Dear Reviewers, we are grateful to you for your comments and suggestions, which have helped to improve our contribution. Below you can find our answers to all your comments, addressing the modifications performed in the paper.

**Reviewer 1 (RC C8634):**

General comments

**Comment:** The method 4 – the genetic algorithm that this study is intended to evaluate – is simply a numeric method and it totally different from the "models 1-3", which in essence are bio-optical models. In a sense, models 1-3 are formulae whereas method 4 is an (advanced) technique seeking solution of a formula. Therefore, there is no comparison between them. Actually, as the authors have already mentioned, the genetic algorithm can be applied to either of those three models in seeking a better solution. In addition, the model-1, as mentioned above, is intended to be used for entire particle populations that are assumed/expected to follow a power-law size distribution, and is fundamentally different from models-2&3, which were developed to apply to a phytoplankton culture (or dominance of one particular phytoplankton species) and require the concurrent measurement of the size distribution. These apple vs. orange comparisons show a poor design of the study.

**Answer:** The differences between the presented models are known and already stated in the text (see for instance end of Subsection 4.1.1). However, regardless of the scenario, they all perform the same task, which is the estimation of the inner refractive index of the small scatterers. The aim of this study is precisely a fair comparison between the different methods in several theoretical situations which, in some cases, have been adapted to fit in a particular model (note, for instance, that the first example uses a power-law distribution in order to apply the Twardowski Model). With this comparison, it is possible to certify that for the test cases presented in this paper, the Twardowski method is not the best option among all the possibilities (mainly, as you state, because it was not designed for isolated cultures). Even though this was an expected result, only after the numerical experiment the inadequate use of this model was objectively evaluated, considering the relative error as a performance indicator.

From our point of view, the results shown in this paper can be useful for other scientists interested in the retrieval of the refractive indices in order to select the most suitable methodology without having to test all of them, at least, when having similar test conditions.

To further clarify it in the text, the following modifications (highlighted in black) have been done in the Introduction:

"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 is used by \citet{Twardowski2001}, to estimate the refractive index **of a bulk oceanic distribution**. It is indeed a fast method if optical backscattering measurements are feasible. \citet{Stramski1988} presented an extension of a model from \citet{Bricaud1986}, **designed for isolated phytoplankton cultures (or dominance of one particular phytoplankton species)**."

And, later, at the end of the introduction, the following paragraph has been added:

"It must also be noted that the models are fundamentally different. The model developed by \citet{Twardowski2001} is intended to be used for entire particle populations that are assumed to follow a power-law size distribution while the other models are developed for single phytoplankton cultures (or dominance of one particular phytoplankton species) and require the concurrent measurement of the size distribution. And, besides, such bio-optical models are compared with a numeric method (i.e., the genetic algorithm) in the same conditions. On the other hand, the methodology applied in this contribution allows to obtain an objective comparison of the results of the different methods in those occasions where it is not clear which methodology is most suitable, and therefore, interesting for the ocean optics community."

**Comments:** As far as optical modeling is concerned, I'm not sure if the super-accurate estimation of the refractive index offered by the genetic algorithm is meaningful. For one particular wavelength, the genetic algorithm was configured to partition complex refractive index into 2000 random values with real parts between 1.02 and 1.15 and imaginary parts between 0 and 0.02. Each of these complex values is tested to find the best refractive index that reproduce the observed absorption and scattering coefficients. Then this procedure is repeated by generated a new sets of random values following a certain rule (e.g., 50% crossover and 20% mutation). This entire process then moves to the next wavelength. The authored showed that the genetic algorithm can provide a solution with an accuracy of 0.08% for the real part of the index (the error was estimated against n-1) and 0.24% for the imaginary part of the index as in the test of spherical particles. Such moot precision can never be verified in an experiment nor can lead to meaningful improvement in optical modeling.

**Answer:** The research presented in this paper comes from the necessity to reproduce the spectra signature of particle absorption and scattering. Using Lorenz-Mie or *T*-Matrix methods to this end requires retrieving previously the refractive index. Our first attempt in this field was using the existing methods of refractive index estimations but, as seen in the paper, they have an associated error. When the results were introduced in the forward simulations (Lorenz-Mie or *T*-Matrix), the desired spectra was never recovered, which made us to develop some improvements by using, for instance, an accurate optimization method such as the genetic algorithm.

In any case, none of the models described in the paper are a realistic representation of real algae where there may be cell walls, chloroplasts, vacuole, nucleus and other internal organelles, each with its own optical properties. The best we can do is a gross approximation (usually using homogeneous spheres) only useful from an optical point of view. The aim of the genetic algorithm is only an attempt to make the approximation a bit closer to the reality. Maybe, using modern techniques (the development of new measurement techniques is obtaining more accurate mappings of refractive-index distributions in live cells and tissues, as seen for instance in *Tomographic Phase Microscopy*, published in Nature Methods By Choi et al. 2007), the verification could be developed.

This has been clarified in the paper by adding the following sentence in the Introduction:

"Although new promising techniques such as Tomographic Phase Microscopy (Choi et al., 2007) may provide in the future measurements of the refractive index in live cells, at present current ocean instrumentation do not directly provide it, so it must be estimated somehow (Aas, 1996)."

And, in the Discussion:

"To conclude, the results presented in Table 2 do not determine which is the best method to estimate the phytoplankton optical properties, since none of them are a realistic representation of real algae where there may be cell walls, chloroplasts, vacuole, nucleus and other internal organelles, each with its own optical properties. However, the assumed particles serve as a first approximation of actual phytoplankton and are useful to extract some preliminary conclusions and to introduce several improvements as an attempt to make the approximations a bit closer to the reality."

Presentation

**Comment:** You used relative error in evaluating the performance, but didn't provide a definition. Since all of the relative errors cited in the text are positive values, I'd assume you used absolute values. But, this should be defined.

**Answer:** This has been clarified adding the definition of the relative error at the beginning of Section 4:

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

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

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

**Comment:** In Page 18739 Lines 3-16 and Figure 5, you compared genetic algorithm with other optimization algorithms, which you did not introduce in the method section. You also mentioned that the BFGS method showed averaged relative error 0.073% for the real part and 0.72% for the imaginary part, which you think are worse than the genetic algorithm. However, this performance measure is actually better than the performance of the genetic algorithm you listed in Table 1 for the real parts of the index.

**Answer:** As you can find in the text (just above the text you pointed out), the error committed by the genetic algorithm is 0.004% for the real part and 0.24% for the imaginary part, which is better than that committed by the BFGS. In Table 1 there was an error in the real part that has also been corrected.

The reason for not introducing BFGS, which is another optimization method as it is the Genetic Algorithm, is because its performance is much worse than the Genetic Algorithm when applied on the examples described in the manuscript. The results were only added in order to state that other optimization algorithms were also tested (a part from BFGS we also tried with Nelder-Mead, Conjugated Gradient, etc.), but still the Genetic Algorithm provided the best solutions. We did not find necessary to make the paper longer with the introduction of a new method that does not present any meaningful improvement. As explained in the text, the Genetic Algorithm presents some disadvantages, being the convergence time the most important one, but its accuracy level is not achievable by any other algorithm.

In order to clarify the reason for not introducing in more detail the BFGS algorithm in the paper, the following text has been added in Section 4.1.3:

"Other optimization algorithms were also applied to determine if similar results can be obtained with a significant reduction of the computation time. **However, since none of them led to any meaningful improvement, they are not introduced here**. As an example, Fig. 7b shows the results…"

**Comment:** I think the test 3 (cylindrical particles) is confusing. First, you used coated spheres to emulate homogeneous cylindrical particles in inversion. Since the cylindrical particles are homogeneous and have only one refractive index, how do you evaluate the results (Fig. 14) of the coated spheres which would give two indices, one for the core and one for the coating. Second, due to the computation constraint, you used equivalent volume spheres to simulate the cylindrical particles in inversion. How can this help you evaluate the performance of the genetic algorithm? It cannot! And it is clearly shown in Fig. 16. Since absorption is proportional to the volume and you used volume equivalent spheres, the inverted imaginary part of the refractive index agree well with the assumed values. However, since scattering depends strongly on the shape of particles, the inverted real part of the refractive index deviate significantly from the assumed values.

**Answer:** As you state, the two complex refractive indices of the coated sphere cannot be compared with the individual one of the homogeneous cylinder. Instead, we need to use the IOPs that are recovered using the estimated refractive indices in the forward model to analyse if coated spheres are useful to emulate homogeneous cylinders. In this particular case, we used the volume scattering function to this end. To clarify this in the text, the following sentences have been added in Section 4.3.1:

"Figure 16a shows the assumed and estimated real part of the inner and outer layers and Figure 16b shows the assumed and estimated imaginary parts. **In this case, they cannot be compared with the assumed individual refractive index of the cylindrical particle. Instead, the IOPs obtained from the estimated refractive indices need to be computed using the forward model to analyze if this model is useful to emulate homogeneous cylinders.**"

Regarding the second comment, the real evaluation of the genetic algorithm is done in Section 4.1.3, since it is where the genetic algorithm uses the same shape as the assumed one to estimate the refractive index. In this example, the same could be done if we didn't have the computation constraints when using cylinders. Indeed, simpler examples using cylinders, not reported in the paper, have already been tested with the genetic algorithm and the original refractive index was accurately recovered. But, for the paper example, we needed to think in a faster estimation since it was not practical for us to leave a PC computing for several days. Thus, this result cannot be considered as a validation (or invalidation) of the genetic algorithm model, but an alternative technique to find the assumed refractive index. For sure, using better computing resources (as for instance, by means of a computer cluster), this problem disappears and the genetic algorithm can be used with its complete potential. To clarify this in the paper, the following sentence has been added in Section 4.3.2:

"For sure, using better computing resources (as for instance, by means of a computer cluster), this problem disappears and the genetic algorithm can be used with its complete potential".

And later, in the same Section:

"Since absorption is proportional to the volume, the inverted imaginary part of the refractive index agree well with the assumed values (volume equivalent spheres are being used). However, since scattering depends strongly on the shape of particles, the inverted real part of the refractive index deviate from the assumed values".

**Comment:** While I can understand the text, the writing needs improvement. Also, some figures are difficult to interpret. I will list some specific examples below regarding the writing, figures and others.

**Answer:** Paper style has been improved using the recommendations described below, the suggestions of Reviewer 2, and others of our own.

<u>Specific comments</u>

**Comment:** 2pi in Eq. (9) is not a normalization factor. It comes naturally from integration w.r.t. the azimuth angle. Sometimes (and often in atmospheric optics), the integration of phase function is normalized to 4pi (representing the total solid angle over entire sphere), in this case, the so-called factor is 1/2.

**Answer:** Last paragraph of Section 2.1 has been re-edited considering this suggestion.

**Comment:** The Bernard et al. 2009 reference, which you have cited multiple times and is the basis for your model-3, was not in the bibliography list.

**Answer:** It did not appear due to a problem of compilation in Latex. It has been solved.

**Comment:** Page 18734, Line 11, attenuation → absorption

**Answer:** It has been corrected.

**Comment:** Page 18740, Line 4, " . . . not agree with a perfect power-law distribution since there is minimum size beneath which there are no cells." Any power law function has to stop somewhere in the lower end!

**Answer:** This sentence tries to state the difference between the PSD of Fig. 4a (performing a classical power-law function), typical in oceanic bulk distributions, and that of Fig. 9, more similar to that of isolated cultures. In order to avoid any confusion, the following text:

"Instead of using the PSD of Fig. 4a for this example, the PSD of an isolated culture was simulated with a concentration of 40 particles $mm^3$ (Rmin = 0.7 µm, Rmax = 12.1 µm). It must be noted that the PSD denotes the external radius (the inner one can be extracted using the $V_V$ value). In this case, the function does not agree with a perfect power-law distribution since there is a minimum size beneath which there are no cells. Thus, the PSD of Fig. 7 (using 31 points) better represents the case of a monoculture PSD.",

has been simplified to:

"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 $mm^3$ (Rmin = 0.7 um, Rmax = 12.1 um and using 31 points), as seen in Fig. 9. It must be noted that the PSD denotes the external radius (the inner one can be calculated using the $V_V$ value)."

**Comment:** The way the volume scattering functions were shown in the figures does not help in evaluating the results. Why not draw VSFs at only a few wavelengths using lines instead of the color map.

**Answer:** VSF figures have been redrawn as suggested. Only three wavelengths (300, 500 and 700 nm) are plotted using intense colors, as seen in the legend, while the rest of wavelengths between 300 and 700 nm in steps of 10 nm are plotted in light grey. A clarification has also been added in the last paragraph of Section 4.1:

"The volume scattering function is shown in Fig. 5b. Only three wavelengths (300, 500 and 700 $nm$) are plotted using intense colors, while the rest of wavelengths between 300 and 700 $nm$ in steps of 10 $nm$ are plotted in light grey."

**Comment:** In section 4.3 Cylindrical-shape particles, you tested coated sphere, but did not mention the size of the core and how did you come up with that size.

**Answer:** The size of the core can be extracted from the relative chloroplast volume $V_V$. As in previous examples, the assumed value was $V_V$ = 30%, since it is a value between that assumed by Bernard et al., 2009, and previous works. This clarification has been added in Section 4.3.1 with:

"As in previous examples, the assumed value was $V_V$ = 30% (an averaged value between that assumed by Bernard et al, 2009, and previous works)."

**Comment:** Cylindrical-shape or spherical-shape should be cylindrical-shaped or spherical-shaped

**Answer:** It has been corrected.

**Comment:** In specifying wavelengths (e.g. Page 18743, line 13), longer or shorter are typically used (e.g., longer wavelength), whereas higher or lower are typically used for frequencies (e.g., lower frequency).

**Answer:** Thank you for this clarification. It has been corrected.

**Comment:** Both "initial" and "synthetic" refractive indices (as in figure captions) are used to represent the assumed values that have been used to simulate the optical properties. Initial values were also used in running the genetic algorithm. Recommend to use "assumed" to avoid confusion

**Answer:** The document has been reviewed and the "initial" and "synthetic" adjectives have been replaced by "assumed" when not referring to the initial values of the genetic algorithm.

**Reviewer 2 (RC C9713):**

<u>Overall comment</u>

**Comment:** The biggest weakness of this paper is that it presents relatively complex concepts and models in a manner that needs a better organization. In particular, a summarizing flowchart or table is necessary that shows the inputs used for each model/algorithm and in each test case, the PSD, the type of particle assumed, etc., as well as the assumptions of the model, the equations used (refer to equations in this paper or elsewhere). Such a flowchart/table will greatly help the reader be able to follow how each model is applied, in a forward or inverse manner. Also, the units are sometimes not given, please give units consistently everywhere, including captions/axes labels.

**Answer:** We have tried to clarify the descriptions provided along the paper by adding new Fig. 1, Fig. 2 (both in the Introduction) and Table 1 (in Section 3):

- Fig. 1 shows some phytoplankton particles and the axially symmetric figure that best characterizes their shape. It also shows the type of shape characterization that can be used with Lorenz-Mie (only spheres) and *T*-Matrix methods (more complex shapes).
- Fig. 2 shows the three steps followed to analyse each method: 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 wavelength-dependent refractive index, $m(\lambda)$, the PSD and the particle shape). Then, the inverse models (described above) are used to estimate the initial refractive index using the IOPs obtained in the first step. Finally, the estimated refractive index along with the assumed refractive index are used to analyze the accuracy of the inverse model.

- Table 1 summarizes the inputs used for each model, their outputs, the type of particle assumed, the assumptions of the model and the equations used.

The description of each figure and table has been introduced in the appropriate place of the text (usually, above each figure and table). Units to all equations and variables have also been added at the end of the equations.

**Comment:** I recommend the addition of a table of variables, symbols and units used. In many cases you discuss methodology mixed with the results. You even introduce new concepts such as BFGS later in the paper. All these are better placed in methods, and/or in a table/flow-chart such as I suggest. Admittedly, sometimes text flows better if you do introduce some of these methods later where you do, so this comment does not always apply.

**Answer:** As stated in the previous answer, several improvements have been added in the form of figures, flowcharts and tables, to clarify some parts of the paper. All units have also been added to the equations and some of the Figures (a table has not been added only for the units since there are only a few).

The reason for not introducing BFGS in more detail, which is another optimization method as it is the Genetic Algorithm, is because its performance is much worse than the Genetic Algorithm when applied on the examples described in the manuscript. The results were only added in order to state that other optimization algorithms were also tested (a part from BFGS we also tried with Nelder-Mead, Conjugated Gradient, etc.), but still the Genetic Algorithm provided the best solutions. We did not find necessary to make the paper longer with the introduction of a new method that does not present any meaningful improvement. As explained in the text, the Genetic Algorithm presents some disadvantages, being the convergence time the most important one, but its accuracy level is not achievable by any other algorithm.

In order to clarify the reason for not introducing in more detail the BFGS algorithm in the paper, the following text has been added in Section 4.1.3:

"Other optimization algorithms were also applied to determine if similar results can be obtained with a significant reduction of the computation time. **However, since none of them led to any meaningful improvement, they are not introduced here**. As an example, Fig. 7b shows the results…"

**Comment:** In the real world, given IOPs of a whole seawater sample, an average complex index of refraction would be retrieved by the presented methods. This average is weighted according to the PSD and the indices of refraction of the individual types of particles present. You should discuss this clearly somewhere and preferably also derive this weighting and state what is actually retrieved. See Eq. 8 in Boss et al. (2001) (see below), and refs. therein. This would be very useful to the ocean color research community.

**Answer:** The major goal of the paper is to point out:

1. the need to use alternative approaches for complex-shaped particles to those based on Lorenz-Mie (i.e. those based on *T*-matrix), and
2. to demonstrate that it is possible to get accurate retrievals of the refractive index with inverse methods based on these alternative methods, using as input parameters *in situ* measurements (IOPs, PSD and shape) that can be obtained routinely with existing oceanographic instrumentation.

In this first paper we have selected simplified synthetic scenarios that demonstrate these ideas. The practical question about how to apply these new ideas in real (and much more complex) scenarios implies many questions to consider, assuming the need to take into account complex-shaped particles: is it reasonable (or useful) to try to get a single averaged value in this case? Since complex particles cannot be characterised with a single volumetric parameter (the radius of the sphere) and there could be several refractive indices associated to each particle, which could be the best averaging approach? Maybe there will not be a single answer to this last question and the averaging methods would depend on the "dominant particle community". All these questions are very important to the ocean color research community, but will require a dedicated effort to try to answer them. Probably they are topics for developing new research lines in the future.

**Comment:** You need to discuss the applicability of these models to remote sensing data. Is it feasible for them to be applied to IOPs derived from ocean color remote-sensing reflectance? The problem with this may be that many operational remote-sensing inversion models for IOPs have in them an implicit or explicit assumption about the index of refraction when they were developed, so it would become a circular reasoning. Retrieving the index of refraction from space would improve our ability to distinguish sources of backscattering from each other in the ocean, so a paragraph in the discussion about that would be really important.

**Answer:** The proposed new inversion method use as input parameter in situ measurements (IOPs, PSD and shape) that can be obtained routinely with existing oceanographic instrumentation. We hope that these approaches will also have further applicability with other type of optical measurements (one case will be remote-sensing reflectance), but this topic is beyond the scope of the paper, as it implies a much more complex inversion scheme.

Nevertheless, we have mentioned this option in the final discussion:

"Besides, all these approaches can also have further applicability with other type of optical measurements, as for instance with remote-sensing reflectance. As an example, Fig. 19 shows the spectral backscattering of the three test cases, the homogeneous sphere, the coated sphere and the homogeneous cylinder Many operational remote-sensing inversion models for IOPs use an implicit or explicit assumption about the refractive index, and, in combination with these methods, could be severely improved. Retrieving the index of refraction from space would improve the ability to distinguish sources of backscattering from each other in the ocean. To this end, a much more complex inversion scheme should be developed".

**Comment:** You discuss the limitations of having limited degrees of freedom in one instance (Sect. 4.2.3). The same applies to multispectral sensors of several wavelengths only. What is the feasibility of retrieval of the bulk m value from space with the advent of hyperspectral sensors such as the planned NASA PACE mission?

**Answer:** In principle, the inverse method is applied iteratively to a single wavelength, so it will be possible to apply with both multi-spectral and hyperspectral in situ instrumentation. Regarding its use with remote sensing instrumentation, this is a two-step inversion problem. First, it is necessary to estimate the IOPs and PSD from the AOPs (RS) measurements (which, as it is mentioned in the previous comment, goes beyond the scope of the paper). Once the IOPs and PSD have been obtained, it is possible to apply the proposed method in the paper. Here, the main constrain is that the inverse method is sensible to inaccuracies of the input parameters, so the quality of the refractive index retrievals will depend strongly on the accuracy of the estimated IOPs and PSD in the first inversion step.

<u>More specific comments</u>

**Comment:** Title needs revision of word order and, more importantly, it needs to better reflect that this paper refers to aquatic optics.

**Answer:** Title has been modified as suggested here and in the pdf to: "Methods to Retrieve the Complex Refractive Index of Oceanic Particles: going beyond simple shapes"

**Comment:** Abstract: Needs major revision. Several sentences need to be added to set the context (aquatic optics), state that complex index of refraction determines IOPs and as such is input to forward models. Refractive indices are not easy to measure, thus are often assumed or retrieved with inverse modeling, etc.

**Answer:** Abstract has been completely rewritten to put the addressed questions and results in context, as well as the first part of the Introduction.

**Comment:** Introduction: You need to state more clearly how this test procedure works. I.e. do you start with particle(s) with known complex refractive indices, and then do a forward model (Mie, T-matrix, specify), than pass the IOPs to the inversion models and compare the results to the known inputs. Is this the scheme of your tests in this paper? It does not become very clear.

**Answer:** Yes, your description is correct. For the forward model, Lorenz-Mie or T-Matrix can be used. Lorenz-Mie, only for homogeneous or coated spheres, and T-Matrix, for the same kind of particles and also for more complex shapes.

To clarify this in the paper, we have added Figure 2 and its explanation in the Introduction:

"The comparison has been done following the three steps presented in Fig. \ref{Fig0b}. 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 wavelength-dependent refractive index, m(λ), the PSD and the particle shape). Then, the inverse models (described above) are used to estimate the refractive index using the IOPs obtained in the first step. Finally, the estimated refractive index is compared with the original one to obtain the accuracy of the inverse model."

Besides, in each example, some comments have been added. In example 1 -Section 4.1- (new comments highlighted in black):

"In particular, the BHMIE code, originally from Bohren and Huffman and modified by B.T. Draine, was used **as a Forward Model** (additional features were added, such as polydispersion and the computation of the Stokes scattering matrix)."

In example 2 (Section 4.2):

"Using this PSD with the previous refractive indices in the BART code from A. Quirantes (Quirantes, 2005) (**a Forward Model** based on the Aden-Kerker theory to calculate light-scattering properties for coated spherical particles), the absorption, scattering and extinction coefficients of Fig. 10a, and the volume scattering function of Fig. 10b were obtained.".

And, in example 3 (Section 4.3):

"In order to find which is the most accurate model for the characterization of such complex shapes, an example considering 100 prolate cylinders per $mm^3$ with a diameter-to-length ratio equivalent to 0.8, the PSD of Fig. 14 (showing the radius of an equivalent volume sphere with a slope parameter \xi = 3, effective radius $r_{eff}$=3.2 µm and effective variance $v_{eff}$=0.005 resulting in $R_{min}$ = 0.8 µm to $R_{max}$ = 3.6 µm), and the assumed refractive index of Fig. 4b was simulated using the T-Matrix algorithm (Mischenko1996, Mischenko1998) **as a Forward Model**".

**Comment:** pg. 18730, line 11, eq. 14 – this needs a better explanation because it is confusing as it is presented. Do you mean that the values in Eq. 14 are already the values relative to seawater, as you use them throughout the paper? Show the actual equation to calculate the relative index, given the complex indices of the particle and the medium.

**Answer:** Yes, this is what it means, as a matter of convention. We have followed the same assumption with respect to the normalization of the refractive index as in other publications on light scattering. See for instance Bernard et all 2009, Twardowski et all 2001, etc.

We have clarified the text just before Section 3.1:

"Note that this paper assumes effective refractive indices relative to seawater, which has a constant value of $m_{water} = 1.334 + i0$ (Hale et al, 1973). Absolute values can be recovered using $m_a = m \times m_{water}$."

**Comment:** Eq. 15 – I believe the Twardowski model retrieves just the real part of the refractive index, not something equivalent to |m|.

**Answer:** It is not clear, since all along the paper only refers to it as the "bulk refractive index". However, we have considered this suggestion (and the one related to the imaginary part, which is assumed to be 0.005), and have applied the following modification:

- Fig. 6a is shows the real part of the refractive index instead the absolute value, and also compares the imaginary part.
- Table 2 shows the error on the real and imaginary parts of the refractive index.
- Descriptions of Fig. 6a (Section 4.1.1) and Table 2 (Discussion).

**Comment:** pg. 18731, line 2: Fig,. 1a caption in Twardowski et al. (2001) says that k was fixed at 0.005, therefore they do not ignore/neglect k in this model, which is a fit to their Fig. 1a.

**Answer:** It has been corrected (see answer above).

**Comment:** pg. 18734, line 3 – you use inconsistent notation for the complex index of refraction – it is 'm' above, and 'n' here. Please use consistent notation everywhere.

**Answer:** I has been corrected.

Additional comments

**Comment:** Sect. 4.2 Why not apply the Twardowski model to these coated particle IOPs and see how the retrieved bulk n compares to the input ones? I.e., why can't the Twardowski model be applied to the cases other than homogeneous spheres?

**Answer:** It must be considered that the Twardowski model was developed to apply to entire particle populations that are assumed to follow a power-law size distribution, and is fundamentally different from the other models, developed to apply to isolated phytoplankton cultures which, in general, follow a different size distribution.

In our contribution, only the first example uses a power-law size distribution, precisely, to apply the Twardowski model and compare its results with the other models. The other two test cases follow particle size distributions more similar to those presented by monocultures, and therefore, the Twardowski model was not used in order to avoid unfair comparisons.

To clarify this in the text, the following sentence has been added just before Section 4.2.1:

"Note that the Twardowski model is not applied to avoid unfair comparisons with the other methods (it was designed to be used with entire particle populations that are assumed to follow a power-law size distribution)."

**Comment:** The Bernard (2009) reference is not given in the list of references.

**Answer:** It did not appear due to a problem of compilation in Latex. It has been solved.

**Comment:** In all figures with the output IOPs, consider showing spectral backscattering as well – would be very useful for remote sensing relevant applications.

**Answer:** Although it is possible to add extra plots showing spectral backscattering, we wandered up to which point it would be really relevant for remote sensing applications, taking into account that all tests cases in the paper were developed using synthetic and probably too simplistic data to compare with real remote sensing products. For the time being, we have only added the spectral backscattering of the three test cases in the Discussion Section.

**Comment:** Sect. 2.1: Whenever you use equations not derived by you, please give citations.

**Answer:** They have been added (except when it is obvious where do they come from; for instance, in Section 3.1, The Twardowski Model, the whole Section is referred to Twardowski et al., 2001. In such cases, references have been omitted).

**Comment:** + or – in the m = n+ik expression? Different sources list it differently. So it would be useful if you clarify this.

**Answer:** The sign of the complex part is a matter of convention. The notation in the paper, which is usually used by physicists, corresponds to waves with time evolution given by $e^{-i\omega t}$.

In the paper, it has been clarified in the second paragraph of Section 3 with:

"The sign of the complex part is a matter of convention (it can also be defined with the negative sign). The notation above corresponds to waves with time evolution given by $e^{-i\omega t}$".

<u>References</u>

**Comment:** Boss E., M.S. Twardowski, and S. Herring, 2001. Shape of the particulate beam attenuation spectrum and its relation to the size distribution of oceanic particles. Applied Optics, 40, 4885-4893

**Answer:** According to our findings, the exact name of this reference is "Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution", as seen in:

https://www.osapublishing.org/ao/abstract.cfm?uri=ao-40-27-4885

This reference already appeared in the manuscript. If it is not correct, please, tell us and we will modify it.

**Comment:** Please also note the supplement to this comment:

http://www.biogeosciences-discuss.net/12/C9713/2016/bgd-12-C9713-2016-supplement.pdf

**Answer:** All comments have been considered and applied, but some specific comments need to be answered. For instance:

- Just before equation 19, "explain what xi is". Xi was already defined in the first paragraph of Section 3.1 (The Twardowski Model).
- In Fig. 3, "say what conv. means." It is already explained in Section 3.4: "After the evaluation, the algorithm may stop if either a maximum number of generations (each

generation is a new vector of solutions) or a satisfactory fitness level have been reached. If the convergence condition is not fulfilled, the best solutions are selected and separated…"

- Section 2.2, "most authors give D+dD, explain your choice". It is a matter of convention. We follow the same definition as in Bricaud and Morel, 1986, Bernard et al, 2009, etc. A reference has been added.
- First paragraph of Section 4, "how does this combination work?" This is only an introduction of the Section. Please, refer to Section 4.2.3 for further details.
- Section 4.1.3, "are these code, or algorithms? Explain the acronyms." As stated, they are frameworks for programming. Acronyms have been explained. References may provide further information for the interested readers.
- Section 4.1.3, "Isn't this a very small number of generations. I think typically many more are used. Justify your choice." That's a good point, since normally it is usual to achieve at least 100 generations. Indeed, it strongly depends on the length of the initial-population vector and cross-over and mutation percentages, among other parameters of the genetic algorithm. In this particular case, using an initial vector of 2000 solutions and 50 and 20% of probability of crossover and mutation, not significant improvements are generally found beyond the tenth generation. It has been clarified in the text at the end of the first paragraph of Section 4.1.3.
- In Section 4.2, "why a peak at 500 nm, Chl does not normally have a peak of absorption there? is it accessory proteins/pigments?". Some examples shown in Stramski et al, 2001, and Bernard et al, 2009, exhibit similar peaks (with a lower amplitude, to be honest) around 500 nm. We tried to emulate such refractive indices in order to have assumed values close to those measured in nature.
- Section 4.1.2 and 4.2: "Explain Aden-Kerker and Hilbert Transform". In order to not further extend the length of the paper, references have been added for both mathematical definitions.

Manuscript prepared for Biogeosciences
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 5 April 2016

# Methods to Retrieve the Complex Refractive Index of Oceanic Particles: going beyond simple shapes

A.-M. Sánchez[1] and J. Piera[1]

[1]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 oceanic particles have many practical applications in oceanic optics. In particular, several methods have been proposed to estimate important features of the particles by inversion from optical measurements, 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, in which 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, the performance of several inversion methods for the retrieval of the refractive indices are compared considering three different synthetic scenarios. The error associated with each method is discussed and analyzed. The obtained results suggest that inverse methods using the $T$-matrix approach are useful to retrieve the refractive indices of complex shapes (i.e., many phytoplankton species) accurately.

## 1 Introduction

Light interactions with oceanic particles are the processes by which changes on the composition of small particles suspended in the water column (such as phytoplankton, sediment or microplastics), cause optical observable phenomena that usually are wavelength dependent (for example, changes on the ocean color or the light extinction with depth) (Andrady, 2011; Cole et al., 2011). Understanding the interaction of light with particles is the central topic in many bio-optical studies in which the water particle composition is inferred from $in$-$situ$ or remote sensing optical observations.

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 with arbitrary size parameter is explained analytically by the Lorenz–Mie (also known as Mie) theory (Lorenz, 1898; Mie, 1908). For more complex shaped particles, scattering can be computed using the $T$-Matrix theory (Waterman, 1965). At present, the $T$-Matrix method is the fastest exact technique for the computation of nonspherical scattering based on a direct solution of Maxwell's equations (Mischenko et al., 1996).

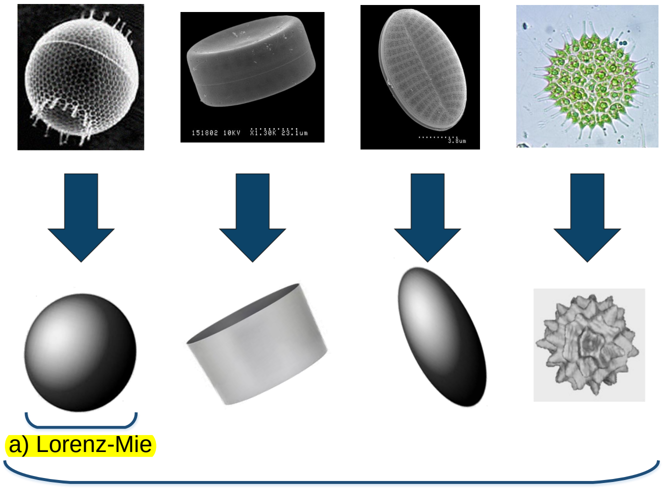



**Figure 1.** Lorenz-Mie (a) can be used to model spherical shapes, while $T$-Matrix (b) can be applied to model more complex shapes such as spheroids, cylinders or Chebyshev particles. As shown by Clavano et al. (2007), aspect ratios of phytoplankton (ratio of the principal axes of a particle) span between 0.4 and 72.

Although there are some geometrical restrictions, such as axial symmetry, it is applicable to many life forms of phytoplankton (Quirantes and Bernard, 2006; Stramski et al., 2001) and suspended mineral particles (Twardowski et al., 2001), as shown in Fig. 1.

Both approximations to solve Maxwell's equations share one important requirement: the inner complex refractive index of the particles must be known. Although new promising techniques such as Tomographic Phase Microscopy (Choi et al., 2007) may provide in the future measurements of the refractive index in live cells, at present current ocean instrumentation do not directly provide it, so it must be estimated somehow (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 is used by Twardowski et al. (2001), to estimate 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 dominance of one particular phytoplankton species). It is based on the anomalous diffraction approximation (ADA), which allows the computation of the real and imaginary parts 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). And 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 as 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

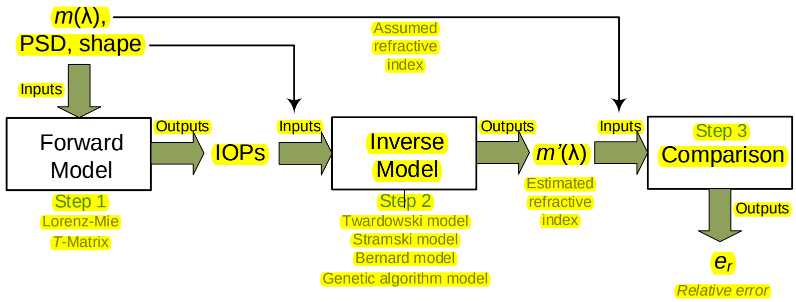



**Figure 2.** Procedure to analyze the accuracy of the inverse models.

of a genetic algorithm with the Lorenz-Mie and $T$-Matrix approaches was used by Sánchez et al. (2014), thereby allowing more complex structures than simple homogeneous or coated spheres. A genetic algorithm is a search heuristic for optimization problems that simulates the process of natural selection using inheritance, mutation, selection, and crossover between different possible solutions. Again, this method only requires the measured attenuation and scattering coefficients, and the PSD to find the complex refractive index. This later method is much slower than the previous ones (in particular, for non-spherical particles), but it can provide very accurate estimations.

In this paper, the refractive index retrieval models mentioned above 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 assumed wavelength-dependent refractive index, $m(\lambda)$, the PSD and the particle shape). Then, the inverse models (described above) are used to estimate the refractive index using the IOPs obtained in the first step along with the PSD and particle shape. Finally, the estimated refractive index is compared with the assumed one to obtain 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 a homogeneous sphere, a coated sphere and a homogeneous cylinder. None of these idealized models is an exact representation of a real algae presenting cell walls, chloroplasts, vacuole, nucleus and other internal organelles, each with its own optical properties. However, they can be considered 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 while the other models are developed for single phytoplankton cultures (or dominance of one particular phytoplankton species) and require the concurrent measurement of the size distribution. And, besides, such bio-optical models are compared with a numeric method (i.e., the genetic algorithm) in the same conditions. On the other hand, the methodology applied in this contribution allows to obtain an objective comparison of the results of the different methods in those occasions where it is not clear which methodology is most suitable, and therefore, interesting for the ocean optics community.

In order to establish the foundations of the work presented in this paper, Section 2 reviews the formulation to obtain the IOPs from Lorenz-Mie and $T$-Matrix characterizations (that perform the forward calculations) for polydispersed algal assemblages. In Section 3, a review of the different inverse approximations to retrieve the refractive index is described. In Section 4, all the models are used to retrieve the refractive index of three assumed particles in polydisperse assemblages. Section 5 discusses the results and finally, the conclusions are outlined in Section 6.

## 2 Model Theory

### 2.1 Size Distributions and Polydispersions

Algal assemblages are typically polydispersed 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 the other IOPs), the absorption efficiency factor representing the mean of a size distribution can be described as (Bricaud and Morel, 1986):

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

The relationship between the absorption efficiency factor and the absorption cross section is:

$$Q_a = \frac{C_a}{G}, \tag{2}$$

being $G$ the geometric cross section of the particle. And the resultant volume absorption coefficient is given by either:

$$a = \frac{\pi}{4}\int_0^\infty Q_a(D)F(D)D^2 d(D) \qquad [m^{-1}], \tag{3}$$

or, if the result of Eq. (1) is used:

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

### 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 (which describe the area of the incident-beam intensity converted to extincted, scattered or absorbed light), $C_{EXT}$, $C_{SCA}$ and $C_{ABS}$ respectively. Using the obtained cross sections (size-averaged in polydisperse concentrations), the wavelength-dependent extinction, scattering and absorption coefficients ($c(\lambda)$, $b(\lambda)$ and $a(\lambda)$ respectively) can be computed as (Quirantes and Bernard, 2006):

$$c(\lambda) = N \cdot \langle C_{EXT}(\lambda) \rangle \qquad [m^{-1}], \tag{5}$$

$$b(\lambda) = N \cdot \langle C_{SCA}(\lambda) \rangle \qquad [m^{-1}], \tag{6}$$

$$a(\lambda) = N \cdot \langle C_{ABS}(\lambda) \rangle \qquad [m^{-1}], \tag{7}$$

where $N$ denotes the number of particles per unit volume and $\lambda$ the wavelength. The relationship between the three parameters is:

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

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

$$\beta(\Psi,\lambda) = \widetilde{\beta}(\Psi,\lambda) \cdot b(\lambda) \qquad [m^{-1} sr^{-1}]. \tag{9}$$

$\Psi$ is the scattering angle (i.e., the angle between the incident and scattered beams) and $\widetilde{\beta}(\Psi,\lambda)$ is the scattering phase function and the first parameter of the Stokes scattering matrix (or Mueller matrix). This matrix transforms the Stokes parameters of the incident light into those of the scattered light and it is obtained with the Lorenz-Mie and $T$-Matrix formulation when the physical characteristics of the particles are known. The integral scattering in all directions, assuming azimuthal symmetry, retrieves the total scattering coefficient $b$:

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

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

$$B_b(\lambda) = \frac{b_b(\lambda)}{b(\lambda)}. \tag{11}$$

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

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

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

$$B_b(\Psi,\lambda) = 2\pi \int_{\frac{\pi}{2}}^{\pi} \widetilde{\beta}(\Psi,\lambda) sin(\Psi) d\Psi. \tag{13}$$

The $2\pi$ factor used in Eq. (10), Eq. (12) and Eq. (13) (which comes naturally from integration with respect to the azimuth angle) is different from the $1/2$ factor used by Mischenko et al. (1996); Mischenko and Travis (1998); Wiscombe and Grams (1976); Mugnai and Wiscombe (1986) (where the integration of phase function is normalized to $4\pi$, representing the total solid angle over entire sphere), but used by Twardowski et al. (2001), Bohren and Huffman (1998), and most of the literature in Ocean Optics (Mobley, 1994), and therefore, applied here.

## 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 method is named using the surname of the lead author of the publication. In order to make the understanding of this section easier, Table 1 shows, for each model, their inputs and outputs, type of particles, as well as the assumptions of the model and the equations used.

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

| Inverse models | Inputs | Outputs | Type of particles | Assumptions | Equations |
|---|---|---|---|---|---|
| **Twardowski Model** | $B_b$, $\xi$ | $n$ | Homogeneous spheres | $k = 0.005$, Power-law PSD, $gamma = \chi - 3$ | (9), (15) |
| **Stramski Model** | 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) |
| **Bernard Model** | $a$, PSD | $n_{chlor}$, $k_{chlor}$ | Coated spheres | $n_{cyto}$, $k_{cyto}$, $V_V$, $1 + \epsilon$ | (16), (17), (20), (21) |
| **Genetic Algorithm Model** | $a$, $c$, PSD | $n$, $k$ | Homogeneous and coated spheres, spheroids, cylinders and Txebixev particles | Particles with axial symmetry | (2), (3), (22) |

The complex refractive index $m(\lambda)$ is defined as:

$$m(\lambda) = n(\lambda) + ik(\lambda), \tag{14}$$

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 (it can also be defined with the negative sign). The notation above corresponds to waves with time evolution given by $e^{-i\omega t}$. Note that this paper assumes effective refractive indices relative to seawater, which has a constant value of $m_{water} = 1.334 + i0$ (Hale and Querry, 1973). Absolute values can be recovered using $m_a = m \times m_{water}$.

### 3.1 The Twardowski Model

The Twardowski model, presented by (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, 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 using a polynomial fit to the output of Lorenz-Mie calculations as:

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

This formulation is only exact for particles which size spans from 0 to infinity with a constant absorption along wavelength ($k$ is held constant at 0.005) and are homogeneous spheres. It was first tested by Boss et al. (2001a) and refined in Boss et al. (2001b). It must be noted that the model is consistent with the measurements obtained from an AC9 with the scattering coefficient $b$ serving as integrated scattering from 0.93 to 180 degrees, which must be considered in Eq. (13). Even though this was not firstly considered in the calculations in Twardowski et al. (2001), it was taken into account in Boss et al. (2004), but the regression was not recomputed.

### 3.2 The Stramski Model

This model is based on the methods presented by Stramski et al. (1988), which is an extension of that 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 wavelength ($\alpha = \frac{\pi D}{\lambda} \gg 1$) and the refractive index is small ($n - 1 \ll 1$ and $k \ll 1$). This method allows the effects of the real and imaginary refractive indices on absorption and scattering to be decoupled. Assuming homogeneous geometry, the ADA expression for the absorption efficiency factor is given by:

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

where $\rho' = 4\alpha k$ is the absorption optical thickness. Eq. (4) 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 part of the real refractive index $1 + \epsilon$. Thus:

$$n(\lambda) = 1 + \epsilon + \Delta n(\lambda). \tag{17}$$

The central value $1+\epsilon$ is estimated by computing the nonabsorbing 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:

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

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

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

The exact value of $\epsilon$ is indicated by such $\bar{Q}_c^{NAE}(\bar{\rho})$ that it equals $\bar{Q}_c(\lambda_\epsilon)$.

This methodology was latterly simplified by Bernard et al. (2001, 2009) by using the Kramers-Kronig relations to compute the spectral variations in the real part of the refractive index instead the Lorentzian oscillators. The Kramers-Kronig relations describe the mutual dependence of the real and imaginary parts of the refractive index through dispersion, as does Ketteler-Helmholtz theory, but they are more simply applied 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. In 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:

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

where $k_{cyto}(400) = 0.0005$. The real refractive index spectra for the cytoplasm, $n_{cyto}(\lambda)$, is 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 the $k_{cyto}(\lambda)$ of Eq. (20), volume equivalent $k_{chlor}(\lambda)$ are determined using the Gladstone-Dale formulation given by:

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

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 previous 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 similarly generated with a Hilbert transform and Eq. (17) with assumed $1+\epsilon$ values between 1.044 and 1.14 depending upon 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 ($a$ and $b$) when using the Lorenz-Mie or $T$-Matrix approaches with the measured PSD. The methodology of the algorithm can be summarized as follows (see Fig. 3). First, a random vector of solutions is generated for a specific wavelength ($[m_1(\lambda_i)$, $m_2(\lambda_i)$, ..., $m_n(\lambda_i)]$, where $\lambda_i$ denotes the selected wavelength and $m_1, m_2, ..., m_n$ the complex refractive indices). If possible, the search space can be bounded in order to maximize the algorithm success. Then, the complete vector is evaluated by the fitness function. This is done by computing the $a$ and $b$ 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 obtained, for instance, as:

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

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 committed error for the absorption coefficient. Using logarithmic values allows a suitable weighting factor when dealing with small errors over small coefficients. The same equation can be used for the scattering coefficient. Both results are finally combined using the quadratic mean, obtaining a single evaluation value that the algorithm tries to minimize.

After the evaluation, the algorithm may stop if either a maximum number of generations (each generation is a new vector of solutions) or a satisfactory fitness level have been reached. If the convergence condition is not fulfilled, the best solutions are selected and separated. 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 avoid becoming stuck at 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 convergence is achieved, the algorithm presents the best solution it has been able to find.

Since 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

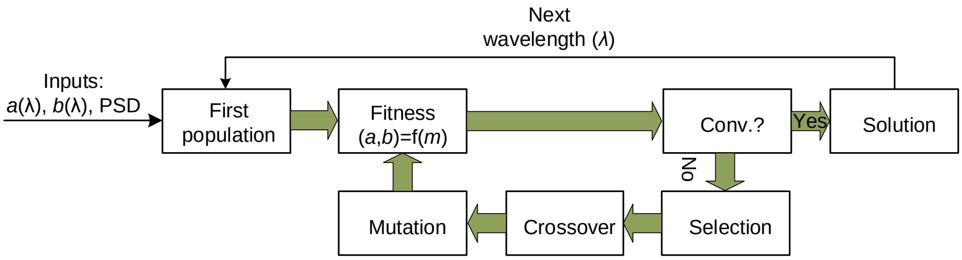



**Figure 3.** Flow chart of the genetic algorithm.

genetic algorithm performs the search procedure at a single wavelength each time. After each convergence, the process starts again with the next wavelength-dependent values, eventually obtaining the complete complex-refractive-index signature.

The main advantage of this model 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, cylinders, etc. Besides it can also be easily combined with other models to improve the results. On the other hand, it must be noted, that some inversions could 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 the aforementioned criteria (Bhandarkar et al., 1994). In that case, techniques such as regularization methods can be applied to improve the results (Mera et al., 2004).

## 4 Experimental 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

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

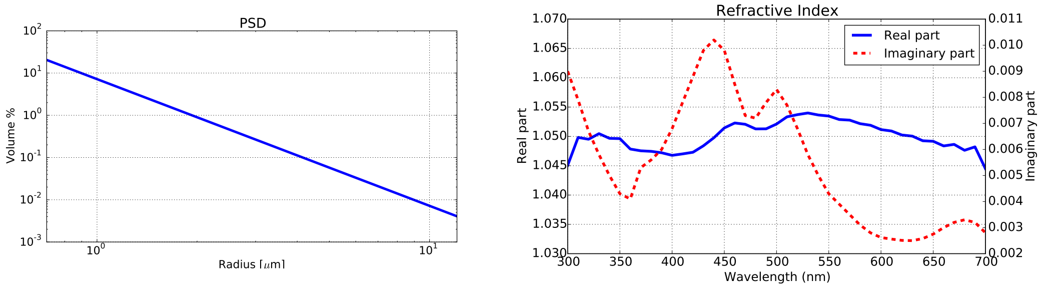
where $x'$ accounts for the estimated real part of the refractive index as $n'(\lambda) - 1$ (the unity is subtracted to only consider the decimals) or the estimated imaginary part of the refractive index as $k'(\lambda)$, and $x$ accounts for the assumed real part of the refractive index as $n(\lambda)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

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

To this end, Subsection 4.1 deals with a simple spherical and homogeneous particle and presents the results provided by the Twardowski, Stramski and Genetic Algorithm models. Such particles, however, are not a fair representation of phytoplankton particles. First, because eukaryotic phytoplankton cells are heterogeneous particles with membrane systems and intracellular organelles, and second, because most of the phytoplankton species are not spherical. As stated by Bernard et al. (2009), the spherical structure mainly fails in the description of the backward scattering and suggests a two-layered spherical geometry as the simplest possible heterogeneous structure capable of reproducing measured algal angular scattering properties. In consequence, Subsection 4.2 presents a coated sphere as the assumed particle and presents the estimated refractive indices provided by the genetic algorithm model, the Bernard model and a combination of both. Finally, Subsection 4.3 uses a cylindrical shape particle with a homogeneous refractive index. This shape was selected to be different from a sphere and similar to that of some species of phytoplankton (as for instance, the diatom $Thalassiosira\ pseudonana$). Although this assumed model does not exactly reproduce the same optical behavior as the actual phytoplankton particle (the micro-details of the cells are neglected), it can serve as a first approximation. Refractive index estimations provided by the combination of the Genetic Algorithm with the Bernard model for coated spheres and the Genetic Algorithm alone are shown.

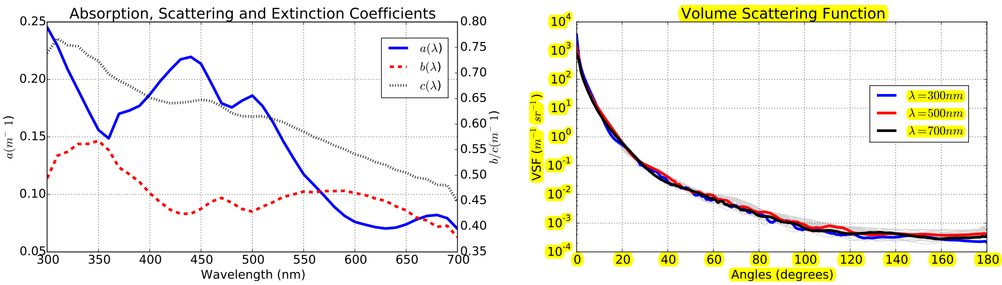
### 4.1 Spherical-Shaped Homogeneous Particles

A concentration of 100 spherical particles per $mm^3$ presenting the PSD of Fig. 4a (based on a power-law distribution -or Junge-type- with 51 points, $R_{min} = 0.7\ um$, $R_{max} = 12.1\ um$, a slope parameter $\xi = 3$, effective radius $r_{eff} = 4\ \mu m$ and effective variance $v_{eff} = 0.6$), along with the assumed complex refractive index of Fig. 4b, was simulated using the Lorenz-Mie scattering theory (Bohren and Huffman, 1998). In particular, the BHMIE code, originally from Bohren and Huffman and modified by B.T. Draine, was used as a Forward Model (additional features were added, such as polydispersion 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 can 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). Although the power-law distribution is not a realistic distribution for single phytoplankton species, it is a fairly good approximation of natural-water composition (even with anomalous natural conditions such as a phytoplankton bloom), as there is always a strong background contribution to the PSD (Twardowski et al., 2001). Besides, it is the only distribution that can be used in the Twardowski model, and therefore, used here. The complex refractive index of 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 on the imaginary part of the refractive index can be found in the Kramers-Kronig relations (Bernard et al., 2001, 2009; Bricaud and Morel,




(a) (b)

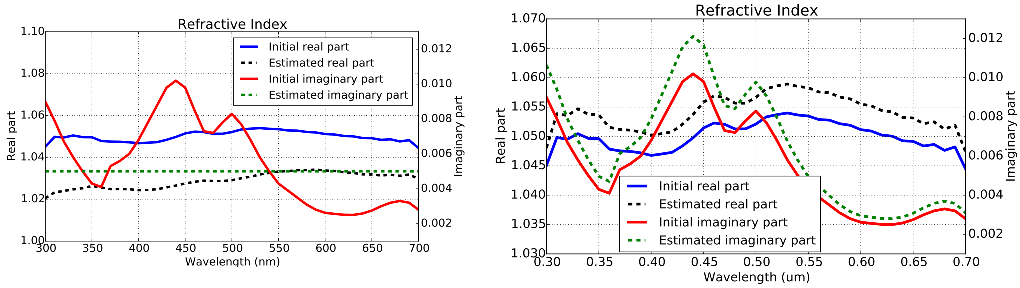
**Figure 4.** (a) PSD of the test run with spherical-shaped homogeneous particles. (b) Assumed complex refractive-index signature of the example with spherical-shaped homogeneous particles.




(a) (b)

**Figure 5.** (a) Absorption ($a$), scattering ($b$) and extinction ($c$) coefficients of the example with spherical-shaped homogeneous particles, and (b) the volume scattering function.

1986; van de Hulst, 1957), which allow the spectral variations in the real refractive index to be calculated as a Hilbert transform of the imaginary refractive index. The central part of the real refractive index was selected as $1+\epsilon = 1.05$ (for phytoplankton it 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), have not been considered. As can be seen, the imaginary part presents three peak values, at 440, 500 and 680 $nm$ (corresponding to the chlorophyll absorption wavelengths), and, as expected, a similar shape is propagated to the absorption coefficient spectra, $a(\lambda)$, of Fig. 5a. The volume scattering function is shown in Fig. 5b. Only three wavelengths (300, 500 and 700 $nm$) are plotted using intense colors, while the rest of wavelengths between 300 and 700 $nm$ in steps of 10 $nm$ are plotted in light grey. As expected (since particles are relatively large regarding to wavelength), the scattering is mainly focused in the forward direction (between 0 and 10 degrees) and smoothly decreases in the backward direction.




(a) (b)

**Figure 6.** (a) Assumed and estimated refractive indices using the model of Twardowski et al. (2001). (b) Assumed and estimated refractive indices using the Bernard model.

#### 4.1.1 The Twardowski Model

Eq. (15) was applied to this example obtaining the results shown in Fig. 6a. To this end, $\gamma = 0$ since the slope parameter of the PSD $\xi = 3$, and the backscatter fraction was computed with Eq. (13) using the volume scattering phase function values given by the modified BHMIE code. As can be seen, for the real part, this model obtains a curve shape similar to the assumed 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 $0.005$, the averaged relative error with the assumed imaginary part of the refractive index is $44\%$. It must be noted that the Twardowski model was designed for a bulk oceanic distribution presenting different physical properties than those of isolated species of phytoplankton (e.g. index of refraction, shape, etc.), and therefore, it is used here in a different scenario than it was designed to.

#### 4.1.2 The Stramski Model

The results obtained with this model are shown in Fig. 6b. As can be seen, this model overestimates both real and imaginary parts on all the analysed spectra, 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 $\sim 10\%$ in comparison to Lorenz-Mie theory (Bernard et al., 2009), and some discrepancies can therefore be expected between ADA and Aden-Kerker derived values (Aden and Kerker, 1951).

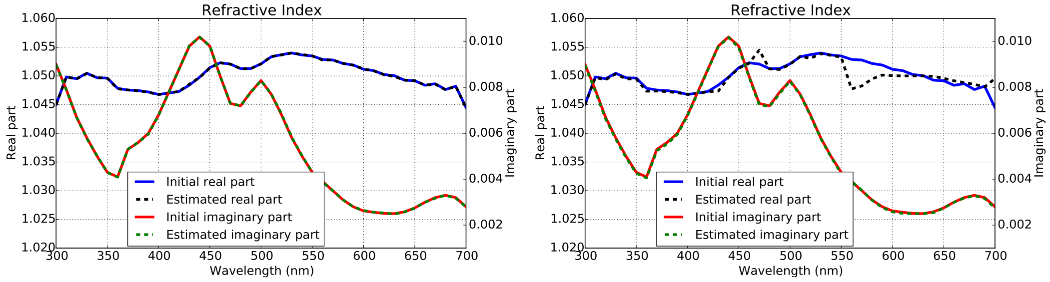
#### 4.1.3 Genetic Algorithm

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 Operations in Python) frameworks to develop evolutionary algorithms and parallel task distribution respectively, were used (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 properties of homogeneous spheres. The coefficients $a$ and $b$ of Fig. 5a were used as inputs of the genetic algorithm model to estimate the assumed complex refractive index and bounding 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 solutions over 10 generations and $50\%$ and $20\%$ of probability of crossover and mutation respectively, obtaining the estimated values shown in Fig. 7a. The good agreement between the assumed complex-refractive-index values and the estimated ones (an averaged relative error of $0.004\%$ for the real part and $0.24\%$ for the imaginary part is obtained, presenting thus the best results in this first example) shows that it is possible to perform accurate estimations with a genetic algorithm. It must be noted that the number of generations needed to have 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. In this particular case, using the described parameters, any significant improvement is generally found beyond the tenth generation.

One disadvantage of the genetic algorithms is that they are relatively slow and require more computation time than other optimization algorithms, since they need to execute the fitness function many more times. Other optimization algorithms were also applied to determine if similar results can be obtained with a significant reduction of the computation time. However, since none of them led to any meaningful improvement, they are not introduced here. As an example, Fig. 7b shows the results obtained with the much faster Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (BFGS is an 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 224 seconds (less than 4 minutes) were needed in front of the 97 minutes used by the genetic algorithm, both in a PC with an Intel Core i7 processor at 3.2 GHz, a 16-GB RAM and running a Windows 8.1. However, although the results are quite satisfactory in general, 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 $0.073\%$ for the real part and $0.72\%$ 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 a worse accuracy than the BFGS, showing that the genetic algorithm is probably the optimal method to solve this problem in terms of accuracy (but not in terms of time).

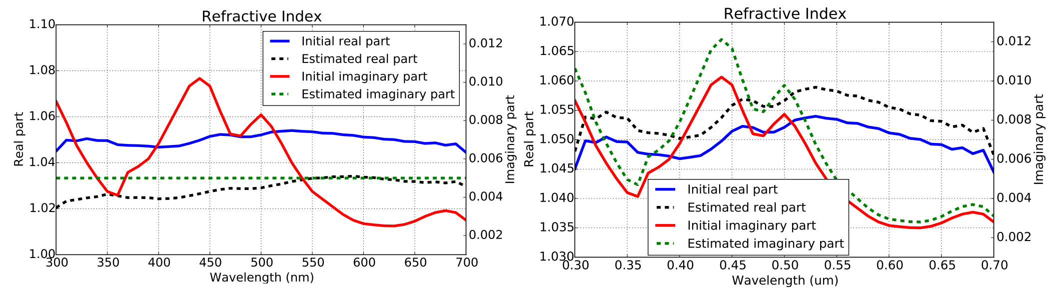
### 4.2 Spherical-Shaped Coated Particles

In order to use the IOPs of a two-layered spherical particle that emulates actual phytoplankton properties, 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




(a) (b)

**Figure 7.** (a)Assumed and estimated refractive indices using the Genetic Algorithm presented in this paper. Note that assumed and estimated values are on top of each other. (b) Assumed and estimated refractive indices using the BFGS algorithm.




(a) (b)

**Figure 8.** (a) Assumed real refractive-index signatures for the inner and outer layers. (b) Assumed imaginary refractive-index signatures for the inner and outer layers.

refractive index of the outer chloroplast was obtained using Eq. (21), with $V_V = 30\%$ (since it is a value between that assumed by Bernard et al., 2009, and previous works), and its real one 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. 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 $mm^3$ ($R_{min} = 0.7\ um$, $R_{max} = 12.1\ um$ and using 31 points), as seen in Fig. 9. It must be noted that the PSD denotes the external radius (the inner one can be calculated using the $V_V$ value). Using this PSD with the previous refractive indices in the BART code from A. Quirantes (Quirantes, 2005) (a Forward Model based on the Aden-Kerker theory to calculate light-scattering properties for coated spherical particles), the absorption, scattering and extinction coefficients of Fig. 10a, and the volume scattering function of Fig. 10b were obtained.

Below, the IOPs presented above are used to estimate their complex refractive indices. First, this is done using the genetic algorithm in order to see if a basic shape such as a homogeneous sphere is useful when modelling more complex particles. If coated particle models better characterize the

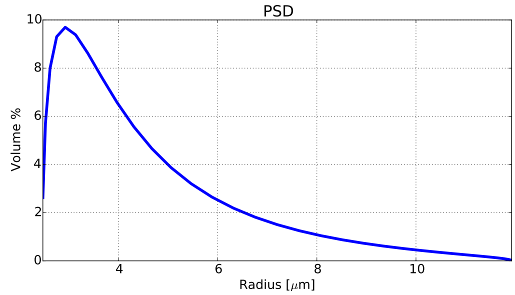
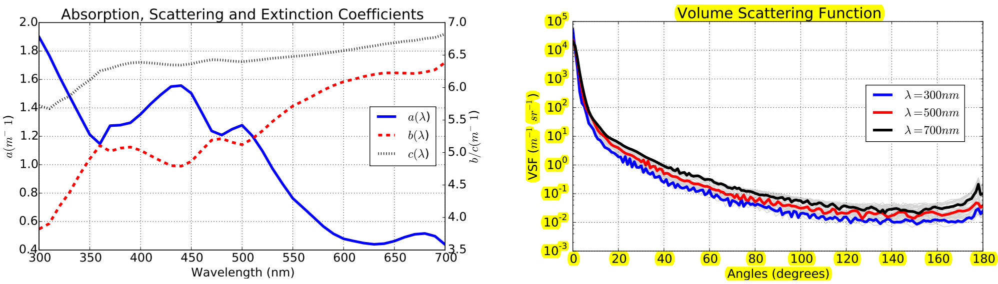



**Figure 9.** PSD that simulates a an isolated culture.



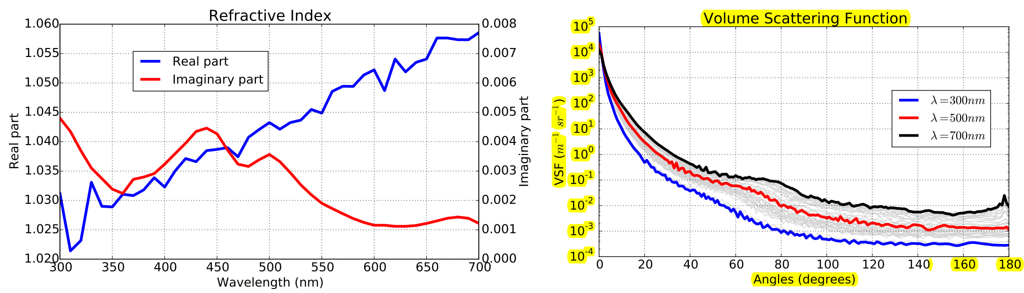

(a) (b)

**Figure 10.** (a) Absorption ($a$), scattering ($b$) and extinction ($c$) coefficients of the coated-particle example, and (b) the volume scattering function.

optical properties of general phytoplankton species, as stated in Bernard et al. (2009), this can be used to estimate the error committed when using spheres. Then, the inner and the 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 is applied to improve the previous results. Note that the Twardowski model is not applied to avoid unfair comparisons with the other methods (it was designed to be used with entire particle populations that are assumed to follow a power-law size distribution).

#### 4.2.1 The Genetic Algorithm

The genetic algorithm model to retrieve the refractive index of spherical-shaped homogeneous particles was applied in order to measure the error committed in such approximation. The same configuration as in the previous example was used (an initial vector of 2000 solutions over 10 generations and 50% and 20% of probabilities for crossovers and mutations respectively). The estimated complex refractive index is shown in Fig. 11a. Both real and imaginary parts present values between the inner and the outer real and imaginary parts of Fig. 8a and Fig. 8b respectively. The volume scattering function generated by the homogeneous particles, as seen in Fig. 11b (obtained by means of a forward model, i.e., Lorenz-Mie, using the estimated complex refractive index and the PSD




(a) (b)

**Figure 11.** (a) Complex refractive-index signature estimated using the genetic algorithm model in the spherical-Shaped coated-particle example. (b) Volume scattering function using the estimated refractive index in the second example.

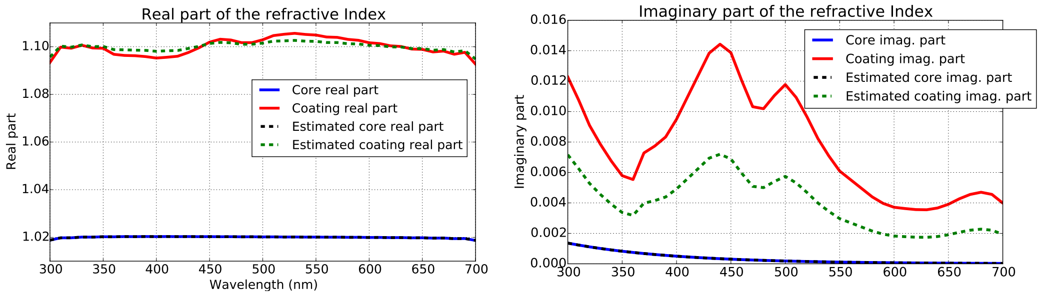
of Fig. 9 as inputs), shows that this model presents similar values in the forward scattering but completely underestimates the backscattering, presenting values far below those of Fig. 10b. This example demonstrates that the common characterization using homogeneous spheres is not a suitable methodology when dealing with complex particles. Even though it is not a surprising result (this is well known and has been discussed for years, by Bohren and Huffman, 1998, in the atmospheric literature, and by Stramski et al., 2004, Clavano et al., 2007, Dall'Olmo et al., 2009, and Bernard et al., 2009), in the oceanic literature), a comparison between the two volume scattering functions manifests that the backscattering can exhibit errors up to one order of magnitude.

#### 4.2.2 The Bernard Model

The Bernard model of Section 3.3 was used to estimate the complex refractive index of the two-layered particle. Figure 12a shows the assumed and estimated real part of the inner and outer layers and Figure 12b shows the assumed and estimated imaginary parts. As expected, the inner refractive index is well estimated (since the same equation is used for both generation and retrieval), but the outer refractive index does not present an accurate agreement. In particular, the imaginary part is significantly underestimated, with an averaged relative error of 51%. On the other hand, the simulation of the estimated refractive indices in coated spheres produce a volume scattering function which is in a better agreement with that of Fig. 10b than the volume scattering function produced by the homogeneous spherical particle (the volume scattering function figure has not been added in this case since errors are not apparent on the graph; a more detailed analysis is done in Section 5).

#### 4.2.3 The Bernard Model combined with Genetic Algorithm

In order to improve the results presented by the Bernard model in the previous subsection, the genetic algorithm, which showed a reasonable performance when applied to homogeneous spherical particles, could be coupled to the BART code (instead the BHMIE code) to try to estimate the two




(a) (b)

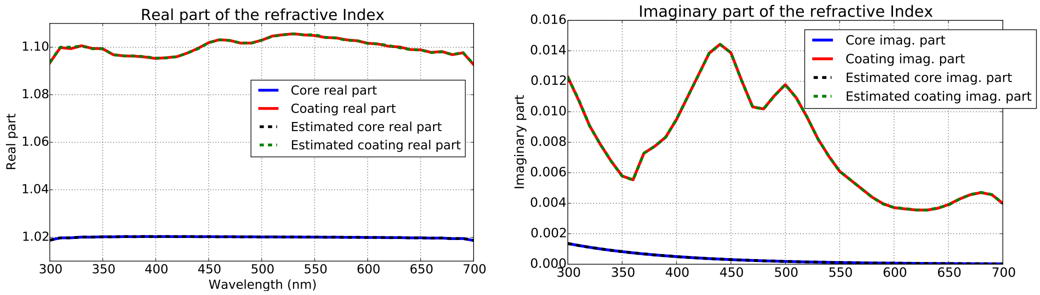
**Figure 12.** (a) Assumed and estimated real part of the refractive indices for the inner and outer layers using the Bernard model. (b) Assumed and estimated imaginary part of the refractive indices for the inner and outer layers using the Bernard model.

complex refractive indices. However, results would hardly be constrained since the solution has more degrees of freedom (the two refractive indices with real and imaginary parts each, that is, four dimensions) than the information data (the attenuation and scattering coefficients, that is, two dimensions), i.e., this is an unconstrained (ill-posed) problem. However, there is the possibility to combine the genetic algorithm with the Bernard model to increase the convergence probability. In this case, the inner refractive index is firstly estimated using the Bernard model, as it was done before, and the outer refractive index is obtained secondly with the genetic algorithm (coupled to the BART code). In this case, the genetic algorithm only has to find a solution with two dimensions (the real and imaginary parts of the outer refractive index).

This method was applied on 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), obtaining the assumed and estimated real part of the inner and outer layers shown in Figure 13a and the assumed and estimated imaginary parts shown in Figure 13b. As it can be seen, accurate results were obtained, meaningfully improving the refractive index estimation for the outer sphere. In this particular case, an average relative error of 0.01% was obtained for the real part and a 0.14% for the imaginary part.

## 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 (as it is the case of the diatom $Thalassiosira\ pseudonana$). In order to find which is the most accurate model for the characterization of such complex shapes, an example considering 100 prolate cylinders per mm$^3$ with a diameter-to-length ratio equivalent to 0.8, the PSD of Fig. 14 (showing the radius of an equivalent volume sphere with a slope parameter $\xi = 3$, effective radius $r_{eff} = 3.2\ \mu m$ and effective variance $v_{eff} = 0.005$ resulting in $R_{min} = 0.8\ \mu m$ to $R_{max} = 3.6\ \mu m$), and the assumed




(a) (b)

**Figure 13.** (a) Assumed and estimated real part of the refractive indices for the inner and outer layers using the Bernard model combined with the genetic algorithm. (b) Assumed and estimated imaginary part of the refractive indices for the inner and outer layers using the Bernard model combined with the genetic algorithm. Note that in both cases, assumed and estimated values are on top of each other.

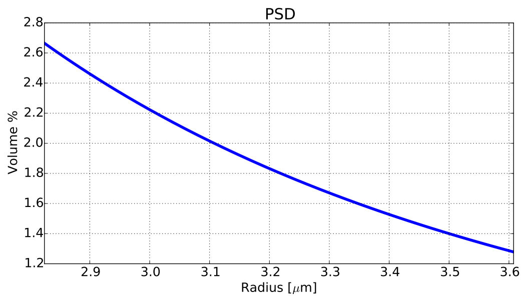
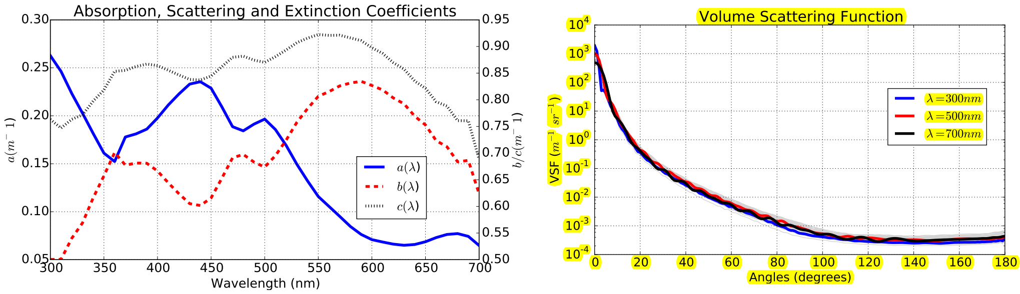



**Figure 14.** PSD of the cylindrical-shaped example.

refractive index of Fig. 4b was simulated using the $T$-Matrix algorithm (Mischenko et al., 1996; Mischenko and Travis, 1998) as a Forward Model. To this end, the code from M. Mischenko Mischenko and Travis (1998) for $T$-Matrix computations on randomly oriented, rotationally symmetric scatterers (cylinders, spheroids and Chebyshev particles) was used. The PSD presents a small effective variance for convergence limitations of the code. The assumed $a(\lambda)$, $b(\lambda)$ and $c(\lambda)$ coefficients are shown in Fig. 15a, and the volume scattering function at each wavelength is shown in Fig. 15b.

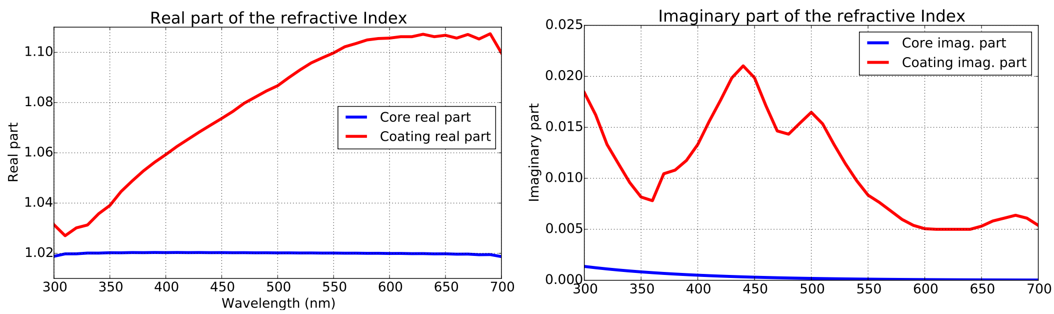
#### 4.3.1 The Bernard Model combined with Genetic Algorithm

Even though these simulated particles are not an exact copy of an actual phytoplankton (for the reasons commented before), the coated sphere model is used here to model the cylindrical shape to analyze their differences. As in previous examples, the assumed value was $V_V = 30\%$ (an averaged value between that assumed by Bernard et al., 2009, and previous works). Figure 16a shows the estimated real part of the inner and outer layers and Figure 16b shows the estimated imaginary parts. In this case, they cannot be compared with the assumed individual refractive index of the cylindrical




(a) (b)

**Figure 15.** (a) Absorption ($a$), scattering ($b$) and extinction ($c$) coefficients of the cylindrical-shaped example. (b) Volume scattering function of the cylindrical-shaped example.




(a) (b)

**Figure 16.** (a) Inner and outer real part of the refractive indices using the Bernard Model combined with Genetic Algorithm in the cylindrical example. (b) Inner and outer imaginary part of the refractive indices using the Bernard Model combined with Genetic Algorithm in the cylindrical example.

particle. Instead, the IOPs obtained from the estimated refractive indices need to be computed using the forward model to analyze if this model is useful to emulate homogeneous cylinders. The volume scattering function, obtained by means of the estimated complex refractive indices and the PSD of Fig. 14 in a forward model, i.e., the $T$-Matrix, is shown in Fig. 17. The committed error in this last figure is noticeable even to the naked eye, especially at longer wavelengths, achieving 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

The genetic algorithm can be combined with the $T$-matrix code in order to consider cylindrical shapes when estimating the inner complex refractive index. However, one simulation of cylindrical shape particles with such dimensions, using the Mischenko code, needs about 67 minutes in a

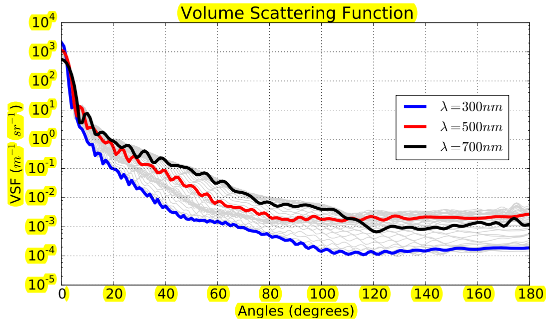
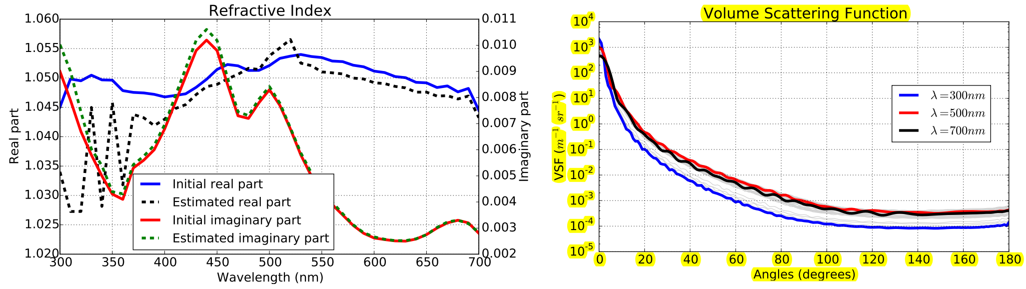



**Figure 17.** Volume scattering function obtained using the Bernard Model combined with Genetic Algorithm in the cylindrical example.

computer with an i7 at 3.20 GHz and running Windows 8.1. This prevents the use of the genetic algorithm in such circumstances, since it needs to execute this simulation several hundreds of times at each wavelength in order to accurately estimate the complex refractive index. That means that several months would be required to estimate the whole refractive index spectra, even using distributed processing. To avoid that, some kind of approximations must be considered. In order to perform fast estimations, equal-volume homogeneous particles with spherical shape are considered instead of the cylinders. This allows using the Lorenz-Mie theory instead the $T$-matrix approach, dramatically improving the simulation time. Then, the estimated refractive index using homogeneous spheres is finally applied on homogeneous cylinders to obtain their IOP, since the volume scattering function values are case sensitive to the particle shape. Although the slow $T$-matrix approach is needed for this simulation, it has to be executed only once. For sure, using better computing resources (as for instance, by means of a computer cluster), this problem disappears and the genetic algorithm can be used with its complete potential.

The methodology was applied on this last example using the same PSD of Fig. 12. The estimated complex refractive index is shown in Fig. 18a. The averaged relative error of the real part is 7.75% and 2.61% for the imaginary part. Since absorption is proportional to the volume, the inverted imaginary part of the refractive index agree well with the assumed values (volume equivalent spheres are being used). However, since scattering depends strongly on the shape of particles, the inverted real part of the refractive index deviate from the assumed values. The major differences are obtained at the lowest wavelengths, which is also noticeable in the volume scattering function, as seen in Fig. 18b, with some artefacts in those wavelengths where abrupt changes of the real part of the refractive index occur (330 and 350 $nm$). However, the averaged relative error committed decreases from 77% in the previous method to 16%. If homogeneous spheres are used instead of cylinders to obtain the IOP, the averaged relative error increases to 22%, which demonstrates that choosing a suitable shape improves the results.




(a) (b)

**Figure 18.** (a) Assumed and estimated refractive indices using the genetic algorithm for spherical-shaped homogeneous particles, and (b) the volume scattering function.

**Table 2.** Averaged relative errors committed in each method

| **Shapes** | **Model** | **$n$ relative error** | **$k$ relative error** | **VSF relative error** |
|---|---|---|---|---|
| Homogeneous sphere | Twardowski model | 42% | 44% | 68% |
| | Stramski model | 8.2% | 15% | 0.17% |
| | Genetic Algorithm | 0.004% | 0.24% | 0.17% |
| Coated sphere | Genetic Algorithm | - | - | 78% |
| | Bernard model | 1.4% | 51% | 52% |
| | Bernard model & GA | 0.1% | 0.14% | 0.2% |
| Homogeneous cylinder | Bernard model & GA | - | - | 77% |
| | Genetic Algorithm[a] | 7.75% | 2.61% | 16% [b] |

[a]The refractive index is estimated using spheres but the IOP is obtained using that refractive index in cylinders.
[b]If the cylindrical shape is not used, the error rises up to 22%.

## 5 Discussion

Table 2 shows the averaged relative errors associated with each method when estimating the real and imaginary parts of the refractive indices and the one committed by the respective volume scattering functions in the three examples of the previous section. In the real part case, the error was obtained using $n - 1$ instead of $n$. Note that the inverse models do not compute the volume scattering function. It is obtained after introducing the estimated complex refractive indices in the suitable forward model, i.e., the Lorenz-Mie or $T$-Matrix theories.

In the homogeneous sphere example, the Twardowski model presents the higher errors, especially when comparing the volume scattering function. It can also be seen in the table that, although the errors of the Stramski model are considerably higher than the ones of the Genetic Algorithm, especially in the imaginary part estimation, similar estimations of the volume scattering function are recovered in both cases. This implies that there is no need of an accurate refractive index estimation

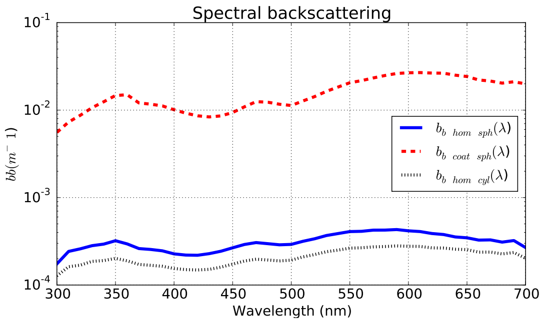



**Figure 19.** Spectral backscattering of the three test cases: homogeneous sphere, coated sphere and homogeneous cylinder.

in this particular example to obtain a suitable characterization of the scattering properties. However, the Genetic Algorithm performs with an 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. Besides, this method presents an important disagreement when computing the volume scattering function. This result shows that, in case the optical behaviour of coated spheres were closer to that of actual phytoplankton particles, as stated in (Bernard et al., 2009), homogeneous spheres would not be a suitable choice to accurately reproduce their optical behaviour. The Bernard model is a fast technique to estimate the inner and outer refractive indices, but mainly fails in estimating the imaginary part of the refractive index (with an error up to $51\%$). This leads to a significant error committed by the forward model when computing the volume scattering function. However, if the Bernard model is combined with the Genetic Algorithm (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, later, for the volume scattering function in the forward model.

Finally, in the homogeneous cylinder case, it can be seen that the optical properties of this kind of particles are not accurately reproduced using a coated sphere the refractive indices of which are obtained with the combination of the Bernard model and the Genetic Algorithm. From the previous results, it could be expected that the optimal retrieval method would be the Genetic Algorithm using cylindrical shapes to obtain an accurate estimation. However, this involves using the slow $T$-Matrix code of Mischenko iteratively, which would require several months to converge (as the particle becomes more aspherical, the convergence time increases considerably). In order to make the retrieval faster, homogeneous spheres with equal volume are used instead of cylinders. The retrieved refractive index is then used to obtain the IOPs using cylinders this time. Using this method, the volume scattering function shows an averaged relative error of $16\%$, improving the result obtained using spheres (22%). Therefore, this result confirms that selecting a suitable shape is important for an improvement of the modelling (at least in this ideal case).

To conclude, the results presented in Table 2 do not determine which is the best method to estimate the phytoplankton optical properties, since none of them are a realistic representation of real algae where there may be cell walls, chloroplasts, vacuole, nucleus and other internal organelles, each with its own optical properties. However, the assumed particles serve as a first approximation of actual phytoplankton and are useful to extract some preliminary conclusions and to introduce several improvements as an attempt to make the approximations a bit closer to the reality. Most of the methods shown in the paper are used for the retrieval of the refractive indices of isolated particles or bulk oceanic distributions, and a comparison of their performance can only be done using well-known models. As it has been shown on each example, the Genetic Algorithm is a versatile technique that alone or combined with other methods improve the accuracy of the estimations. However, it is not a fast technique, and several minutes are required for each estimation (when using spherical shapes; slower with aspherical particles) as compared to the few seconds required by the other methods, the Twardowski model being the faster of them.

Further work must be done in order to study their performance when using the optical properties of actual phytoplankton species and bulk oceanic measurements. New shapes may be required to improve the results, as for instance coated cylinders to model algae with a cylindrical shape (as stated by Bernard et al.,2009, in the spherical case, coated particles generate backscattering functions closer to those produced by actual phytoplankton particles) or other outlines more similar to the actual shape of the particle (the $T$-Matrix approach allows the computation of particle shapes exhibiting axial symmetry, Sun et al.,2016). Besides, all these approaches can also have further applicability with other type of optical measurements, as for instance with remote-sensing reflectance. As an example, Fig. 19 shows the spectral backscattering of the three test cases. Many operational remote-sensing inversion models for IOPs use an implicit or explicit assumption about the refractive index, and, in combination with these methods, could be severely improved. Retrieving the index of refraction from space would improve the ability to distinguish sources of backscattering from each other in the ocean. To this end, a much more complex inversion scheme should be developed. There is, however, one important disadvantage shared by all the methods described in this paper, and it is their strong dependence on the accuracy of the measurements. Attenuation and scattering coefficients are needed as inputs for all of the retrieval methods, and if they are not accurate, the retrieved refractive indices will not be as well. As stated by Ramírez-Pérez et al. (2015), the acceptance angle of the optical instruments affect severely on the amplitude of the measurements. By comparing the extinction coefficient of two different instruments with different acceptance angles, different magnitude values were obtained, showing an averaged ratio of 0.67. This is a key issue that must be considered and dealt in order to improve the fidelity of the whole methodology.

## 6 Conclusions

A performance analysis was carried out in order to examine the accuracy of different inverse methods that use the optical properties of small scatterers and their particle size distribution to retrieve their refractive indices. 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. Results indicated that those methods using a genetic algorithm to optimize the inversion were the most accurate ones, but also the slowest. In particular, an excellent agreement between estimated and actual refractive indices and volume scattering functions was obtained for the homogeneous and coated sphere cases, and a fair agreement for the homogeneous cylinders. These results suggest that better characterizations could be obtained for the actual phytoplankton optical properties. Therefore, the next step is a further analysis of the performance of these methods when applied on 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) and the MESTRAL project (CTM2011-30489-C02-01). The authors would also like to show their gratitude to Emmanuel Boss and two anonymous reviewers for their comments and suggestions, which helped to enhance the quality of the manuscript.

## 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.

Andrady, A.: Microplastics in the marine environment, Marine Pollution Bulletin, 62, 1596–1605, 2011.

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.

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.

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.

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.

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.

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.

Cole, M., Lindeque, P., Halsband, C., and Galloway, T.: Microplastics as contaminants in the marine environment: A review, Marine Pollution Bulletin, 62, 2588–2597, 2011.

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.

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.

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 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.

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.

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.

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

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

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

Wiscombe, W. and Grams, G.: The backscatered 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.