Research Article

Vol. 25, No. 21 | 16 Oct 2017 | OPTICS EXPRESS 25486

Optics EXPRESS

Simplified model of spectral absorption by non-algal particles and dissolved organic materials in aquatic environments

B. B. CAEL[1,2,*] AND EMMANUEL BOSS[3]

[1]*Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*

[2] *Woods Hole Oceanographic Institution, 86 Water Street, Woods Hole, MA 02543, USA*

[3] *School of Marine Sciences, University of Maine, 5706 Aubert Hall, Orono, ME 04473, USA*

snail@mit.edu

http://cael.space

Abstract: Absorption by non-algal particles (NAP, a_d) and colored dissolved organic matter (CDOM, a_g) are frequently modeled by exponential functions of wavelength, either separately or as a sum. We present a new representation of NAP-plus-CDOM absorption a_{dg} based on the stretched exponential function $a_{dg}(\lambda) = A \exp\{-[s(\lambda - \lambda_o)]^\beta\}$, whose parameter β can be considered a measure of optical heterogeneity. A double exponential representation of a_{dg} can be fit extremely well by a stretched exponential for all plausible parameter combinations, despite having one fewer free parameter than a double exponential. Fitting two published compilations of *in situ* a_{dg} data – one at low spectral resolution ($n = 5$, $\lambda = 412$-555 nm) and one at high spectral resolution ($n = 201$, $\lambda = 300$-700 nm) – the stretched exponential outperforms the single exponential, double exponential, and a power law. We thereby conclude that the stretched exponential is the preferred model for a_{dg} absorption in circumstances when NAP and CDOM cannot be separated, such as in remote sensing inversions.

References and links

1. C. S. Roesler, M. J. Perry, and K. L. Carder, "Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters," Limnol. Oceanogr. **34**(8), 1510–1523 (1989).
2. C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters* (Academic, 1994).
3. C. S. Roesler and M. J. Perry, "In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance," J. Geophys. Res. **100** 13279–13294 (1995).
4. S. A. Garver and D. A. Siegel, "Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation 1. Time series from the Sargasso Sea," J. Geophys. Res. **102**, 18607–18625 (1997).
5. Z.P. Lee, K. L. Carder, and R. Arnone, "Deriving inherent optical properties from water color: a multi-band quasi-analytical algorithm for optically deep waters," Appl. Opt. **41**, 5755–5772 (2002).
6. P. J. Werdell, B. A. Franz, S. W. Bailey, G. C. Feldman, E. Boss, V. E. Brando, M. Dowell, T. Hirata, S. J. Lavender, Z. P. Lee, H. Loisel, S. Maritorena, F. Mélin, T. S. Moore, T. J. Smyth, D. Antoine, E. Devred, O. Hembise Fanton d'Andon, and A. Mangin, "Generalized ocean color inversion model for retrieving marine inherent optical properties," Appl. Opt. **52**(10) 2019–2037 (2013).
7. M. S. Twardowski, E. Boss, J. M. Sullivan, and P. L. Donaghay, "Modeling the spectral shape of absorption by chromophoric dissolved organic matter," Marine Chemistry **89**(1), 69–88 (2004).
8. K. S. Shifrin, *Physical Optics of Ocean Water* (Springer Science & Business Media, 1988).
9. R. Kohlrausch, "Theorie des elektrischen Rückstandes in der Leidner Flasche," Annalen der Physik un Chemie **91** 56–82 (1854).
10. D. C. Johnston, "Stretched exponential relaxation arising from a continous sum of exponential decays," Phys. Rev. B **74** 184430 (2006).
11. K. L. Carder, R. G. Steward, G. R. Harvey, and P. B. Ortner, "Marine humic and fulvic acids: Their effects on remote sensing of ocean chlorophyll," Limnol. Oceanogr. **34**(1), 68–81 (1989).
12. M. Babin, D. Stramski, G. M. Ferrari, H. Claustre, A. Bricaud, G. Obolensky, and N. Hoepffner, "Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe," J. Geophys. Res. **108**(C7), 3211 (2003).
13. Z. P. Lee, ed. "Reports of the International Ocean-Colour Coordinating Group, No. 5, Remote sensing of inherent optical properties: fundamentals tests of algorithms, and applications," IOCCG, Dartmouth, NS, Canada (2006).

14. A. Bricaud, M. Bavin, H. Claustre, J. Ras, and F. Tieche, "Light absorption properties and absorption budget of South East Pacific waters," J. Geophys. Res. **115**(C08) 009 (2010).
15. "ac-s In Situ Spectrophotometer Datasheet," Sea-bird Scientific (WET Labs), Philomath, OR, USA (2016).

1. Introduction

The total absorption spectrum $a(\lambda)$ (in units of m^{-1}, with λ denoting wavelength, in units of nm) in aquatic environments is frequently decomposed as the sum of four components (e.g. [1])

$$a(\lambda) = a_w(\lambda) + a_\phi(\lambda) + a_d(\lambda) + a_g(\lambda) \tag{1}$$

where $a_w(\lambda)$ is the absorption by water, $a_\phi(\lambda)$ is the absorption by phytoplankton pigments, $a_d(\lambda)$ is the absorption by non-algal particles (NAP), and $a_g(\lambda)$ is the absorption by colored dissolved organic materials (CDOM). Here we are interested in the last two components: absorption by NAP and CDOM.

$a_d(\lambda)$ and $a_g(\lambda)$ are routinely approximated by exponential functions

$$a_d(\lambda) = A_d(\lambda_o) \exp[-s_d(\lambda - \lambda_o)], \quad a_g(\lambda) = A_g(\lambda_o) \exp[-s_g(\lambda - \lambda_o)] \tag{2}$$

for the purpose of radiative transfer computations [2]; by extension, their sum $a_{dg}(\lambda) = a_d(\lambda) + a_g(\lambda)$ is routinely approximated by a double exponential function (the sum of two exponentials). In applications where it is not possible to separate the signals, such as in semi-analytical inversions of ocean-color remote sensing [3–6], they are typically lumped into a single exponential function due to the similar spectral shapes of $a_d(\lambda)$ and $a_g(\lambda)$. (Other models are also used; cf. Ref. [7], their Table 3.)

The first approach treats NAP and CDOM as distinct and independently varying, while the second treats NAP-plus-CDOM as a single pool. Operationally, CDOM is defined as material which passes through a given filter (most often with pore size 0.2 μm) and NAP is defined as that which does not (other than soluble pigments, which are separated into a_ϕ), forming a continuum of material in terms of size [8]. This suggests it would be useful to model the absorption by NAP and CDOM as a the absorption of a continuum of heterogeneous material.

The objective of this paper is to propose a model for $a_{dg}(\lambda)$ that considers NAP and CDOM as a continuum, and to demonstrate that this model is a superior one for $a_{dg}(\lambda)$ and therefore may be useful in remote sensing signal inversion. We propose modeling $a_{dg}(\lambda)$ with the stretched exponential (a.k.a. Kohlrausch function [9]):

$$a_{dg}(\lambda) = A \exp\{-[s(\lambda - \lambda_o)]^\beta\}, \quad \beta \in [0, 1]. \tag{3}$$

The stretched exponential function can be considered the sum of many simple exponential functions with different exponents, where the wider the distribution of exponents, the smaller the value of β [10]. The stretched exponential reduces to a single exponential when the exponents being summed over are identical, in the limit where $\beta = 1$ (hence β being limited to the range [0,1]; $\beta = 1$ is the limiting case where the distribution of s_p values is a delta function, and $\beta = 0$ is the limiting case of an infinitely wide distribution of s_p values). Therefore, modeling $a_{dg}(\lambda)$ by a stretched exponential function considers NAP and CDOM as a collection of many sub-pools of material (e.g. many size classes), each of which has an exponential absorption spectrum with exponents s_p that may differ.

The parameter β then corresponds to the optical heterogeneity of NAP-plus-CDOM, where smaller values of β indicate more variance between the spectral shapes of absorption by the sub-pools. This implies that β may contain useful biogeochemical information (cf. Ref. [11], their Fig. 3); waters with a more heterogeneous NAP-plus-CDOM pool should have smaller β values. In terms of NAP and CDOM, smaller β values should indicate similar magnitudes and dissimilar spectral shapes of NAP and CDOM absorption.

2. Stretched exponential fit to double exponential

We first demonstrate that for any double exponential curve with plausible values for the parameters (A_d, s_d, A_g, s_g), there exists a stretched exponential curve that closely fits that double exponential. Therefore using the stretched exponential as a model for $a_{dg}(\lambda)$ yields no loss of flexibility in modeling absorption curves, despite having one fewer parameter.

We fit a stretched exponential to a suite of artificial $a_{dg}(\lambda)$ data derived from double exponential curves:

$$a_{i,j,k}(\lambda) = A_n \exp[-s_i(\lambda - \lambda_o)] + A_n R_j \exp[-s_k(\lambda - \lambda_o)] \qquad (4)$$

where s_i and s_k correspond to the exponents for NAP and CDOM respectively, R_j corresponds to the ratio of the amplitudes A_g/A_d, and A_n is a normalization constant set so that the maximum of $a_{i,j,k}(\lambda)$ is equal to 1 m^{-1}. We vary s_i from 0.01 to 0.026 nm^{-1} and s_k from 0.008 to 0.018 nm^{-1}, each by increments of 0.001 nm^{-1} (corresponding to the ranges reported in Ref. [12]), and vary $R_j = A_g(\lambda_o)/A_d(\lambda_o)$ from 1/16 to 16 by factors of two (i.e. ranging from $a_d \gg a_g$ to $a_d \ll a_g$), and generate artificial $a_{i,j,k}(\lambda)$ data for each combination of (s_i, R_j, s_k) at 1 nm resolution from 300-700 nm, totaling $17 \times 11 \times 9 = 1,683$ curves.

We use two metrics to assess the quality of fit of the stretched exponential function to the double exponential function: the coefficient of determination (r^2), and the root-mean-square-error (RMSE; the square root of the sum of squared residuals divided by the degrees of freedom of the regression). The fits (as well as those in Section 3) were performed in MATLAB (R2014b; all code used for all analyses herein will be made available at http://cael.space should this manuscript be accepted) using Nonlinear Least Squares, with the constraint $\beta \in [0, 1]$ according to the definition of the stretched exponential function. We set $\max[a_{i,j,k}(\lambda)] = 1$ in all cases so that the RMSEs are comparable across different parameter combinations. Setting $\max[a_{i,j,k}(\lambda)] = 1$ is immaterial; only the ratio $R_j = A_g(\lambda_o)/A_d(\lambda_o)$ is relevant for the fitting procedure, as the stretched exponential can be rescaled by any factor C by multiplying its amplitude by C. In Equation 4, we use $\lambda_o = 300$ nm, but this value is also immaterial for the same reason; for a single or double exponential, the choice of λ_o is equivalent to a rescaling of the amplitude. However, the stretched exponential is not well-defined for wavelengths $\lambda < \lambda_o$. For simplicity, throughout the manuscript we have therefore chosen λ_o equal to the lowest wavelength present in any given analysis, e.g. $\lambda_o = 412$ nm when analyzing the data from Ref. [13], and $\lambda_o = 300$ nm when analyzing the data from Ref. [14].

Figure 1 shows an example fit of a stretched exponential fit to a double exponential where the means reported in Ref. [12] are used for $s_d = 0.0123$ nm^{-1} and $s_g = 0.0176$ nm^{-1}. The stretched exponential fit ($A = 1.005$ m^{-1}, $s = 0.0165$ nm^{-1}, $\beta = 0.974$) is extremely good, with $r^2 > 0.9999$ and RMSE = 8.2 $\times 10^{-4}$; all deviations are within 0.005 m^{-1}, the largest occurring at the endpoint $\lambda = 300$ nm. Note that 0.005 m^{-1} is typically within the measurement uncertainty of the instruments used to measure absorption such as the WETLabs' ac-s [15].

Table 1 shows summary statistics from fitting the 1,683 $a_{i,j,k}(\lambda)$ curves. In each case, the fit was similar to Fig. 1, e.g. $r^2 \geq 0.9985$. The largest residual was always at the endpoint $\lambda = 300$ nm as in Fig. 1. This demonstrates that for any plausible combination of parameters, a double exponential can be extremely well-fit by a stretched exponential, even though the stretched exponential has one fewer parameter. By extension, any $a_{dg}(\lambda)$ data that can be well-described by a double exponential should also be well-described by a stretched exponential.

3. Fits to *in situ* data

The above statistical exercise illustrates the capacity of the stretched exponential to fit a double exponential, but this means little if it does not translate into improvements in fitting $a_{dg}(\lambda)$ data. We next demonstrate that the stretched exponential is indeed a superior model in fitting $a_{dg}(\lambda)$ data, at both low and high spectral resolution.
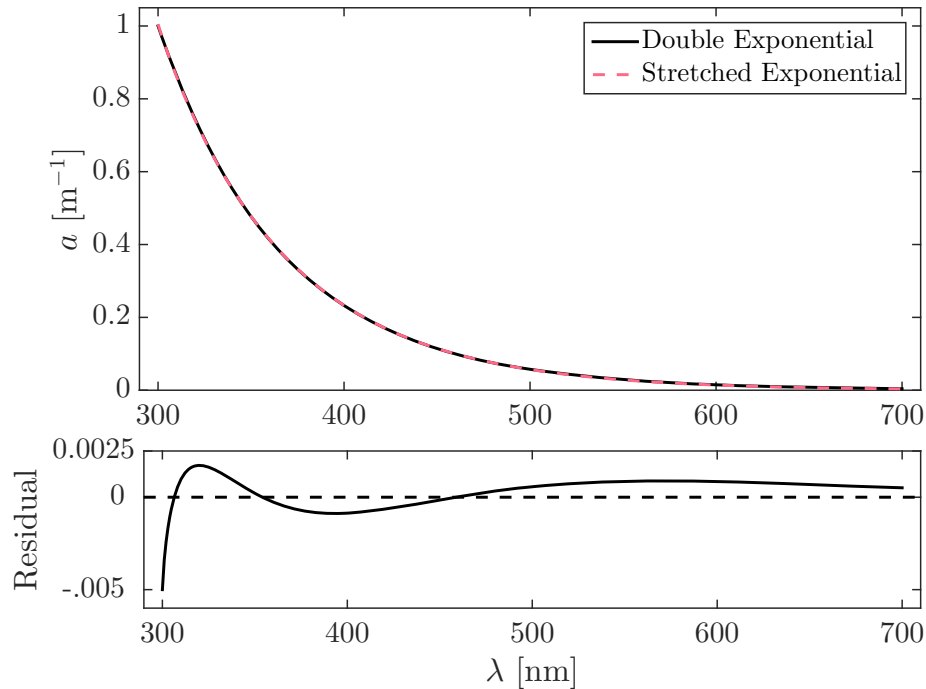
Fig. 1. a) Example of a stretched exponential (dashed line) fit to a double exponential (solid line). b) Residual difference (solid line) between the two curves in Fig. 1(a).

Table 1. Summary statistics for stretched exponential fits to the sum of two exponential functions. Across fits to all parameter combinations, the median, minimum, and maximum r^2 and RMSE are reported, as well as the median, minimum, and maximum of s and β.

Quantity	Median (Range)
r^2	1.0000 (0.9985 − 1)
RMSE	3.3×10^{-4} ($1.4 \times 10^{-5} − 8.0 \times 10^{-3}$)
β	0.986 (0.755 − 1)
s [nm^{-1}]	0.0165 (0.0082 − 0.0510)

We fit a single, stretched, and double exponential to two compilations of *in situ* a_{dg} data. We also fit a power-law function of wavelength, as has been shown to be a better model for multispectral CDOM absorption data [7]. For each $a_{dg}(\lambda)$ curve, we fit all four functions, and compare the stretched exponential to the other three by their RMSE and by their r^2. Note that the RMSE is arguably a better metric by which to compare these models because it incorporates a regression's degrees of freedom, i.e. it suitably penalizes models for having additional fitting parameters. For each data compilation, we say one function outperformed another if it most often had the lower RMSE (or higher r^2).

The first data compilation was developed to inter-compare different remote-sensing inversions [13]. We generate $n = 656$ curves of $a_{dg}(\lambda)$ by adding the measured $a_d(\lambda)$ and $a_g(\lambda)$ data in this compilation. As these data are measured at five wavelengths ($\lambda = 412, 443, 490, 510, 555$ nm), they correspond to the low-resolution limit, where the degrees of freedom are very small;

the number of free parameters in the regression is close to the spectral resolution, and the RMSE will therefore harshly penalize the addition of free parameters.

The second data compilation is from the field program BIOSOPE [14]. We generate $n = 260$ curves of $a_{dg}(\lambda)$ by adding the measured $a_d(\lambda)$ and $a_g(\lambda)$ data in this compilation. As these data are measured every 2 nm from 300-700 nm, i.e. at 201 wavelengths, they correspond to the high-resolution limit, where the degrees of freedom are large, and the RMSE and r^2 are therefore likely to behave similarly.

We found that the stretched exponential outperformed each of the three other functions in both cases. Table 2 shows the number of cases for each data compilation where the stretched exponential had a better (lower) RMSE and better (higher) r^2. The stretched exponential outperformed the single exponential, double exponential, and power law, in terms of both RMSE and r^2, for both datasets. In the low-resolution dataset, in 128 cases the regression found $\beta = 1$ for the stretched exponential, reducing it to the single exponential and thereby making their r^2 and RMSE the same (hence the stretched exponential having a higher r^2 in only 528 cases).

Table 2. Summary statistics for exponential fits to *in situ* data. First column corresponds to data compilation; 'High-resolution' data are those compiled in Ref. [14], and 'Low-resolution' data are those compiled in Ref. [13]. Second column corresponds to which function the stretched exponential is being compared: single exponential, double exponential, and power law. Third column reports the number of cases where the stretched exponential had a higher r^2. Fourth column reports the number of cases where the stretched exponential had a lower RMSE.

a_{dg} data	Stretched vs.	Better r^2	Better RMSE
High-resolution	Single	260	260
($n = 260$)	Double	145	145
	Power	257	257
Low-resolution	Single	528	367
($n = 656$)	Double	610	656
	Power	439	352

4. Discussion

The results considered herein were insensitive to the spectral range considered; we repeated the analysis in Section 2 for the ranges 300-750 nm, 400-700 nm, and 400-750 nm, which yielded no substantive difference. While no ocean color satellite radiometer to date measures below 400 nm, future ones will – NASA's Plankton, Aerosols, Cloud and Ecosystem (PACE) is planned to have UV band extending at least to 350nm – so different spectral ranges are of interest.

Interestingly, the regression estimated different values for β in the two data compilations; see Fig. 2. For the high-resolution data compilation, the β estimates were in the range $\beta \in [.56, .96]$; for the low-resolution data compilation, a majority of β estimates were > 0.96. These data compilations differ not only in terms of their spectral resolution and range, but also in terms of the water types they sample – the high-resolution data are taken from clear ocean waters [14], whereas the low-resolution data are taken from a range of water types. To examine whether spectral resolution and range can explain this difference in the estimated β values, we subsampled the high-resolution data at the same wavelengths as the low-resolution data and repeated the regressions on the subsampled data. The subsampling did not appreciably affect the β estimates, resulting in an average difference of 0.02 and a maximum difference of 0.07 (n.b. the stretched exponential outperformed the other models on the subsampled data). This suggests that these differences in β may result from optical differences between water samples, corrobo-
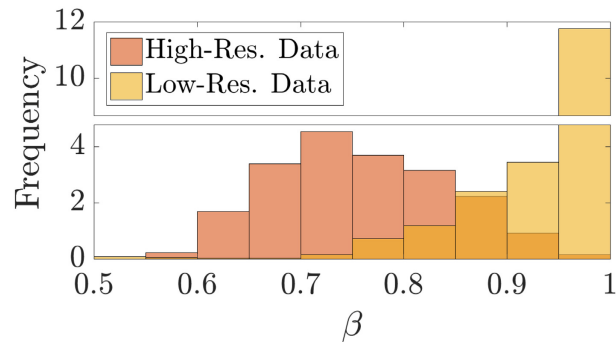
Fig. 2. Histogram of β estimates from the regression to the high-resolution (BIOSOPE) data and to the low-resolution (IOCCG) data.

rating the possibility mentioned in the Introduction that β may contain useful biogeochemical information (ideally when comparing data with good and similar spectral resolution and range). β and s are also anti-correlated throughout all analyses described in this paper, suggesting that a reduced-parameter expression with β as a function of s, or vice versa, could be employed; further research is required to determine the global distributions of β and s, as well as the relationship between these two parameters.

5. Summary

We found, using published *in situ* data compilations of CDOM and NAP absorption spectra, that the stretched exponential function (Equation 3) performed better as an analytical approximation of NAP-plus-CDOM absorption than a single or double exponential or a power law at both high and low spectral resolution. We also found that, for any plausible parameter combination, a double exponential representation of a_{dg} can be very well fit by a stretched exponential. These results favor the stretched exponential model for parameterizing NAP-plus-CDOM absorption, especially in applications where their signals cannot be separated.

We expect that the strongest potential application for this new model is its incorporation into semi-analytical inversions, such as the model of Ref. [6]. We expect it to provide improved inversions, particularly when incorporating data in the UV. We hope our findings here might encourage other researchers to test this model against their own and other community data, to further evaluate its utility in other applications and study regions.

Disclosures

The authors declare that there are no conflicts of interest related to this article.