Modeling biological fluorescence emission spectra using Lorentz line shapes and nonlinear optimization

Paul D. Nation,[1,2] A. Q. Howard,[3,*] and Lincoln J. Webb[1]

[1]Dugway Proving Ground Office, Physics Department, Utah State University, Logan, Utah 84322-4415, USA
[2]Department of Physics and Astronomy, Dartmouth College, Hanover, New Hampshire 03755, USA
[3]West Desert Test Center, Dugway Proving Ground, Dugway, Utah 84022, USA

*Corresponding author: allen.howard@us.army.mil

Received 29 September 2006; revised 6 June 2007; accepted 15 June 2007;
posted 18 June 2007 (Doc. ID 75539); published 20 August 2007

Using the Levenberg–Marquardt nonlinear optimization algorithm and a series of Lorentzian line shapes, the fluorescence emission spectra from BG (*Bacillus globigii*) bacteria can be accurately modeled. This method allows data from both laboratory and field sources to model the return signal from biological aerosols using a typical LIF (lidar induced fluorescence) system. The variables found through this procedure match individual fluorescence components within the biological material and therefore have a physically meaningful interpretation. The use of this method also removes the need to calculate phase angles needed in autoregressive all-pole models.

OCIS codes: 280.3640, 260.2510, 300.2530, 070.6020, 190.1900.

1. Introduction

The ability of many amino acids and coenzymes to fluoresce under stimulation with ultraviolet light has been exploited by many in the biological sciences. In the search and identification of biological aerosols, this property has been exploited in the use of fluorescence lidar systems. The ability of such systems to distinguish biological aerosols from a host of other constituents relies solely on the ability to detect these characteristic markers. It is therefore of great interest to determine whether these signs exist in the measured lidar return signal. Thus, the goal here is to develop a routine that is able to determine the underlying components of fluorescence emission spectra, and their time dependence, from signals gathered by a lidar system. This method can also be used to develop an accurate model for any fluorescence spectrum in question.

2. Theory

In the use of fluorescence lidar systems, the return signal of a biological cloud is measured in terms of received energy. In order to gain information about the emission spectrum of the aerosol, one needs to use the fluorescence lidar equation [1]:

$$E(\lambda, R) = E_L K_0(\lambda) T(R) \xi(R) \frac{A_0}{R^2} N(R) \frac{\sigma^F(\lambda_L, \lambda)}{4\pi} \frac{c\tau_d}{2}. \quad (1)$$

For a given range, R, and laser output wavelength, λ_L, the amount of energy returned depends on the wavelength of interest. Excluding the filter function, $K_0(\lambda)$, the only other wavelength-dependent term is the fluorescence cross section which is a property of the aerosol, or aerosols, being measured. Expanding the fluorescence cross section,

$$\sigma^F(\lambda_L, \lambda) = \sigma^A(\lambda_L) Q^F \mathcal{L}^F(\lambda), \quad (2)$$

gives the wavelength-dependent fluorescence profile function $\mathcal{L}^F(\lambda)$. This is defined as the total fluorescence emission per wavelength interval, $d\lambda$, and therefore is a normalized fluorescence emission spectrum. In addition to acquiring field data, spectral data can be gathered directly in the laboratory. A continuous function modeling the data not only allows for numerical simulations but also gives an in-

0003-6935/07/246192-04$15.00/0

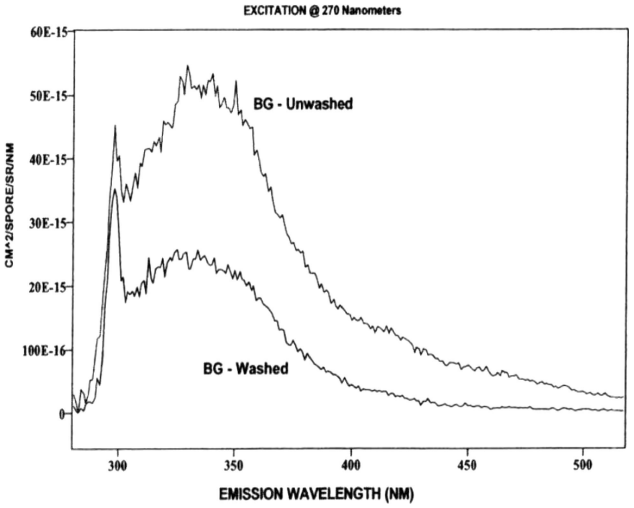

Fig. 1. BG emission spectra at 270 nm in washed and unwashed states.

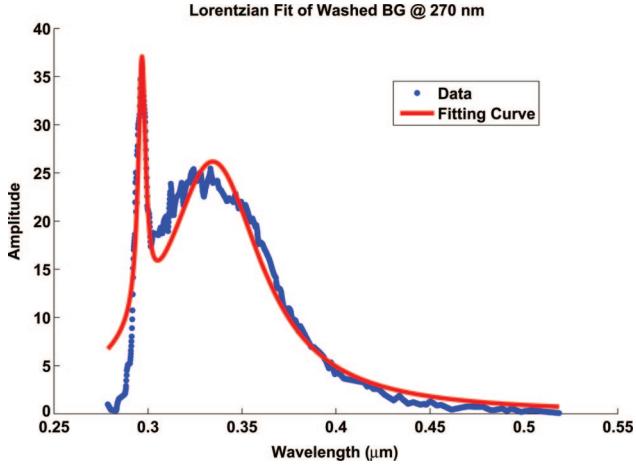

Fig. 3. (Color online) Best fit of washed BG at 270 nm using the Marquardt method.

dication of how individual fluorescence species make up the complete signal. This model of the spectra uses a sum of Lorentzian line shapes:

$$\mathcal{F}(\lambda) = \sum_{i=1}^{n} \frac{A_i}{(\lambda - \lambda_{0i})^2 + \beta_i}. \tag{3}$$

Here, the parameters are the amplitudes, peak wavelengths, and width factors represented by A_i, λ_{0i}, and β_i, respectively. The Lorentz line shape model, or Lorentz approach, to emission is a standard model incorporating natural and collisional broadening effects as well as the limited lifetime of the fluorescing states. These effects are represented in both the amplitude and width terms in the series. Here the extension includes a number of elements that are emissive. As a much used approach, the Lorentz model has been studied in-depth. Our variation however, allows for feedback control and monitoring of parameter changes between measurements, details of which will be discussed later. Our model can also be seen as a reduced Voigt function where the integral is replaced by a discrete summation. We have also ignored the effects of Doppler broadening which accompanies the Voigt model, the validity of which stems from the large mass of our fluorescence samples. Since a large number of line shapes guarantees a good fit, the goal is to accurately model with the fewest number of terms.

Due to the form of the Lorentzian, using optimization techniques requires a nonlinear method. This method uses Levenberg–Marquardt for the fitting algorithm [2], in particular the MATLAB version of the Marquardt method by Hans Bruun Nielsen [3]. The Marquardt method uses unconstrained optimization and therefore will optimize a variable over the entire real line. However, in fitting the emission spectra, the parameters must be non-negative. This is accomplished by following transformations to map each variable from their given interval (a, b) to the real line and back again after optimization is complete:

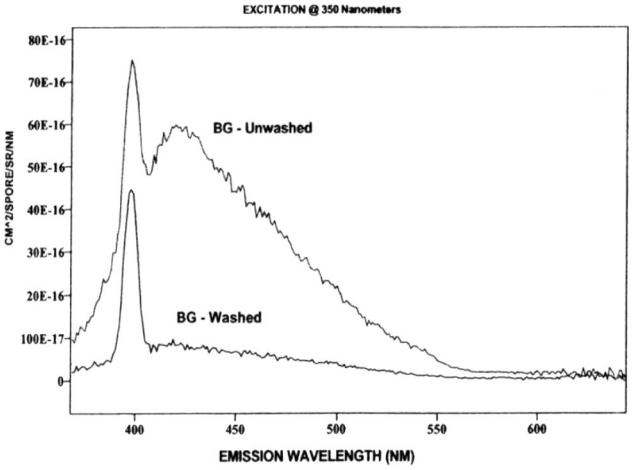

Fig. 2. BG emission spectra at 350 nm in washed and unwashed states.

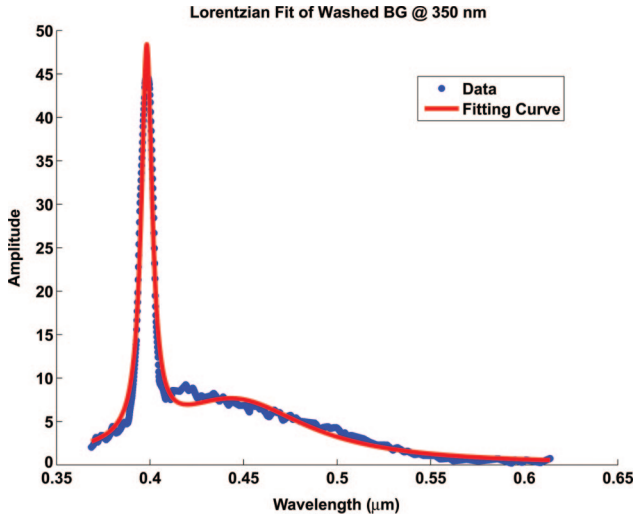

Fig. 4. (Color online) Fitting washed BG at 350 nm.

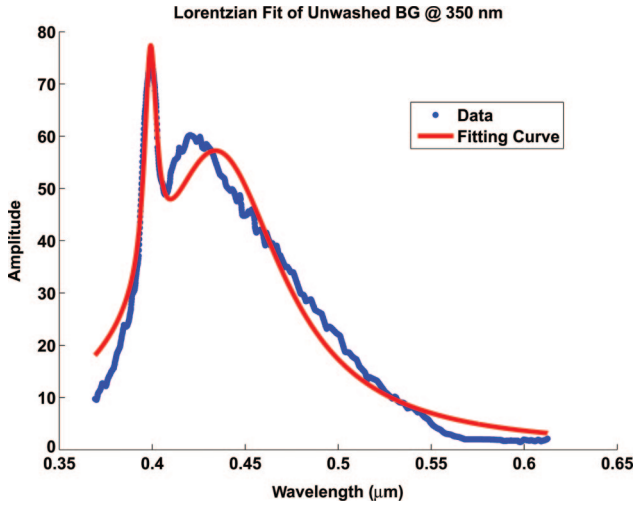

Fig. 5. (Color online) Fitting unwashed BG at 350 nm.

$$g(x)_{(-\infty,\infty)} = \ln\left(\frac{\lambda - a}{b - \lambda}\right), \tag{4}$$

$$f(\lambda)_{(a,b)} = \frac{a + e^x b}{1 + e^x}. \tag{5}$$

Using these two maps, the routine tries to minimize the difference between Eq. (3) and the emission spectra data. As with any other optimization routine, a set of initial conditions must be specified. This approach lets the user graphically choose the peaks in the data and use that information in the initial estimates. As alluded to previously, we can also use feedback methods to give initial values for our fitting function. On repeated measurements, our parameter estimates are fed back into the routine for the new data set under analysis. This removes the need for user intervention on subsequent measurements. This feedback also allows for monitoring the time dependence of the calculated fitting parameters, which is valuable for some lidar measurements but is outside the scope of this paper. In addition to the optimization algorithm used above, the use of Lorentz line shapes also provides for the use of autoregressive (AR), or maximum entropy, methods. As an all-pole model, AR methods should be able to determine the variables in Eq. (3). However, for an AR approach to Lorentzian line shape determination, the underlying time sequence phase needs to be reconstructed. To recover this information one can make use of the complex cepstrum to recover the minimum phase information. While in theory this is straightforward, such a numerical procedure may be more prone to parameter estimate errors.

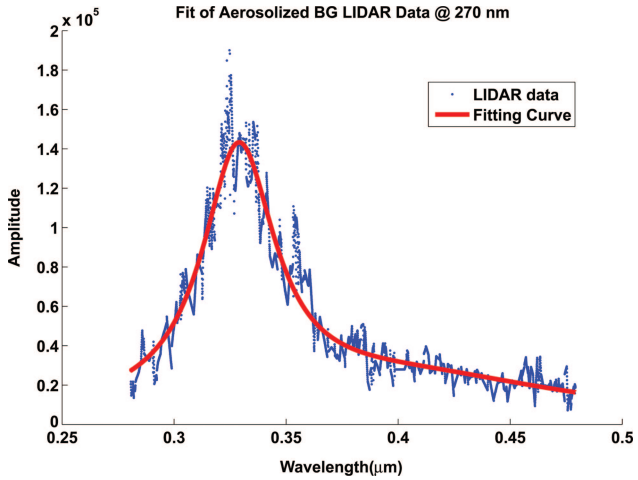

Fig. 6. (Color online) Fitting lidar data of aerosolized BG at 270 nm.

Table 1. Best Fit Variables for Washed BG at 270 nm

Variable	First Term	Second Term
A	1.50×10^{-4}	2.60×10^{-2}
λ_0 (μm)	0.297	0.335
β	5.67×10^{-6}	9.96×10^{-4}

3. Data Fitting

To test the fitting procedure, several data sets of *Bacillus globigii* (BG) were chosen [4]. Included in these data sets are two different laser wavelengths, 270 nm (Fig. 1) and 350 nm (Fig. 2), as well as washed and unwashed preparations.

The data sets from these graphs are obtained through PLOTDIGITIZER [5], an open-source program designed to extract data from technical graphs. The data set text files are then processed in MATLAB, where an interpolation table is created and the user is asked to graphically estimate the heights and widths of the individual Lorentzian terms. The interpolation table is necessary since data sets created by digitizing graphs tend to be unevenly spaced and correcting this provides a better fit.

4. Results

The graphical output from the routine is given for several of the BG data sets (Figs. 3–5), as well as lidar data gathered from an aerosol cloud of BG (Fig. 6). Note that the lidar data set has unitless amplitude, unlike the other data sets.

To verify that the Lorentz terms are representative of physical phenomena, compare the peak wavelengths, λ_0, to those values given for individual fluorescence materials of which the sample is composed. Comparing the peak wavelengths for the two BG samples at 270 nm shows the calculated values represent the Tyrosine and Tryptophan components of the material; see Table 1, Table 2, and Fig. 7 [6]. The large fluorescent intensities for these two species indeed make up most of the emission spectra tested.

Table 2. Variables for One Term Fit for Aerosol BG Lidar Data

Variable	Term
A	1.10×10^2
λ_0 (μm)	0.331
β	8.43×10^{-4}

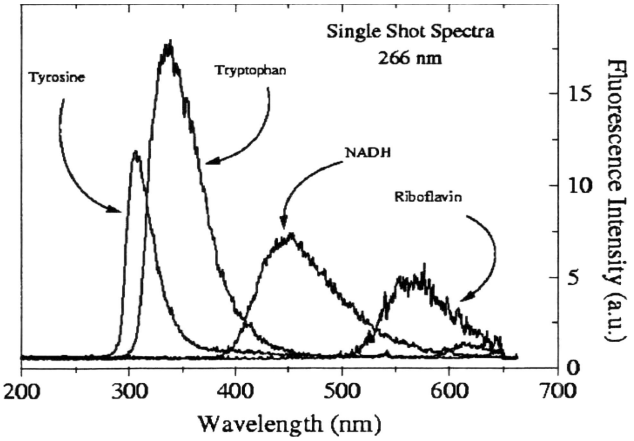

Fig. 7. Single shot emission spectra of various cellular components at 266 nm.

Additionally, one finds the same relationships as those above for BG at 350 nm (Tables 3 and 4). Now, having identified which terms in the series correspond to the fluorescing cellular materials, the amplitudes and widths give additional information on broadening effects and fluorescing lifetimes.

Table 3. Best Fit Variables for Washed BG at 350 nm

Variable	First Term	Second Term
A	5.45×10^{-4}	1.70×10^{-2}
λ_0 (μm)	0.398	0.445
β	1.22×10^{-5}	2.29×10^{-3}

Table 4. Best Fit Variables for Unwashed BG at 350 nm

Variable	First Term	Second Term
A	6.73×10^{-4}	1.06×10^{-1}
λ_0 (μm)	0.399	0.434
β	1.55×10^{-5}	1.88×10^{-3}

5. Conclusions

As has been shown, the optimization approach to calculating emission spectra creates physically accurate models on which to base simulations or other analyses. This approach is an alternative to calculating the phase angles required for AR techniques and relies solely on the Lorentz line shape model. Since this model is used in the AR approach, no additional information is needed for this technique save initial conditions. Indeed the success of this approach relies on good initial values; the modeled spectra tend to fail unless the initial input is close to the actual values. The graphical method of picking points is the only routine found to work in all of the trial cases. The feedback routine automatically provides accurate initial values for all measurements after the initial input. A completely automated version of the same basic method using MATHEMATICA was also investigated. This implementation differs somewhat from the MATLAB version and as of yet does not fully function but completely automates our process. However, even with the current limitation imposed by the need for initial user input, the developed routine provides a robust method by which emission spectra may be quantitatively modeled for use in various scientific investigations.

References

1. R. M. Measures, *Laser Remote Sensing: Fundamentals and Applications* (Krieger, 1984).
2. D. M. Bates and D. G. Watts, *Nonlinear Regression and Its Applications* (Wiley, 1988).
3. H. B. Nielsen, "Immoptibox: A Matlab Toolbox for Optimization and Data Fitting," retrieved 09 May 2005, www2.imm.dtu.dk/~hbn/immoptibox.
4. K. K. Ong and S. D. Christesen, "Biological fluorescence database," in *Proceedings of The Fifth Joint Conference on Standoff Detection for Chemical and Biological Defense* (U.S. Army, 2001).
5. J. A. Huwaldt, "Plotdigitizer," retrieved 28 December 2004, http://plotdigitizer.sourceforge.net.
6. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy* (Plenum, 1999).