first approximation suitable for the purposes of the tests presented in this contribution.

It must also be noted that the models are fundamentally different. The model developed by Twardowski et al. (2001) is intended to be used for entire particle populations that are assumed to follow a power-law size distribution whereas the other models are developed for single phytoplankton cultures (or dominated by one particular phytoplankton species) and require the concurrent measurement of the size distribution. These bio-optical models are also compared with a numerical method (i.e., the genetic algorithm) in the same conditions. On the other hand, our approach allows an objective comparison of the results from the different methods in those occasions when no single optimum methodology is clearly identifiable, a tool of potentially high interest for the ocean optics community.

In order to establish the foundations of the work here presented, Section 2 reviews the formulation used to obtain the IOPs from the Lorenz-Mie and  $T$ -Matrix characterizations (which perform the forward calculations) for poly-dispersed algal assemblages. In Section 3, a review of the different inverse approximations to retrieve the refractive index is carried out. In Section 4, all the models are used to retrieve the refractive index of particles, with three different proposed shapes, in poly-disperse assemblages. Section 5 discusses the results and, finally, the conclusions are drawn in Section 6.

## 2 Model theory

### 2.1 Size distributions and poly-dispersions

Algal assemblages are typically poly-dispersed with regard to size, and can be described by a PSD  $F(D)$ , where  $D$  is the particle diameter, and  $F(D)d(D)$  is the number of particles per unit volume in the size range  $D \pm 1/2d(D)$  (Bricaud and Morel, 1986). Using absorption as an example (analogous expressions may be used for other IOPs, i.e., those related to light extinction and scattering), the absorption-efficiency factor representing the mean of a size distribution can be described by (Bricaud and Morel, 1986)

$$\bar{Q}_a(\lambda) = \frac{\int_0^\infty Q_a(\lambda, D) F(D) D^2 d(D)}{\int_0^\infty F(D) D^2 d(D)}, \quad (1)$$

where  $\lambda$  is the wavelength. The total absorbed power per unit incident irradiance and unit volume of water, i.e., the absorption coefficient, is then either given by

$$a(\lambda) = \frac{\pi}{4} \int_0^\infty Q_a(\lambda, D) F(D) D^2 d(D) \quad [m^{-1}], \quad (2)$$

or, using the result of Eq. (1), by

$$a(\lambda) = \frac{\pi}{4} \bar{Q}_a(\lambda) \int_0^\infty F(D) D^2 d(D) \quad [m^{-1}]. \quad (3)$$

## 2.2 Inherent optical properties

Lorenz-Mie and  $T$ -Matrix theories are powerful methods to formulate an analytical solution to electromagnetic scattering by spherical and non-spherical particles. Both rely on the expansion of the incoming light into spherical harmonics and use an intensive formulation to compute the coefficients that link the incident field with the scattered and transmitted ones. The complete Lorenz-Mie derivation is reviewed by Bohren and Huffman (1998), and the  $T$ -Matrix approach is described by Mischenko et al. (1996). Both theories provide the particle specific optical properties, i.e., the extinction, scattering and absorption cross sections that describe the fraction of the incident-beam intensity converted to extinct, scattered or absorbed light,  $C_e$ ,  $C_s$  and  $C_a$ , respectively, in terms of effective area. The relationship between wavelength-dependent efficiency factors and cross sections is

$$Q_x(\lambda) = \frac{C_x(\lambda)}{\langle G \rangle}, \quad (4)$$

being  $x$  any of the subindices  $e$ ,  $s$  or  $a$  to denote extinction, scattering or absorption, and  $G$  the geometric cross section of the particle (in the case of nonspherical particles,  $\langle G \rangle$  is the averaged cross-sectional area over all orientations). The cross sections obtained from the Lorenz-Mie and  $T$ -Matrix approaches (size-averaged in poly-disperse concentrations as described above) can also be used to compute the extinction, scattering and absorption coefficients ( $c(\lambda)$ ,  $b(\lambda)$  and  $a(\lambda)$ , respectively) as (Quirantes and Bernard, 2006)

$$c(\lambda) = N \cdot \langle C_e(\lambda) \rangle \quad [m^{-1}], \quad (5)$$

$$b(\lambda) = N \cdot \langle C_s(\lambda) \rangle \quad [m^{-1}], \quad (6)$$

$$a(\lambda) = N \cdot \langle C_a(\lambda) \rangle \quad [m^{-1}], \quad (7)$$

where  $N$  denotes the number of particles per unit volume. The relationship between the three parameters is

$$c(\lambda) = a(\lambda) + b(\lambda) \quad [m^{-1}]. \quad (8)$$

Scattering can be further characterized in terms of the angular distribution of the scattered light using the volume scattering function ( $\beta$ ) as (Mobley, 1994)

$$130 \quad \beta(\Psi, \lambda) = \tilde{\beta}(\Psi, \lambda) \cdot b(\lambda) \quad [m^{-1} sr^{-1}]. \quad (9)$$

where  $\Psi$  is the scattering angle, i.e., the angle between the incident and scattered beams, and  $\tilde{\beta}(\Psi, \lambda)$  is the scattering phase function and the (1,1) element of the Stokes scattering matrix (or Mueller matrix). This matrix, obtained with the Lorenz-Mie and  $T$ -Matrix formulation when the physical characteristics of the particles are known, transforms the Stokes parameters of the incident  
 135 light into those of the scattered light. By integrating the volume scattering function in all directions, assuming azimuthal symmetry, the total scattering coefficient  $b$  is obtained as

$$b(\lambda) = 2\pi \int_0^\pi \beta(\Psi, \lambda) \sin(\Psi) d\Psi \quad [m^{-1}], \quad (10)$$

which can be partitioned into its forward and backward components (respectively  $b_f$  and  $b_b$ ) by limiting the integration limits from 0 to  $\pi/2$  and from  $\pi/2$  to  $\pi$ , respectively. The backscatter fraction,  
 140 defined by

$$B_b(\lambda) = \frac{b_b(\lambda)}{b(\lambda)}, \quad (11)$$

gives the fraction of scattered light that is deflected through the scattering angles above  $\pi/2$ . Given Eq. (9) and Eq. (10), the normalization condition for the volume scattering phase function is:

$$2\pi \int_0^\pi \tilde{\beta}(\Psi, \lambda) \sin(\Psi) d\Psi = 1. \quad (12)$$

145 This normalization implies that the backscatter fraction can be computed using the volume scattering phase function as

$$B_b(\Psi, \lambda) = 2\pi \int_{\pi/2}^\pi \tilde{\beta}(\Psi, \lambda) \sin(\Psi) d\Psi. \quad (13)$$

The  $2\pi$  factor used in Eq. (10), Eq. (12) and Eq. (13) arises naturally after integration with respect to the azimuth angle. Notice that the same  $2\pi$  factor is also used by Twardowski et al. (2001), Bohren  
 150 and Huffman (1998) and in most of the literature on ocean optics (Mobley, 1994), but differs from the  $1/2$  factor used by Mischenko et al. (1996), Mischenko and Travis (1998), Wiscombe and Grams (1976) and Mugnai and Wiscombe (1986), where the integration of phase function is normalized to  $4\pi$ , representing the total solid angle over the whole sphere.

### 3 Review of refractive index retrieval models

In this section, a review of the different approximations to retrieve the refractive index (inverse models) is presented. Each model is named after the lead author of the publication. The complex refractive index  $m(\lambda)$  is defined as

$$m(\lambda) = n(\lambda) + ik(\lambda), \quad (14)$$

where the real part  $n(\lambda)$  determines the phase velocity of the propagating wave, and the imaginary part  $k(\lambda)$  determines the flux decay. The sign of the complex part is a matter of convention as it may also be defined with a negative sign. The above notation corresponds to waves with time evolution given by  $e^{-i\omega t}$ . Notice that, throughout this paper, the effective refractive indices are computed relative to seawater, which has a constant value of  $m_{water} = 1.334 + i0$  (this value varies with wavelength, salinity, temperature and pressure, Hale and Querry, 1973). Thus, values relative to free vacuum can be obtained by  $m_a = m \times m_{water}$ .

#### 3.1 The Twardowski model

The Twardowski model (Twardowski et al., 2001) is based on Volz (1954) as cited in van de Hulst (1957). It is derived using the Lorenz-Mie theory and the relationship between the particulate spectral attenuation ( $c_P(\lambda)$ ) and the size distribution to retrieve the bulk particulate refractive index from *in situ* optical measurements. In particular, it assumes that  $\gamma = \xi - 3$  ( $\gamma$  is the hyperbolic slope of the attenuation coefficient and  $\xi$  is the power-law slope of the PSD). It only considers power-law distributions that fulfill the conditions  $2.5 \leq \xi \leq 4.5$  and  $0 \leq B_b \leq 0.03$ . The bulk refractive index is obtained from a polynomial fit to the Lorenz-Mie calculations

$$\hat{n}(B_b, \gamma) = 1 + B_b^{0.5377 + 0.4867\gamma^2} (1.4676 + 2.2950\gamma^2 + 2.3113\gamma^4). \quad (15)$$

The link between  $\gamma$  and  $\xi$  is only exact for homogeneous sphere particles with sizes from 0 to infinity, having wavelength-independent constant absorption ( $k$  is held constant at 0.005). It was first tested by Boss et al. (2001a) and refined in Boss et al. (2001b). It should be noted that the calculations took into account the acceptance angle of the AC9, a WET Labs instrument used for the measurement of the extinction coefficient, which is 0.93 degrees. The total scattering coefficient value from the *insitu* measurements was derived by subtracting the measured absorption from the measured extinction, so the model was formulated to be consistent with the measurements (with  $b$  serving as an integrated scattering from 0.93 to 180 degrees instead of 0 to 180 degrees). Even though this was not considered by Twardowski et al. (2001), it was later taken into account by Boss et al. (2004), but without recomputing the regression. This means that some inaccuracies can be expected when using ideal models.

### 3.2 The Stramski model

This model is based on the methods presented by Stramski et al. (1988), which are an extension of those developed by Bricaud and Morel (1986). It is based on the ADA, first described in van de Hulst (1957). The ADA offers approximations to the absorption and attenuation optical efficiency factors using relatively simple algebraic formulae, based on the assumptions that the particle is large relative to the wavelength ( $\alpha = \frac{\pi D}{\lambda} \gg 1$ ) and that the refractive index is small ( $n - 1 \ll 1$  and  $k \ll 1$ ). This method allows decoupling the effects of the real and imaginary refractive indices on absorption and scattering. Assuming an homogeneous geometry, the ADA expression for the absorption efficiency factor is given by

$$Q_a(\rho') = 1 + \frac{2e^{-\rho'}}{\rho'} + 2\frac{e^{-\rho'} - 1}{\rho'^2}, \quad (16)$$

where  $\rho' = 4\alpha k$  is the absorption optical thickness. Eq. (3) and Eq. (16) are then used iteratively to determine the homogeneous imaginary part of the refractive index ( $k(\lambda)$ ) in conjunction with measured algal absorption and PSD data. According to the Ketteler-Helmholtz theory of anomalous dispersion (van de Hulst, 1957), a variation in  $k$  induces variations in  $n$ , quantified with a series of oscillators (representing discrete absorption bands) based on the Lorentz-Lorentz equations (Stramski et al., 1988; Bricaud and Morel, 1986). These spectral variations (denoted as  $\Delta n(\lambda)$ ) vary around a central value of the real refractive index,  $1 + \epsilon$ . Thus

$$n(\lambda) = 1 + \epsilon + \Delta n(\lambda). \quad (17)$$

The central value  $1 + \epsilon$  is estimated by computing the non-absorbing equivalent-population attenuation efficiency factor ( $\bar{Q}_c^{NAE}$ ) at those wavelengths where  $\Delta n(\lambda_\epsilon) = 0$ . Considering polydispersion, this is done according to

$$\bar{Q}_c^{NAE}(\bar{\rho}) = \frac{\int_0^\infty Q_c(\rho) F(\rho) \rho^2 d(\rho)}{\int_0^\infty F(\rho) \rho^2 d(\rho)}, \quad (18)$$

where  $\rho = 2\alpha(n - 1)$ , and  $F(\rho)$  is obtained from the experimental size distribution by replacing  $D$  by  $\rho$ , and calculating  $Q_c(\rho)$  with the van de Hulst's formula assuming  $\xi = 0$  (van de Hulst, 1957)

$$Q_c(\rho) = 2 - \frac{4}{\rho} \sin \rho + \frac{4}{\rho^2} (1 - \cos \rho). \quad (19)$$

The exact value of  $\epsilon$  is indicated by a value of  $\bar{Q}_c^{NAE}(\bar{\rho})$  that is equal to  $\bar{Q}_c(\lambda_\epsilon)$ .

This methodology was subsequently simplified by Bernard et al. (2001, 2009) using the Kramers-Kronig relations, instead of employing the Lorentzian oscillators, to compute the spectral variations

in the real part of the refractive index on the basis of the imaginary part. The Kramers-Kronig relations describe the mutual dependence of the real and imaginary parts of the refractive index through dispersion, as does the Ketteler-Helmholtz theory, but they are simpler than the tedious and sometimes inaccurate use of summed oscillators (the real part is the Hilbert transform of the imaginary part, van de Hulst, 1957).

### 3.3 The Bernard model

Meyer (1979) and Bernard et al. (2009) suggested that two-layered spherical geometry models reproduce more accurately the measured algal angular scattering properties. For Bernard et al. (2009) the outer layer accounts for the chloroplast and the inner layer for the cytoplasm. Refractive index values are assumed for the cytoplasm, with a spectral imaginary part modelled as

$$k_{cyto}(\lambda) = k_{cyto}(400)e^{[-0.01(\lambda-400)]}, \quad (20)$$

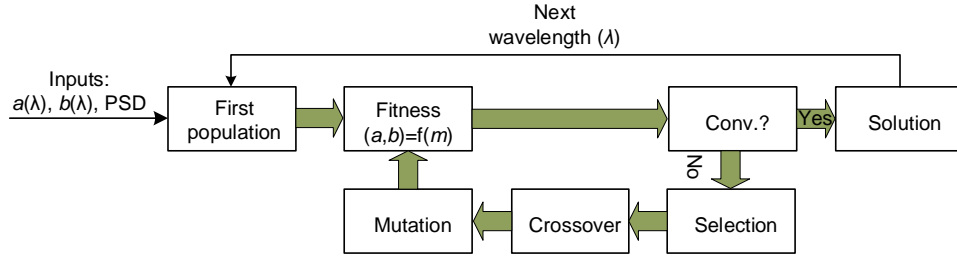
where  $k_{cyto}(400) = 0.0005$ . The real refractive index spectra for the cytoplasm,  $n_{cyto}(\lambda)$ , are obtained using the Hilbert transform (absorption has an influence on scattering and attenuation, expressed through the Kramers-Kronig relations) and Eq. (17) with  $1 + \epsilon = 1.02$ . Using  $k_{cyto}(\lambda)$  from Eq. (20), the volume equivalent values of  $k_{chlor}(\lambda)$  is determined using the Gladstone-Dale formulation

$$k_{chlor}(\lambda) = \frac{k_h(\lambda) - k_{cyto}(\lambda)V_V}{1 - V_V}, \quad (21)$$

where  $k_h(\lambda)$  is the imaginary part of the refractive index considering homogeneous cells and obtained using Eq. (16), and  $V_V$  is the relative chloroplast volume. According to Bernard et al. (2009), a  $V_V$  value of 20% can be considered as a first approximation for a spherical algal geometry, although higher values should be considered for the large celled dinoflagellate and cryptophyte samples. Other studies have employed relative chloroplast volumes of  $V_V = 41\%$  (Zaneveld and Kitchen, 1995),  $V_V = 58\%$  (Latimer, 1984), and  $V_V = 27\%$  to  $66\%$  (Bricaud et al., 1992). The real refractive index spectra for the chloroplast  $n_{chlor}(\lambda)$  is then generated by a Hilbert transform and using Eq. (17) with  $1 + \epsilon$  values between 1.044 and 1.14, depending on the sample.

### 3.4 The genetic algorithm model

The model presented by Sánchez et al. (2014) uses a genetic algorithm to find the refractive index that produces the desired scattering and absorption coefficients ( $a(\lambda)$  and  $b(\lambda)$ ), when using the Lorenz-Mie or  $T$ -Matrix approaches with the measured PSD. The methodology of the algorithm may be summarized as follows (Fig. 3). First, a random vector of solutions is generated for a specific wavelength ( $[m_1(\lambda_i), m_2(\lambda_i), \dots, m_n(\lambda_i)]$ , where  $\lambda_i$  denotes the selected wavelength



**Figure 3.** Flow chart for the genetic algorithm.

245 and  $m_1, m_2, \dots, m_n$  the complex refractive indices); if possible, the search space is bound in order to maximize the algorithm success. Next, the complete vector is evaluated by the fitness function. This is done by computing the  $a(\lambda)$  and  $b(\lambda)$  coefficients corresponding to each refractive index (using the Lorenz-Mie or  $T$ -Matrix formulation and Eq. (6) – Eq. (7)) and evaluating the weighted Euclidean distance between the calculated and desired coefficients. This can be done, for instance,  
 250 as

$$e_a(\lambda_i) = |20\log(\hat{a}_k(\lambda_i)) - 20\log(a_m(\lambda_i))|, \quad (22)$$

where  $a_k$  is the calculated absorption coefficient of the  $n_k$  refractive index,  $a_m$  is the desired (or measured) attenuation coefficient, and  $e_a$  is the error in the absorption coefficient. The use of logarithmic weighting improves the performance when dealing with small errors over small coefficients.

255 The same equation can be applied for the scattering coefficient. Both results are finally combined using a quadratic mean, thus obtaining a single evaluation value that is minimized by the algorithm.

After the evaluation, the algorithm stops if a satisfactory fitness level or a maximum number of generations is reached (each generation is a new vector of solutions). If the convergence condition is not fulfilled, the best solutions are selected and taken apart. Part of this elite is then recombined (crossover) and randomly mutated to provide genetic diversity and broaden the search space (crossover and mutation introduce the diversity needed to ensure that the entire sample space is reachable and to avoid suboptimal solutions, Greenhalgh and Marshall, 2000). The new set of solutions is re-evaluated and inserted again into the solutions' vector, which completes the cycle. After converging, the algorithm presents the best possible solution.

265 Since the Lorenz-Mie and  $T$ -Matrix algorithms can only be executed for single wavelengths and the refractive index is also wavelength dependent, with different values at different wavelengths, the genetic algorithm completes the search procedure at a single wavelength each time. After each convergence, the process starts over with the next wavelength-dependent value, eventually obtaining the complete complex-refractive-index signature.

270 The main advantage of this method is that it can be easily adapted to different Lorenz-Mie or  $T$ -Matrix codes, as for instance those developed for homogeneous spheres, coated spheres and cylin-

**Table 1.** Summary of the refractive index retrieval models

Inverse models	Inputs	Outputs	Type of particles	Assumptions	Equations
<b>Twardowski Model</b>	$B_b, \xi$	$n$	Homogeneous spheres	$k = 0.005$ , Power-law PSD, $\gamma = \xi - 3$	(9), (15)
<b>Stramski Model</b>	PSD, $a$ and $c$ (or $Q_a$ and $Q_c$ )	$n, k$	Homogeneous spheres	$\alpha \gg 1, n - 1 \ll 1, k \ll 1$	(13), (16), (17), (18), (19)
<b>Bernard Model</b>	$a$ , PSD	$n_{chlor}, k_{chlor}$	Coated spheres	$n_{cyto}, k_{cyto}$ , relative chloroplast volume ( $V_V$ ), $1 + \epsilon$	(16), (17), (20), (21)
<b>Genetic Algorithm Model</b>	$a, c$ , PSD	$n, k$	Homogeneous and coated spheres, spheroids, cylinders and Chebyshev particles	Particles with axial symmetry	(2), (3), (22)

ders. Besides, it can also be easily combined with other models to improve the results. However, it should be noted that some inversions may be ill-posed. A constrained optimization problem is considered to be well-posed (in the sense of Haddamard) if (a) a solution exists, (b) the solution is unique, and (c) the solution is well-behaved, i.e. varies continuously with the problem parameters. An ill-posed problem fails to satisfy one or more of these criteria (Bhandarkar et al., 1994); in this case, techniques such as regularization methods can be applied to improve the results (Mera et al., 2004).

### 3.5 Summary of the refractive-index retrieval models

This section has been summarized in Table 1, showing, for each model, its inputs and outputs, type of particles, as well as the assumptions of the model and the equations used.

## 4 Simulations

The models described in the previous section are used here to retrieve the refractive index of well-known particles in order to determine their accuracy by means of the averaged relative error defined

as:

$$e_{rx}(\%) = \frac{1}{N} \sum_{n=1}^N \left| \frac{x'(\lambda_n)}{x(\lambda_n)} - 1 \right| \times 100, \quad (23)$$

where  $x'$  is either the real part of the refractive index, estimated as  $n'(\lambda) - 1$  (the unity is subtracted to only consider the decimals), or the estimated imaginary part of the refractive index,  $k'(\lambda)$ , and  $x$  accounts for the postulated real part of the refractive index,  $n(\lambda) - 1$ , or the postulated imaginary part of the refractive index,  $k(\lambda)$ . For the volume scattering function, the error is also averaged with respect to its angular contribution as:

$$e_{rVSF}(\%) = \frac{1}{N \cdot M} \sum_{n=1}^N \sum_{m=1}^M \left| \frac{VSF'(\lambda_n, \theta_m)}{VSF(\lambda_n, \theta_m)} - 1 \right| \times 100, \quad (24)$$

First, Subsection 4.1 presents the results of the Twardowski, Stramski and genetic algorithm models applied to a simple spherical and homogeneous particle. Such particles, however, are not a good representation of phytoplankton particles. This is both because eukaryotic phytoplankton cells are heterogeneous particles, with membrane systems and intracellular organelles, and because most of the phytoplankton species are not spherical. As stated by Bernard et al. (2009), the spherical structure fails mainly at describing the backward scattering, which suggests that a two-layered spherical geometry is the simplest possible heterogeneous structure capable of properly reproducing the measured algal angular scattering properties. Therefore, Subsection 4.2 simulates the particle with a coated sphere and presents the estimated refractive indices provided by the genetic algorithm and Bernard models, and a combination of both. Finally, Subsection 4.3 uses a cylindrical-shaped particle with a homogeneous refractive index. This shape was selected because it is substantially different from a sphere and similar to that of some species of phytoplankton, e.g., the diatom *Thalassiosira pseudonana*. Although this particle design does not exactly lead to the same optical behavior as the actual phytoplankton particle -the micro-details of the cells are neglected- it can serve as a first approximation. The refractive indices are estimated from the combination of the genetic algorithm with the Bernard model for coated spheres and from the genetic algorithm alone.

### 4.1 Spherical-shaped homogeneous particles

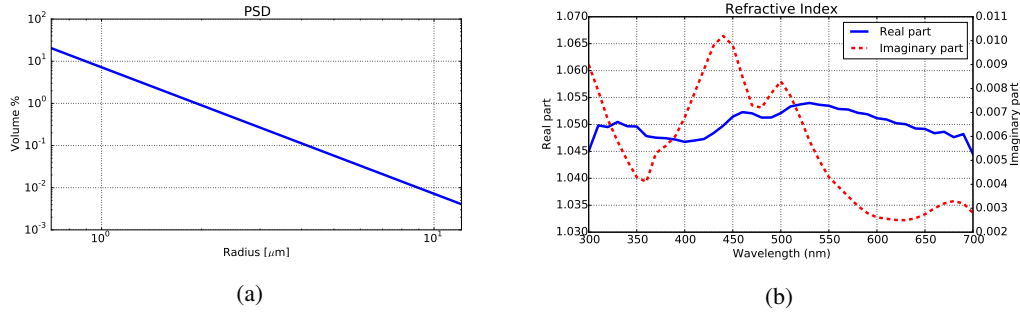
A concentration of 100 spherical particles per  $mm^3$ , with the PSD shown in Fig. 4a and the complex refractive index of Fig. 4b), was simulated using the Lorenz-Mie scattering theory (Bohren and Huffman, 1998); this PSD is based on a power-law distribution -or Junge-type- with 51 points,

$R_{min} = 0.7 \text{ } \mu m$ ,  $R_{max} = 12.1 \text{ } \mu m$ , a slope parameter  $\xi = 3$ , effective radius  $r_{eff} = 4 \text{ } \mu m$  and effective variance  $v_{eff} = 0.6$ . In particular, the BHMIE code, originally from Bohren and Huffman (Bohren and Huffman, 1998) and modified by B.T. Draine, was used as a forward model (additional features were added, such as poly-dispersion and the computation of the Stokes scattering matrix). The computed IOPs from this forward model, i.e. the  $a(\lambda)$ ,  $b(\lambda)$  and  $c(\lambda)$  coefficients, are shown in Fig. 5a. As may be observed, the concentration was selected in order to obtain IOP coefficients similar to those measured by Twardowski et al. (2001) and Stramski et al. (2001). The power-law distribution is used in our calculations for two main reasons. First, despite it is not a realistic one for single phytoplankton species, it is a fairly good approximation to natural-water composition even for anomalous conditions such as phytoplankton blooms, as there is always a strong background contribution to the PSD (Twardowski et al., 2001). And, second, because it is the only distribution that can be used in the Twardowski model.

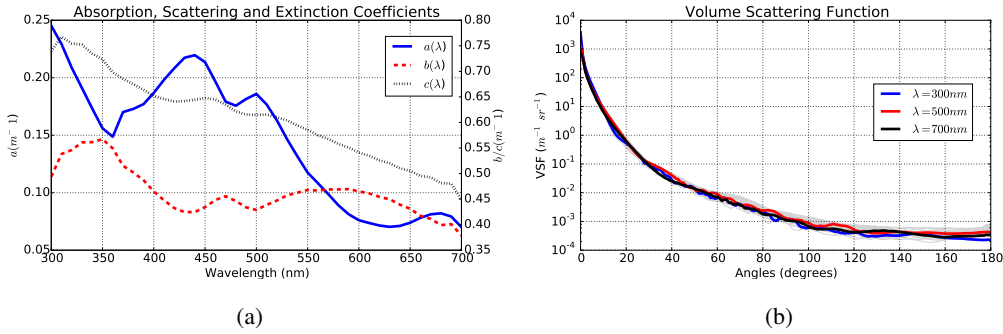
The complex refractive index in Fig. 4b was synthetically generated using imaginary values similar to those presented in Bernard et al. (2009) for phytoplankton species, derived from sample algal assemblages and considering homogeneous spheres. The dependence of the real part of the refractive index on the imaginary part can be found in the Kramers-Kronig relations (Bernard et al., 2001, 2009; Bricaud and Morel, 1986; van de Hulst, 1957), which allow the spectral variations in the real part of the refractive index to be calculated as the Hilbert transform of its imaginary part. The central part of the real refractive index was set to  $1 + \epsilon = 1.05$ ; for phytoplankton, this index typically ranges from 1.02 to 1.15, relative to water, as stated in Morel, 1973, Carder et al., 1974, Aas, 1996, Bernard et al., 2001. The effects due to normal dispersion, as described in Aas (1996), were not considered. The imaginary part presents peak values at 440, 500 and 680 nm, corresponding to the absorption wavelengths of the three chlorophyll (a,b and c) and accessory pigments; as expected, a similar shape is propagated to the absorption coefficient spectra,  $a(\lambda)$  (Fig. 5a). The volume scattering function is shown in Fig. 5b. As particles are relatively large relative to the wavelength, the scattering is mainly focused in the forward direction (between 0 and 10 degrees) and smoothly decreases in the backward direction.

#### 4.1.1 The Twardowski model

The spherical-shaped particle idealization was first examined with the Twardowski model. The use of Eq. (15) led to the results shown in Fig. 6a;  $\gamma$  is set to zero, as the slope parameter of the PSD,  $\xi = 3$ , and the backscatter fraction were computed with Eq. (13) using the volume scattering phase function values given by the modified BHMIE code. For the real part, the model leads to a curve similar to the postulated complex refractive index, but with a slight negative offset, presenting an averaged relative error of 42%. Since this model assumes a constant imaginary refractive index value of 0.005 (used for the development of the model), it has been considered as an output and compared with the postulated imaginary part of the refractive index, obtaining that the averaged relative error is



**Figure 4.** (a) PSD for the test run with spherical-shaped homogeneous particles. (b) Complex refractive-index signature postulated for spherical-shaped homogeneous particles.



**Figure 5.** (a) Absorption ( $a(\lambda)$ ), scattering ( $b(\lambda)$ ) and extinction ( $c(\lambda)$ ) coefficients and (b) the volume scattering function for spherical-shaped homogeneous particles (three wavelengths, 300, 500 and 700 nm, are plotted using intense colors, whereas the other wavelengths between 300 and 700 nm are plotted in light gray).

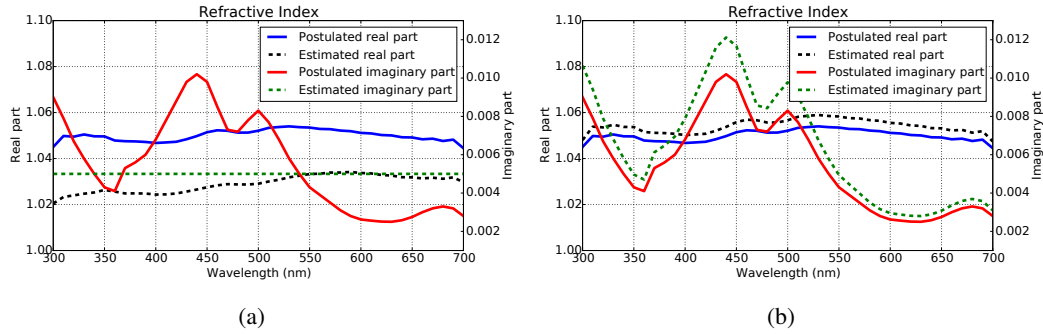
44%. It should be noted that the Twardowski model was designed for a rather different scenario: a bulk oceanic distribution presenting different physical properties than those of isolated species of phytoplankton, e.g., index of refraction and shape.

#### 4.1.2 The Stramski model

This model overestimates both the real and imaginary parts for all analysed spectra (Fig. 6b), showing an averaged relative error of 0.4% for the real part and a 15% for the imaginary part. It should be remembered that the imaginary part of the refractive index,  $k_h$ , is calculated with the ADA, known to give errors of about 10% when compared with results from the Lorenz-Mie theory (Bernard et al., 2009). Some discrepancies can therefore be expected between the ADA and Aden-Kerker derived values (Aden and Kerker, 1951).

#### 4.1.3 The genetic algorithm model

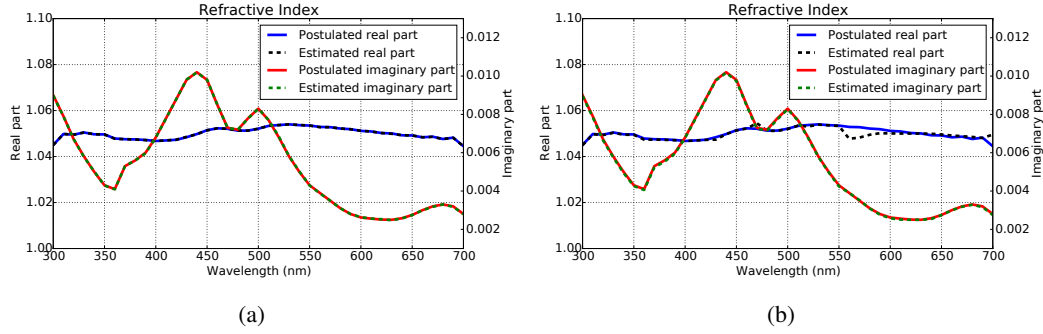
In order to implement the genetic algorithm described in Section 3.4, the tools provided by the DEAP (Distributed Evolutionary Algorithms in Python) and SCOOP (Scalable CONcurrent Opera-



**Figure 6.** Postulated and estimated refractive indices using (a) the Twardowski and (b) the Stramski models.

tions in Python) frameworks were used, respectively aimed at developing evolutionary algorithms and parallel task distribution (Fortin et al., 2012; Hold-Geoffroy et al., 2014). The fitness function was implemented using the fast subroutines of BHMIE to compute the absorption and scattering  
 365 properties of homogeneous spheres. The coefficients  $a(\lambda)$  and  $b(\lambda)$  of Fig. 5a were used as inputs to the genetic algorithm model in order to obtain the postulated complex refractive index, and limiting conditions were applied to facilitate the convergence (typical values for the real part of the phytoplankton refractive indices fall within 1.02 and 1.15 relative to water, and the bulk value of the imaginary part is always below 0.02). The genetic algorithm was configured with a vector of 2000  
 370 solutions over 10 generations and 50% and 20% of probability of crossover and mutation, respectively, leading to the values shown in Fig. 7a. The good agreement between the postulated complex refractive-index values and the estimated ones shows that it is possible to perform accurate estimations with a genetic algorithm (averaged relative errors of 0.0% for the real part and 0.2% for the imaginary part represent the best results for spherical homogeneous particles). It should be noted  
 375 that the number of generations necessary for a suitable convergence strongly depends on the length of the initial-solution vector and the cross-over and mutation percentages, among other parameters of the genetic algorithm. Adopting the previously described parameters, no significant improvement is generally found beyond the tenth generation.

One disadvantage of the genetic algorithms is that they are relatively slow and require more com-  
 380 putation time than other optimization algorithms, as they compute the fitness function many more times. Other optimization algorithms were also applied to determine whether similar results can be obtained with a significant reduction of the computation time. However, since none of them led to any meaningful improvement, no further description is provided. As a single example, Fig. 7b shows the results obtained with the much faster Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm -an  
 385 iterative method for solving unconstrained nonlinear optimization problems (Zhu et al., 1997)- executed using the same bounding conditions as in the genetic algorithm case. In this case, only 4 minutes were needed instead of 97 minutes when using the genetic algorithm. Both computations were performed using a PC with an Intel Core i7 processor at 3.2 GHz, with 16 GB of RAM. Al-



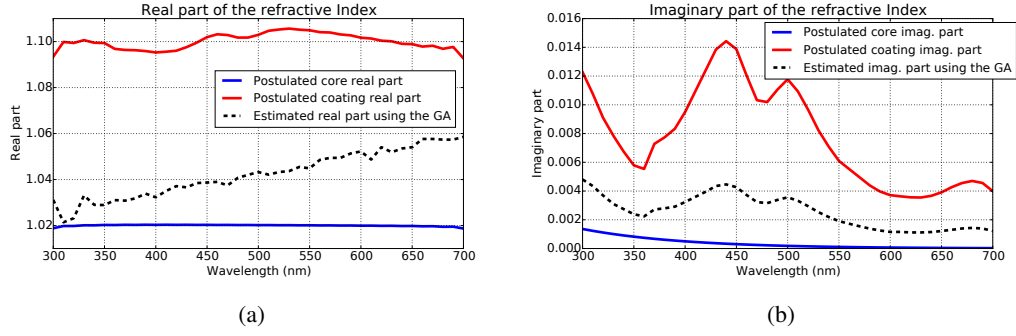
**Figure 7.** (a) Postulated and estimated refractive indices using the genetic algorithm; notice that the postulated and estimated values lie on top of each other. (b) Postulated and estimated refractive indices using the BFGS algorithm.

though the BFGS results are generally satisfactory, some of the wavelengths present a significant error in the real part (mainly between 550 and 600 nm and above 680 nm). The averaged relative error is 0.0% for the real part and 0.7% for the imaginary part. Other optimization algorithms, such as the conjugated gradient algorithm (Nocedal and Wright, 1999), were also tested. The results (not shown) exhibited worse accuracy than the BFGS, indicating that the genetic algorithm is probably the best method to solve this problem (in terms of accuracy but not in terms of executing time).

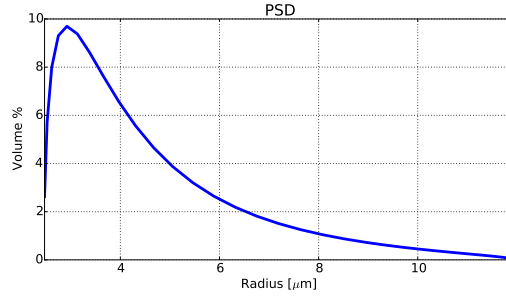
## 4.2 Spherical-shaped coated particles

In order to use the IOPs of a two-layered spherical particle that emulates actual phytoplankton organisms, its complex refractive index was generated using the description presented in Bernard et al. (2009). The imaginary refractive index of the inner cytoplasm was obtained using Eq. (20) and its real one using the Hilbert transform (Hahn, 1996) and Eq. (17) with  $1 + \epsilon = 1.02$ . The imaginary refractive index of the outer chloroplast was obtained using Eq. (21), with  $V_V = 30\%$ , a value that lies between those previously used by Bernard et al., 2009 and other authors, and the real refractive index was calculated using the Hilbert transform and Eq. (17) with  $1 + \epsilon = 1.1$ . Fig. 8a and Fig. 8b show the results for the real and imaginary parts, respectively (postulated values). In this example, instead of using a PSD describing a power-law function as in Fig. 4a, the PSD of an isolated culture was simulated with a concentration of 40 particles per  $mm^3$  ( $R_{min} = 0.7 \mu m$ ,  $R_{max} = 12.1 \mu m$  and using 31 points), as seen in Fig. 9. Notice that the PSD denotes the external radius while the inner one can be calculated using the  $V_V$  value. The absorption, scattering and extinction coefficients (Fig. 10a) and the volume scattering function (Fig. 10b) were obtained introducing these PSD and refractive indices in the BART code of A. Quirantes (a forward model based on the Aden-Kerker theory to calculate light-scattering properties for coated spherical particles, Quirantes (2005)).

The above set of IOPs can now be used to estimate the corresponding complex refractive indices. First, the genetic algorithm is used in order to see whether a basic shape such as a homogeneous

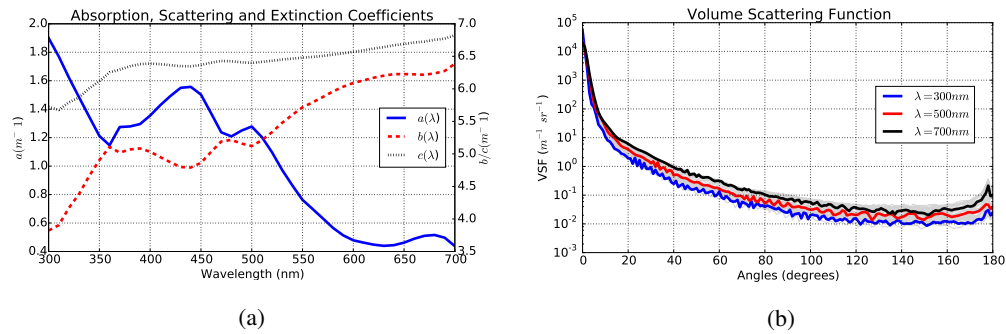


**Figure 8.** Postulated (a) real and (b) imaginary refractive-index signatures for the inner and outer layers of spherical-shaped particles, together with the estimated (a) real and (b) imaginary refractive-index signature calculated using the genetic algorithm model.

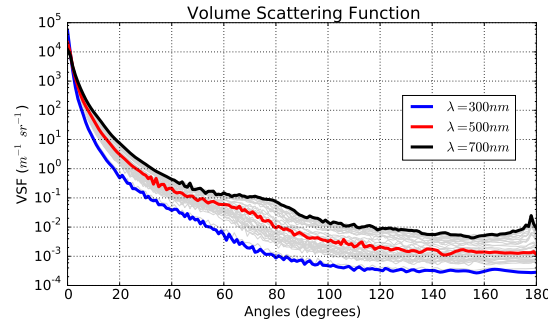


**Figure 9.** PSD simulating an isolated culture.

sphere is useful when modelling more complex particles. If coated particle models do better at characterizing the optical properties of general phytoplankton species, as stated in Bernard et al. (2009), this can be used to estimate the error associated with using homogeneous spheres. Next, the inner and outer complex refractive indices of the original particle are retrieved using the Bernard model for coated particles. Finally, a combination of the genetic algorithm and the Bernard model



**Figure 10.** (a) Absorption ( $a(\lambda)$ ), scattering ( $b(\lambda)$ ) and extinction ( $c(\lambda)$ ) coefficients and (b) volume scattering function for spherical-shaped coated particles.



**Figure 11.** Volume scattering function using the estimated refractive index for spherical-shaped coated particles.

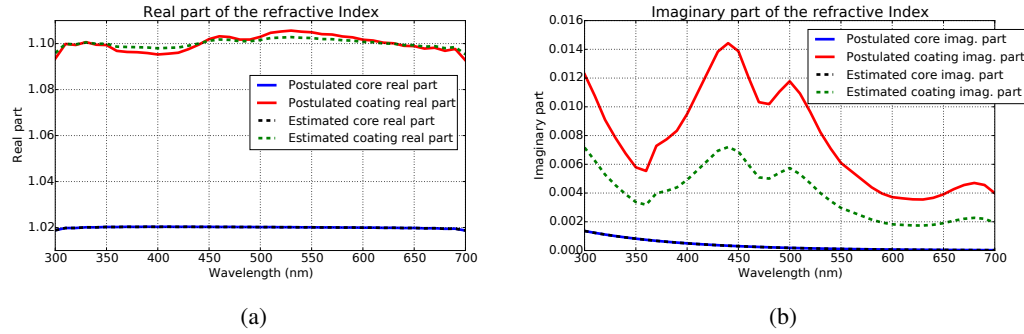
is applied to improve the previous results. Notice that the Twardowski model is not applied in order to avoid an inconsistent comparison with the other methods, as it was originally designed to be used with entire particle populations that are assumed to follow a power-law size distribution.

#### 4.2.1 The genetic algorithm model

The genetic algorithm model, previously employed to retrieve the refractive index of spherical-shaped homogeneous particles, was used for the spherical-shaped coated particles. The same configuration was used, i.e., an initial vector of 2000 solutions over 10 generations and 50% and 20% of probabilities for crossovers and mutations, respectively. The real part of the homogeneous case lies between the real values for the inner and outer regions in the coated case (Fig. 8a), and similarly for the imaginary values (Fig. 8b). The volume scattering function for the homogeneous particles (Fig. 11) is obtained by means of the Lorenz-Mie (forward model), using as inputs the estimated complex refractive index and the PSD (Fig. 9). This model presents similar values in the forward scattering but completely underestimates the backscattering, with values far below those in Fig. 10b. This example demonstrates that the common characterization using homogeneous spheres is not a suitable methodology when dealing with complex particles. This is not a surprising result, as it has been discussed by Bohren and Huffman, 1998, in the atmospheric context, and by Stramski et al., 2004, Clavano et al., 2007, Dall’Olmo et al., 2009, and Bernard et al., 2009, in the oceanic one, but the comparison between the two volume scattering functions highlights that the backscattering can exhibit errors of up to one order of magnitude.

#### 4.2.2 The Bernard model

The Bernard model, described in Section 3.3, was used to estimate the complex refractive index of the two-layered particles. Figure 12a shows the postulated and estimated real part of the inner and outer layers and Figure 12b shows the respective postulated and estimated imaginary parts. The inner refractive index is well estimated, an expected result as the same equation is used for both generation and retrieval, but the outer refractive index is not in accurate agreement. In particular, the imaginary



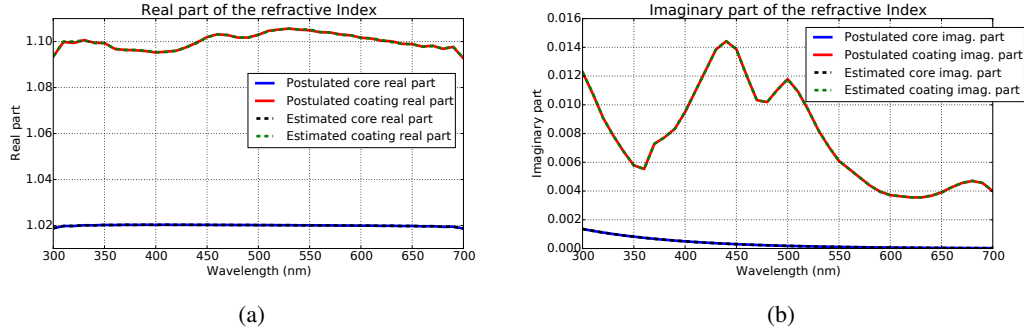
**Figure 12.** Postulated and estimated (a) real and (b) imaginary parts of the refractive indices for the inner and outer layers of spherical-shaped coated particles using the Bernard model.

part is significantly underestimated, with an averaged relative error of 51%. On the other hand, the simulation for the estimated refractive indices in coated spheres gives a volume scattering function that fits better the postulated values (Fig. 10b) than the one produced by the homogeneous spherical particle (the volume scattering function figure is not presented as the errors do not show up in the graph; a more detailed analysis is performed in Section 5 for this case).

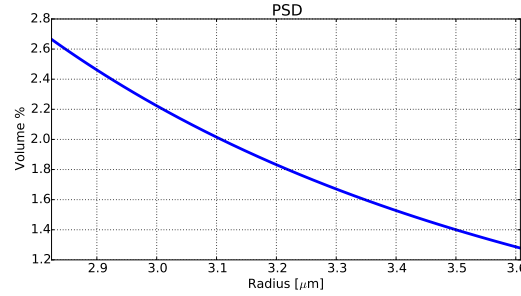
#### 4.2.3 The Bernard model combined with the genetic algorithm model

One possible modelling improvement would be to couple the genetic algorithm, which showed a reasonable performance when applied to homogeneous spherical particles, with the BART code (instead of the BHMIE code) in order to estimate the two complex refractive indices. However, the results would hardly be constrained as the solution has more degrees of freedom (the two refractive indices with their real and imaginary parts give four degrees of freedom) than the available data (two degrees of freedom, corresponding to the attenuation and scattering coefficients), i.e., this is an unconstrained (ill-posed) problem. Alternatively, the Bernard model could be combined with the genetic algorithm to increase the probability of convergence. In this case, the inner refractive index is again estimated using the Bernard model and the outer refractive index is then obtained with the genetic algorithm (coupled to the BART code). In this case, the genetic algorithm only has to find a solution with two degrees of freedom (the real and imaginary parts of the outer refractive index).

This method was applied to the coated particle example using the coefficients of Fig. 10a as input data and configured using an initial vector of 2000 solutions, 10 generations, 50% of probability for crossovers and 20% for mutations. The postulated and estimated real parts of the inner and outer layers are shown in Figure 13a and the corresponding postulated and estimated imaginary parts are presented in Figure 13b. Accurate results were obtained, certainly improving the refractive index estimate for the outer sphere. In this particular case, average relative errors of 0.0% and 0.1% were respectively obtained for the real and imaginary parts.



**Figure 13.** Postulated and estimated (a) real and (b) imaginary parts of the refractive indices for the inner and outer layers of spherical-shaped coated particles using the Bernard model combined with the genetic algorithm. Notice that in both cases the postulated and estimated values lie on top of each other.



**Figure 14.** PSD for cylindrical-shaped particles.

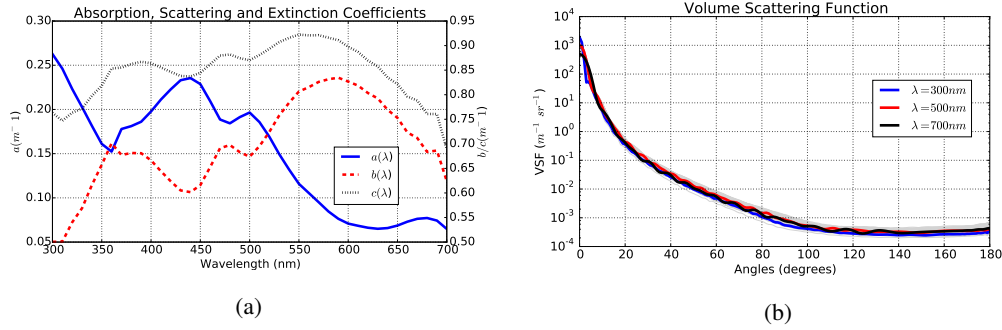
### 4.3 Cylindrical-shaped particles

As a final example, a cylindrical shape particle has been chosen. As commented above, phytoplankton species usually present complex shapes, far from perfect homogeneous or coated spheres.

470 This is the case, for example, of the centric diatom with cylindrical shapes (to name a few genera: *Thalassiosira*, *Aulacoseira*, *Skeletonema*, *Melosira*, ... ). In order to find out the most accurate model for the characterization of such complex shapes, we considered an example consisting of 100 prolate cylinders per  $\text{mm}^3$  with a diameter-to-length ratio equivalent to 0.8 and the PSD of Fig. 14 (showing the radius of an equivalent-volume sphere with slope parameter  $\xi = 3$ , effective radius

475  $r_{eff} = 3.2 \mu\text{m}$  and effective variance  $v_{eff} = 0.005$ , resulting in  $R_{min} = 0.8 \mu\text{m}$  to  $R_{max} = 3.6 \mu\text{m}$ ). The postulated refractive index of Fig. 4b was simulated using the  $T$ -Matrix algorithm from M. Mischenko (Mischenko and Travis, 1998) for randomly oriented, rotationally symmetric scatterers (cylinders, spheroids and Chebyshev particles). The PSD presents a small effective variance for convergence limitations of the code. The postulated  $a(\lambda)$ ,  $b(\lambda)$  and  $c(\lambda)$  coefficients are shown in

480 Fig. 15a, and the volume scattering function at each wavelength is shown in Fig. 15b.



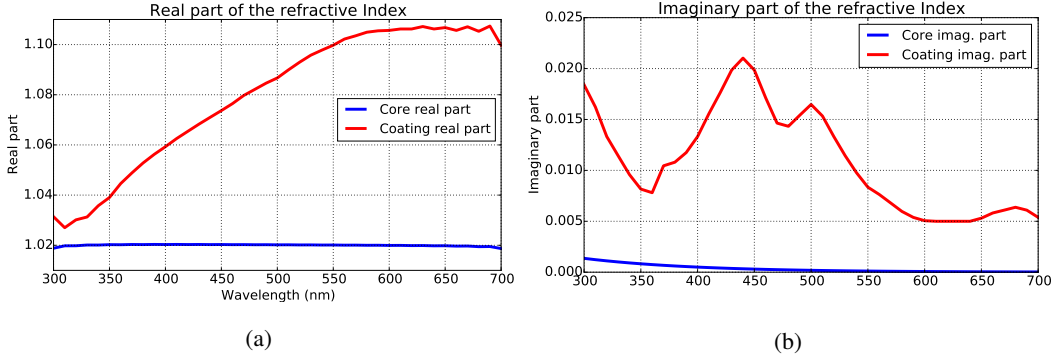
**Figure 15.** (a) Absorption ( $a(\lambda)$ ), scattering ( $b$ ) and extinction ( $c$ ) coefficients for cylindrical-shaped particles. (b) Volume scattering function for cylindrical-shaped particles.

#### 4.3.1 The Bernard model combined with the genetic algorithm model

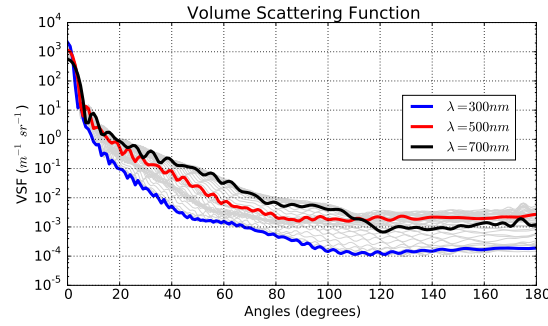
As previously discussed, the simulated cylindrical particles are not exact duplicates of the actual phytoplankton organisms, so it is useful to compare the results using this and the coated sphere design (usually used on all kind of phytoplankton shapes). As in previous examples, the postulated value of  $V_V$  for the coated sphere was 30%, which is an average value between that assumed by Bernard et al., 2009, and previous works. Figure 16a shows the estimated real part of the refractive index of the inner and outer layers and Figure 16b shows their corresponding estimated imaginary parts. However, these results cannot be directly compared with the postulated individual refractive index for the cylindrical particle. Instead, the IOPs obtained from the estimated refractive indices is computed using the forward model in order to analyze if the two designs, coated spheres and homogeneous cylinders, are comparable. The volume scattering function is obtained (Fig. 17) by means of the estimated complex refractive indices and the PSD of Fig. 14 in a  $T$ -Matrix forward model. The error in this last figure is large, especially at long wavelengths, reaching an averaged relative error of 77%. It should be noted that these differences may decrease when using real phytoplankton, since backscattering of heterogeneous particles is different from that of homogeneous particles.

#### 4.3.2 The genetic algorithm model

The genetic algorithm can be combined with the  $T$ -matrix code in order to consider cylindrical-shaped particles when estimating the inner complex refractive index. However, when using the Mishchenko code, one simulation of cylindrical-shaped particles needs about 67 minutes in a computer with an i7 processor at 3.20 GHz. This prevents using the genetic algorithm, as an accurate estimate the complex refractive index would require executing this simulation several hundreds of times at each wavelength, or several months for the entire refractive index spectra. An alternative approach for obtaining fast estimates is to use spherical homogeneous particles with the same volume as the cylinders. This allows using the Lorenz-Mie theory rather than the  $T$ -matrix approach, dramatically



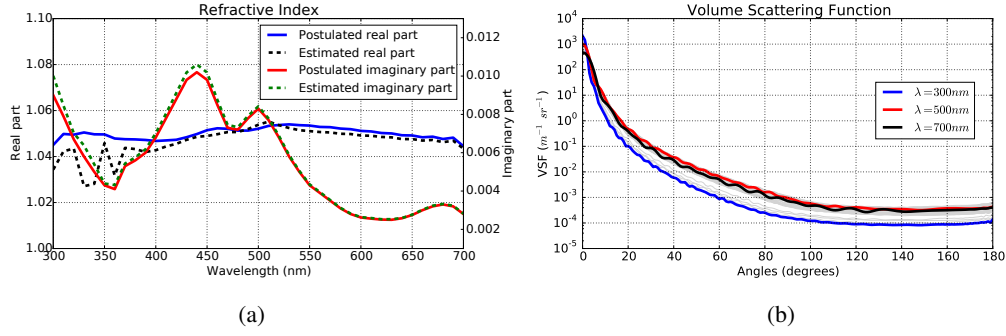
**Figure 16.** Inner and outer (a) real and (b) imaginary parts of the refractive indices using the Bernard model combined with the genetic algorithm for cylindrical-shaped particles.



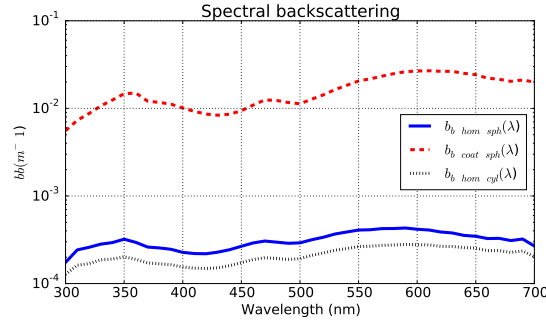
**Figure 17.** Volume scattering function obtained using the Bernard model combined with the genetic algorithm for cylindrical-shaped particles.

505 reducing the simulation time. The refractive index estimated using the equal-volume homogeneous spheres may then be applied for homogeneous cylinders in order to obtain their IOP, since the volume scattering function values are case sensitive to the particle shape. Although this calculation uses the slow  $T$ -matrix approach, it has to be executed only once. Certainly, much better computing resources (such as a computer cluster) would remove the above computing limitation and the genetic  
510 algorithm could be used with its complete potential.

The above methodology was applied using the same PSD as in Fig. 12. The estimated complex refractive index is shown in Fig. 18a. The averaged relative error is 8% for the real part and 3% for the imaginary part. Since the absorption is proportional to the volume, the inverted imaginary part of the refractive index agrees well with the postulated values (using equal-volume spheres).  
515 However, since scattering depends largely on the shape of the particles, the inverted real part of the refractive index deviates from the postulated values. The major differences are obtained at the lowest wavelengths, also noticeable in the volume scattering function (Fig. 18b) with some artifacts in those wavelengths where the real part of the refractive index changes abruptly (330 and 350 nm). However, the average relative error is 16%, much less than the 77% error for the Bernard method



**Figure 18.** (a) Postulated and estimated refractive indices using the genetic algorithm for the equal-volume spherical-shaped homogeneous particle and (b) the volume scattering function for the cylindrical-shaped particle.



**Figure 19.** Spectral backscattering for the three test cases: homogeneous sphere, coated sphere and homogeneous cylinder.

520 combined with the genetic algorithm. If the IOP is obtained using homogeneous spheres instead of  
cylinders, the averaged relative error increases to 22%, which demonstrates that choosing a suitable  
shape improves the results.

## 5 Discussion

The average relative errors of the real and imaginary parts of the estimated refractive indices, together  
525 with their respective volume scattering functions, are shown for each method applied to the three  
case examples considered in the previous section (Table 2); notice that the real part errors were  
computed for  $n - 1$  instead of  $n$ . Also notice that the inverse models do not compute the volume  
scattering function, rather it is obtained after introducing the estimated complex refractive indices  
on the suitable forward model, i.e., the Lorenz-Mie or  $T$ -Matrix theories.

530 In the homogeneous sphere example, the Twardowski model presents the highest errors, especially  
when comparing the volume scattering function. Although the Stramski model leads to complex re-  
fractive index errors considerably higher than the genetic algorithm (particularly for the imaginary

**Table 2.** Averaged relative errors for each method

Shapes	Model	$n$ relative error	$k$ relative error	VSF relative error
Homogeneous sphere	Twardowski model	42%	44%	68%
	Stramski model	8%	15%	0.2%
	Genetic Algorithm	0.0%	0.2%	0.2%
Coated sphere	Genetic Algorithm	-	-	78%
	Bernard model	2%	51%	52%
	Bernard model & GA	0.0%	0.1%	0.2%
Homogeneous cylinder	Bernard model & GA	-	-	77%
	Genetic Algorithm <sup>a</sup>	8%	3%	16% <sup>b</sup>

<sup>a</sup>The refractive index is estimated for equal-volume spheres but the IOP is obtained after using that refractive index on cylinders.

<sup>b</sup>If the cylindrical shape is not used, the error rises to 22%.

part), comparable estimates of the volume scattering function are recovered in both cases. This implies that, for this particular example, there is no need of accurate refractive index estimates in order to obtain a suitable characterization of the scattering properties. However, the genetic algorithm performs with excellent accuracy for the refractive-index retrieval.

In the coated sphere example, the genetic algorithm approximates the coated particle to a homogeneous one with a single complex refractive index. Therefore, errors for the inner and outer refractive indices cannot be obtained; additionally, this method differs substantially when computing the volume scattering function. Hence, if the optical response of coated spheres was similar to the response of actual phytoplankton particles (Bernard et al., 2009) then homogeneous spheres would not be a suitable choice for calculating optical properties. The Bernard model is a fast technique to estimate the inner and outer refractive indices, but fails mainly at estimating the imaginary part of the refractive index (error up to 51%). This leads to a significant error associated with the forward model when computing the volume scattering function. However, if the Bernard model is combined with the genetic algorithm model, i.e., the Bernard model is used to estimate the inner refractive index and the Genetic Algorithm to retrieve the external one, accurate values are obtained for the complex refractive indices and, after using the forward model, for the volume scattering function.

Finally, the optical properties of homogeneous cylinders are not accurately reproduced using coated spheres when their refractive indices are obtained combining the Bernard model and the genetic algorithm. It is likely that the optimal retrieval method would be the genetic algorithm using cylindrical shapes to obtain accurate estimates of the refractive indices. However, this involves using the slow  $T$ -Matrix code of Mischenko iteratively, which would require several months of computer time to converge (as the particles become more aspherical, the convergence time increases considerably). In order to make the retrieval faster, homogeneous spheres with equal volume are used instead

of cylinders, and the retrieved refractive index is then used to obtain the IOPs of cylinders. Using this method, the volume scattering function shows an average relative error of 16%, improving the result obtained when using spheres (22%). This result confirms that the proper selection of particle shapes is an important requirement in the modelling of optical properties.

This study has pointed at three important lines for future research:

1. All the test cases are synthetic examples that, presumably, are simpler than real life. Further work is necessary in order to study the performance of algorithms when using the optical properties of actual phytoplankton species and bulk oceanic measurements. A more complex inversion method remains to be developed in order to deal with a mixture of shapes and refractive indices. This scenario is currently dealt using only spherical particles of different sizes (Boss et al., 2001b, and references therein); however, besides the particle diameter, the *T*-Matrix method opens the possibility to consider other parameters associated with the particle shape. Thus, a new inversion algorithm based on the *T*-Matrix and an optimization technique, such as a genetic algorithm, could be developed in order to estimate the proportion of each phytoplankton morphological type and their own refractive index. Certainly, such algorithm would require an enormous computational effort, although this is a problem that becomes less critical as computer technology evolves. Additionally, other methods could be combined with the genetic algorithm to make the estimation easier, as the one presented by Ciotti et al. (2002), that found a relationship between the dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient.

When dealing with actual phytoplankton, a critical issue is the instrumental accuracy. The attenuation and scattering coefficients are key inputs for all retrieval methods, in order to retrieve valid refractive indices. However, as stated by Ramírez-Pérez et al. (2015), the acceptance angle of optical instruments severely affects the amplitude of the measurements. By comparing the extinction coefficient of two instruments with different acceptance angles, disparate magnitude values were obtained, with an average ratio of 0.67. Accurate measurements are a requirement for obtaining reliable results with the presented methodology.

2. The accuracy of the inversion methods could be improved by applying the *T*-Matrix method to new particle shapes. For instance, coated cylinders could represent algae with cylindrical shape (coated-spherical particles generate backscattering functions closer to those produced by actual phytoplankton particles, Bernard et al., 2009) or other axially-symmetric designs could replicate the actual shape of some phytoplankton particles (as commented before, the *T*-Matrix approach allows using particle shapes with axial symmetry, Sun et al., 2016).
3. Ocean optics goes from research at a microscopic scale (as shown in this paper) to remote sensing, measuring the reflected or backscattered radiation in large areas. The inversion methods based on Lorenz-Mie and *T*-Matrix could be extended to consider other type of optical

measurements, aside the IOPs, such as the remote-sensing reflectance. As an example, Fig. 19 shows the spectral backscattering for the three test cases: the homogeneous sphere, the coated sphere and the homogeneous cylinder. Many operational remote-sensing inversion models for IOPs use implicit or explicit assumptions on the refractive index. Hence, their output would likely improve when combined with the inversion methods here presented. Retrieving the index of refraction from space would improve the ability to distinguish different oceanic sources of backscattering, but certainly requiring a much more complex inversion scheme.

To conclude, the results presented in this paper, and summarized in Table 2, do not determine which is the best method to estimate the phytoplankton optical properties, since none of them is a realistic representation of real algae with cell walls, chloroplasts, vacuoles, nuclei and other internal organelles, each with its own optical properties. However, the assumed particles serve as a first approximation to actual phytoplankton and are useful to extract some preliminary conclusions and to introduce improvements in order to obtain approximations closer to reality. Most of the methods shown in this paper are already being used for the retrieval of the refractive indices of isolated particles or bulk oceanic distributions, and their performance can be compared using well-known models. It has been shown that the genetic algorithm model is not a fast technique, since it requires several minutes for each estimation (when using spherical shapes, and longer for aspherical particles) as compared to the few seconds generally required by other methods, with the Twardowski model being the faster one. However, the genetic algorithm is a versatile technique that alone, or combined with other methods, improves the accuracy of the results to a level not achieved by any other method.

## 6 Conclusions

The accuracy of different inverse methods retrieving refractive indices from the optical properties of small scatterers, and their particle size distribution, has been analyzed. To this end, three different synthetic examples were constructed, each one with a different shape and distribution. The selected shapes were homogeneous spheres, coated spheres and homogeneous cylinders. The results indicate that the most accurate methods are those using a genetic algorithm to optimize the inversion, although they were also the slowest ones. In particular, an excellent agreement was obtained between the estimated and actual refractive indices and volume scattering functions for the homogeneous and coated sphere cases, and a fair agreement for the homogeneous cylinders. The results suggest that even better characterizations could be obtained for the actual phytoplankton optical properties. A next step should be the analysis of the performance of these methods when applied to measurements of isolated cultures of phytoplankton.

*Acknowledgements.* This work was supported by the Spanish National Research Council (CSIC) under the EU Citclops Project (FP7-ENV-308469), the MESTRAL project (CTM2011-30489-C02-01), and the CSIC

ADOICCO project (Ref 201530E063). The authors would also like to show their gratitude to Laura Pelegrí, Jimena Uribe, Josep Lluís Pelegrí and Miquel Ribó for the English revision, and to Emmanuel Boss and two anonymous reviewers for their comments and suggestions, which helped enhancing the quality of the manuscript.

## 630 **References**

- Aas, E.: Refractive index of phytoplankton derived from its metabolite composition, *Journal of Plankton Research*, 18, 2223–2249, 1996.
- Aden, A. and Kerker, M.: Scattering of electromagnetic waves from two concentric spheres, *J. Appl. Phys.*, 22, 1242–1246, 1951.
- 635 Bernard, S., Probyn, T., and Barlow, R.: Measured and modelled optical properties of particulate matter in the southern Benguela, *South African Journal of Science*, 97, 410–420, 2001.
- Bernard, S., Probyn, T., and Quirantes, A.: Simulating the optical properties of phytoplankton cells using a two-layered spherical geometry, *Biogeosciences Discuss.*, 6, 1–67, 2009.
- Bhandarkar, S., Zhang, Y., and Potter, W.: An edge detection technique using genetic algorithm-based optimization, *Pattern Recognition*, 27, 1159 – 1180, 1994.
- 640 Bohren, C. and Huffman, D., eds.: *Absorption and scattering of light by small particles*, New York : Wiley, Oxford, 1998.
- Boss, E., Pegau, W., Gardner, W., Zaneveld, J., Barnard, A., Twardowski, M., Chang, G., and Dickey, T.: Spectral particulate attenuation and particle size distribution in the bottom boundary layer of a continental shelf, *J. Geophys. Res.*, 106, 9509–9516, 2001a.
- 645 Boss, E., Twardowski, M., and Herring, S.: Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution, *Appl. Opt.*, 40, 4885–4893, 2001b.
- Boss, E., Pegau, W., Lee, M., Twardowski, M., Shybanov, E., Korotaev, G., and Baratange, F.: Particulate backscattering ratio at LEO 15 and its use to study particles composition and distribution, *J. Geophys. Res.*, 109, 2004.
- 650 Bricaud, A. and Morel, A.: Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling, *Appl. Opt.*, 25, 571–580, 1986.
- Bricaud, A., Zaneveld, J., and Kitchen, J.: Backscattering efficiency of coccol-25 ithophorids: use of a three-layered sphere model, In: *Ocean Optics XI Proc SPIE*, p. 27–33, 1992.
- 655 Carder, K., Betzer, P., and Eggimann, D. W., eds.: *Physical, chemical, and optical measures of suspended particle concentrations: Their intercomparison and application to the west African shelf*, in *Suspended Solids in Water*, Springer US, Oxford, 1974.
- Choi, W., Fang-Yen, C., Badizadegan, K., Oh, S., Lue, N., Dasari, R., and Feld, M.: Tomographic phase microscopy, *Nature Methods*, 4, 717–719, 2007.
- 660 Ciotti, A., Lewis, M., and Cullen, J.: Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient, *Limnol. Oceanogr.*, 47(2), 404–417, 2002.
- Clavano, W., Boss, E., and Karp-Boss, L.: Inherent optical properties of non-spherical marine-like particles - From theory to observation, *Oceanography and Marine Biology: An Annual Review*, 45, 1–38, 2007.
- 665 Dall’Olmo, G., Westberry, T., Behrenfeld, M., Boss, E., and Slade, W.: Significant contribution of large particles to optical backscattering in the open ocean, *Biogeosciences*, 6, 947–967, 2009.
- Fortin, F., Rainville, F. D., Gardner, M., Parizeau, M., and Gagné, C.: DEAP: Evolutionary algorithms made easy, *Machine Learning Research*, 13, 2171–2175, 2012.

- Gordon, H. and Morel, A.: Remote assessment of ocean color for interpretation of satellite visible imagery: A review, Springer Science and Business Media, 2012.
- Greenhalgh, D. and Marshall, S.: Convergence Criteria for Genetic Algorithms, *J. on Computing*, 30, 269–282, 2000.
- Hahn, S. L.: Hilbert transforms in signal processing, Artech House on Demand, 1996.
- Hale, G. and Querry, M.: Optical constants of water in the 200-nm to 200- $\mu$ m wavelength region, *Appl. Opt.*, 12, 555–563, 1973.
- Hold-Geoffroy, Y., Gagnon, O., and Parizeau, M.: Once you SCOOP, no need to fork, in: Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment, p. 60, ACM, 2014.
- Kirk, J. T.: Light and photosynthesis in aquatic ecosystems, Cambridge University Press, 1994.
- Latimer, P.: Light scattering by a homogeneous sphere with radial projections, *Appl. Opt.*, 23, 442–447, 1984.
- Lorenz, L.: Sur la lumière réfléchie et réfractée par une sphère (surface) transparente, vol. I of *Oeuvres scientifiques de L. Lorenz. Revues et annotées par H. Valentiner*, Librairie Lehmann et stage, Copenhagen, 1898.
- Mera, N., Elliott, L., and Ingham, D. B.: A multi-population genetic algorithm approach for solving ill-posed problems, *Computational Mechanics*, 33, 254–262, 2004.
- Meyer, R.: Light-scattering from biological cells - Dependence of backscatter radiation on membrane thickness and refractive index, *Appl. Optics*, 18, 585–588, 1979.
- Mie, G.: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen, *Annalen der Physik*, 330, 377–445, 1908.
- Mischenko, M. and Travis, L.: Capabilities and limitations of a current Fortran implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers, *Quant. Spectrosc. Radiat. Transfer*, 60, 309–324, 1998.
- Mischenko, M., Travis, L., and Mackowski, D.: T-matrix computations of light scattering by nonspherical particles: a review, *Quant. Spectrosc. Radiat. Transfer*, 55, 535–575, 1996.
- Mobley, C., ed.: Light and Water: Radiative Transfer in Natural Waters, Academic Press, 1994.
- Morel, A.: Diffusion de la lumière par les eaux de mer; résultats expérimentaux et approche théorique, *Optics of the Sea, AGARD Lecture Ser.*, 61, 3.1.1.–3.1.76, 1973.
- Mugnai, A. and Wiscombe, W.: Scattering from nonspherical Chebyshev particles. I: cross sections, single-scattering albedo, asymmetry factor, and backscattered fraction, *Appl. Opt.*, 25, 1235–1244, 1986.
- Nocedal, J. and Wright, S., eds.: Numerical Optimization, New York : Springer, 1999.
- Quirantes, A.: A T-matrix method and computer code for randomly oriented, axially symmetric coated scatterers, *Quant. Spectrosc. Radiat. Transfer*, 92, 373–381, 2005.
- Quirantes, A. and Bernard, S.: Light scattering by marine algae: two-layer spherical and nonspherical models, *Quant. Spectrosc. Radiat. Transfer*, 89, 311–321, 2004.
- Quirantes, A. and Bernard, S.: Light-scattering methods for modelling algal particles as a collection of coated and/or nonspherical scatterers, *Quant. Spectrosc. Ra.*, 100, 315–324, 2006.
- Ramírez-Pérez, M., Röttgers, R., Torrecilla, E., and Piera, J.: Cost-effective hyperspectral transmissometers for oceanographic applications: performance analysis, *Sensors*, 15, 20 967–20 989, 2015.

Sánchez, A., Zafra, E., and Piera, J.: Hyperspectral characterization of marine particles based on Mie-Lorentz and T-matrix codes and a genetic algorithm, in: Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), IEEE, 2014.

710 Stramski, D., Morel, A., and Bricaud, A.: Modeling the light attenuation and scattering by spherical phytoplanktonic cells: a retrieval of the bulk refractive index, *Appl. Opt.*, 27, 3954–3956, 1988.

Stramski, D., Bricaud, A., and Morel, A.: Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community, *Appl. Opt.*, 40, 2929–2945, 2001.

715 Stramski, D., Boss, E., Bogucki, D., and Voss, K.: The role of seawater constituents in light backscattering in the ocean, *Progress in Oceanography*, 61, 27–56, 2004.

Sun, B., Kattawar, G., Yang, P., Twardowskic, M., and Sullivan, J.: Simulation of the scattering properties of a chain-forming triangular prism oceanic diatom, *Quant Spectrosc Radiat Transfer*, Available online, 2016.

Twardowski, M., Boss, E., Macdonald, J., Pegau, W., Barnard, A., and Zaneveld, J.: A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters, *Geophysical Research: Oceans*, 106, 14 129–14 142, 2001.

720 van de Hulst, H., ed.: *Ligh scattering by small particles*, New York : Wiley, Oxford, 1957.

Volz, F.: *Die optik und meteorologie der atmosphärischen trubung*, *Ber. Dtsch. Wetterdienstes*, 13, 1–47, 1954.

Waterman, P.: Matrix formulation of electromagnetic scattering, *Proceedings of the IEEE*, 53, 805–812, 1965.

725 Wiscombe, W. and Grams, G.: The backscattered fraction in two-stream approximations, *Journal of the Atmospheric Sciences*, 33, 2440–2451, 1976.

Zaneveld, J. and Kitchen, J.: The variation in the inherent optical properties of phytoplankton near an absorption peak as determined by various models of cell structure, *Geophys. Res.*, 100(C7), 13 309–13 320, 1995.

Zhu, C., Byrd, R., and Nocedal, J.: L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization, *ACM Transactions on Mathematical Software*, 23, 550–560, 1997.

730