Journal of Instrumentation

PAPER • **OPEN ACCESS**

An alternative SNR-based weighted-LSM algorithm to classify and measure the concentration of Biological Agents from Laser-Induced Fluorescence

Jinst

Published by IOP Publishing for Sissa Medialab

Received: *December 15, 2022*
Accepted: *April 3, 2023*
Published: *May 4, 2023*

6th International Conference Frontiers in Diagnostics Technologies
ENEA Frascati Research Centre, Frascati, Italy
19–21 October 2022

An alternative SNR-based weighted-LSM algorithm to classify and measure the concentration of Biological Agents from Laser-Induced Fluorescence

A. Puleio,* L. Martellucci, R. Rossi, N. Rutigliano, I. Wyss and P. Gaudio

Department of Industrial Engineering, University of Rome Tor Vergata, Via del Politecnico, 1, 00133-Rome, Italy

E-mail: alessandro.puleio@uniroma2.it

Abstract: Optical spectroscopic techniques, such as Laser-Induced Breakdown Spectroscopy (LIBS) or Laser-Induced Fluorescence (LIF), have already been used to study and detect Biological Agents (BAs). Unfortunately, BAs usually share similar-shaped emitted spectra and low-signal intensities, making their detection and classification difficult to assess.

Least-Square Minimisation (LSM) based algorithms are usually deployed to measure the concentration of agents from spectra. Recently, it has been shown how the use of *ad hoc* weights can help in improving the performance of the concentration evaluation. More specifically, it has been observed that the "weight matrix" should be modelled as a function of the boundary conditions of the problem.

This work proposes a new weight matrix that is based on the Signal-to-Noise Ratio (SNR) of the measurements. The idea is based on the fact that more noisy data are less reliable and therefore weight should be lowered.

The paper, after a brief introduction and review of the LSM applied to spectra, will show the new methodology. A systematic analysis of the new algorithm is done and the comparison with the other LSM algorithms is presented. The results clearly show that there is a range of parameters for which the new algorithm performs better.

Keywords: Analysis and statistical methods; Data processing methods; Instruments for environmental monitoring, food control and medical use; Optical sensory systems

*Corresponding author.

https://doi.org/10.1088/1748-0221/18/05/C05004

2023 JINST 18 C05004

Contents

1 Introduction

The use of Laser-Induced Fluorescence (LIF) is largely investigated as optical technique to classify the Biological Agents (BAs) in a sample [1–4]. The fluorescence phenomena are electromagnetic emission originated by a molecule through its excitation with a specific wavelength [4, 5].

LIF is a helpful technique to classify and measure the concentration of BAs. The different combination of fluorophores composing each specific biological agent results in different emitted fluorescence spectra [1]. In this way, the collected spectra contain much information about the composition of the biological sample, and therefore the identities of agents contained inside [1].

However, LIF spectra are usually quite similar, their intensities are contained, and therefore noise worsen the quality of classification and concentration measurement. For these reasons, it is necessary to analyse the spectra using advanced and performant algorithms. For example, many researchers developed machine-learning and deep-learning algorithms to improve the classification performances [2, 6]. Unfortunately, artificial intelligence algorithm requires a lot of data and works correctly only under the i.i.d. hypothesis.

Recently, Gabbarini et al. [1] had investigated the possibilities to use the Least-Square Minimisation (LSM) method to classify and measure the BAs in a mixture sample using a reference database. This new method has the advantage that it requires only reference spectra and not mixture ones.

An important variable that must be taken into consideration is the detector noise that may be a function of the wavelength value. For example, the CCD sensitivity at each wavelength can be different with consequent variation in the Signal-to-Noise Ratio (SNR).

In this work, an alternative method to implement the LSM algorithm considering the possible variation of SNR (named W_{SNR}-LSM) is presented. In this way, it can be possible to extract important features for the classification and measurements avoiding the obstacles created by more noising wavelengths.

As in the cited work [1], synthetic spectra to test the performances of the proposed algorithm (W_{SNR}-LSM) in comparison with the other (Classical-LSM and W_{DIF}-LSM) is performed.

2 Methods

2.1 Synthetic spectra generation: noise generation

To perform numerical tests comparable with the cited work of Gabbarini et al. [1] it was decided to generate the synthetic spectra with the same mathematical methods. Also in this case, two reference spectra (wavelength simulated range from 190 nm to 785 nm with a discretization of 1044 values) have been generated for the database. For each test, a synthetic spectrum is generated to simulate the fluorescence signal that can be produced by the mixture sample given the concentrations of two different hypothetical BAs [1]:

$$I_{\text{final}}\left(\lambda_i\right) = \sum_{j=1}^{m} I_j\left(\lambda_i\right) c_j \tag{2.1}$$

where I_{final} is the synthetic spectrum described by the index j for the j-th BA and i to i-th wavelength.

To better test the proposed new implementation of the W-LSM with the Weight values matrix based on the SNR values it was needed to change the method to simulate the noise in the synthetic spectra. In this case, we added a non-homogeneous random noise signal assuming different noise at different wavelengths.

Specifically, it was decided to simulate two values of noise along the spectra, dividing it in two parts, the first affected by a high noise value ($\text{Noise}_{\text{Lv-MAX}}$) and the second affected by a low noise value ($\text{Noise}_{\text{Lv-MIN}}$) (figure 1). To generate the noise values for each wavelength was used the following equation:

$$I_{\text{noisy},j}\left(\lambda_i\right) = I_j\left(\lambda_i\right) + \left(\text{Noise}_{\text{Lv}} * \left(\text{random}\left(-0.5, 0.5\right) + 1\right)\right) \tag{2.2}$$

where Noise_{Lv} corresponded to the value of noise that distinguished the two portions of spectrum in function of the noise amplitude.

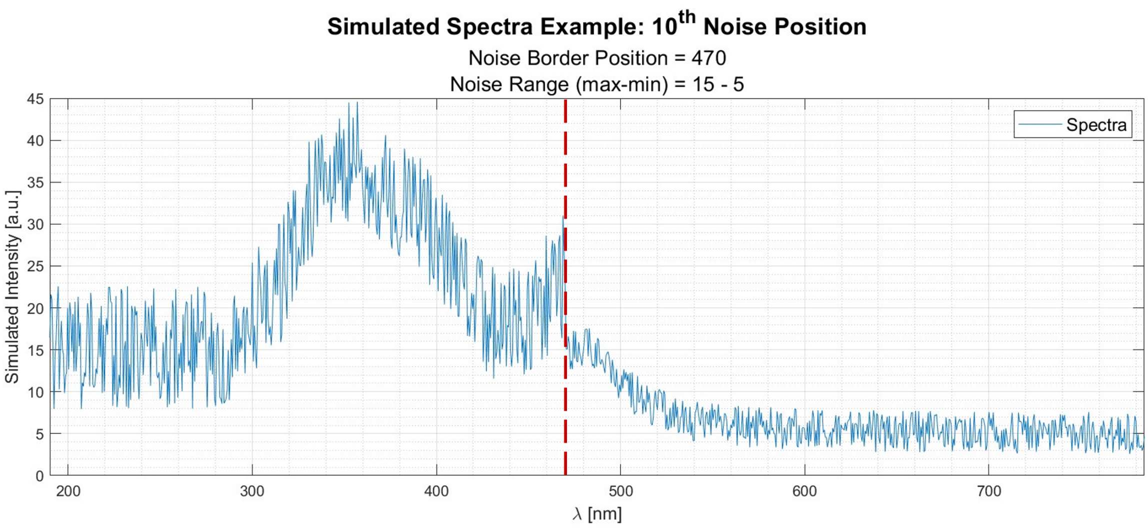

Figure 1. The plot shows one example of the simulated fluorescence spectrum generated through the numerical approach with the particular level of noise variation. Specifically, in the plot is reported the simulated spectra originated by a spectra database with a similitude factor of $f_{\text{sim}} = 66.5\%$.

2.2 W_{SNR}-LSM method: weight matrix based on the Signal-to-Noise Ratio (SNR)

The proposed method has been based on the W-LSM algorithm investigated by Gabbarini et al. [1]. The alternative algorithm proposed in this work has been based on the use of SNR data as weight matrix, to increase the possibility of classification and measurement in case of a non-homogeneous noise level, such as in the real case, that can hide the principal features in the LIF spectra useful for the aim.

In detail, the weight matrix created for the numerical test to use the W_{SNR}-LSM algorithm in the tests has been computed as follow:

$$W_{SNR} = \begin{bmatrix} \frac{1}{\mathrm{Noise}_{\mathrm{Lv}_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\mathrm{Noise}_{\mathrm{Lv}_i}} \end{bmatrix} \tag{2.3}$$

where $\mathrm{Noise}_{\mathrm{Lv}}$ was the variable value of the noise imposed to simulate the noise signal, for n-th acquired wavelengths, that can vary between the maximum imposed value and the minimum value in relation to the position in the synthetic spectrum. The weight matrix has been calculated as shown to calculate the concentrations values through the W-LSM algorithm:

$$c = \left(\Sigma^T W_{SNR} \Sigma\right)^{-1} \left(\Sigma^T W_{SNR} I\right) \tag{2.4}$$

where Σ is the database of LIF reference spectra, W_{SNR} is the weight matrix and I is the acquired LIF spectrum.

2.3 Numerical tests

To study the W_{SNR}-LSM behavior it has been decided to develop a numerical parametric study to compare the proposed algorithm with each one of the algorithms reported in bibliography: C-LSM and W_{DIF}-LSM [1].

Three different databases of reference spectra have been created, composed by a couple of spectra for each database. The 3 databases differed for the similitude factor between the database's spectra (f_{sim} = 33.4%, f_{sim} = 66.5% and f_{sim} = 97.5%) to investigate the influence on the algorithm originated by the similitude between the reference spectra.

Starting with each database, the theoretical spectrum obtained by a mixture sample composed by a concentration of 1 a.u. for the first agent and 2 a.u. for the second agent has been generated.

For each database, it has been created a dataset of 150 spectra, summing the two agent's spectra and adding each one with the noise, generated using the equation (2.2). An example has been shown in figure 1. In addition, to better understand the performance of the proposed W_{SNR}-LSM, a parametric study varying the noise level threshold (or border) has been performed. Thus, it has been decided to locate the border on 20 different positions, starting from 200 nm and varying its position with a 30 nm step.

Thus, the results obtained by the algorithms are an average of 150 simulated measurements for 20 groups, one for each different position of the noise level variation. To complete the parametric study, this approach has been repeated for 4 different ranges of noise ($\mathrm{Noise}_{\mathrm{Max/Min}}$ = 7.5/5; $\mathrm{Noise}_{\mathrm{Max/Min}}$ = 10/5; $\mathrm{Noise}_{\mathrm{Max/Min}}$ = 15/5 and $\mathrm{Noise}_{\mathrm{Max/Min}}$ = 30/5).

The residuals between the methods in comparison have been calculated as the difference between the "measured" concentration and the imposed concentrations [1]. To compare the W_{SNR}-LSM's performances against each other methods has been computed for each test group the Gain factor (G) [1].

2023 JINST 18 C05004

3 Results

Figure 2 shows the comparison between the C-LSM [1] and the W_{SNR}-LSM methodologies. Each row reports the comparison for two spectrum couple (A: Similitude Factor (SF) = 33.4%, B: SF = 66.5%, and C: SF: 97.5%). On the horizontal axis is reported the "noise border" position (in terms of wavelength), while the vertical axis reports the gain value. A gain larger than one means that the W_{SNR}-LSM perform better than C-LSM. The black line shows the difference intensity between the two spectra (defined as the difference of the two spectra). The four coloured lines with markers show the gain values in four different cases of noise ratios (NR).

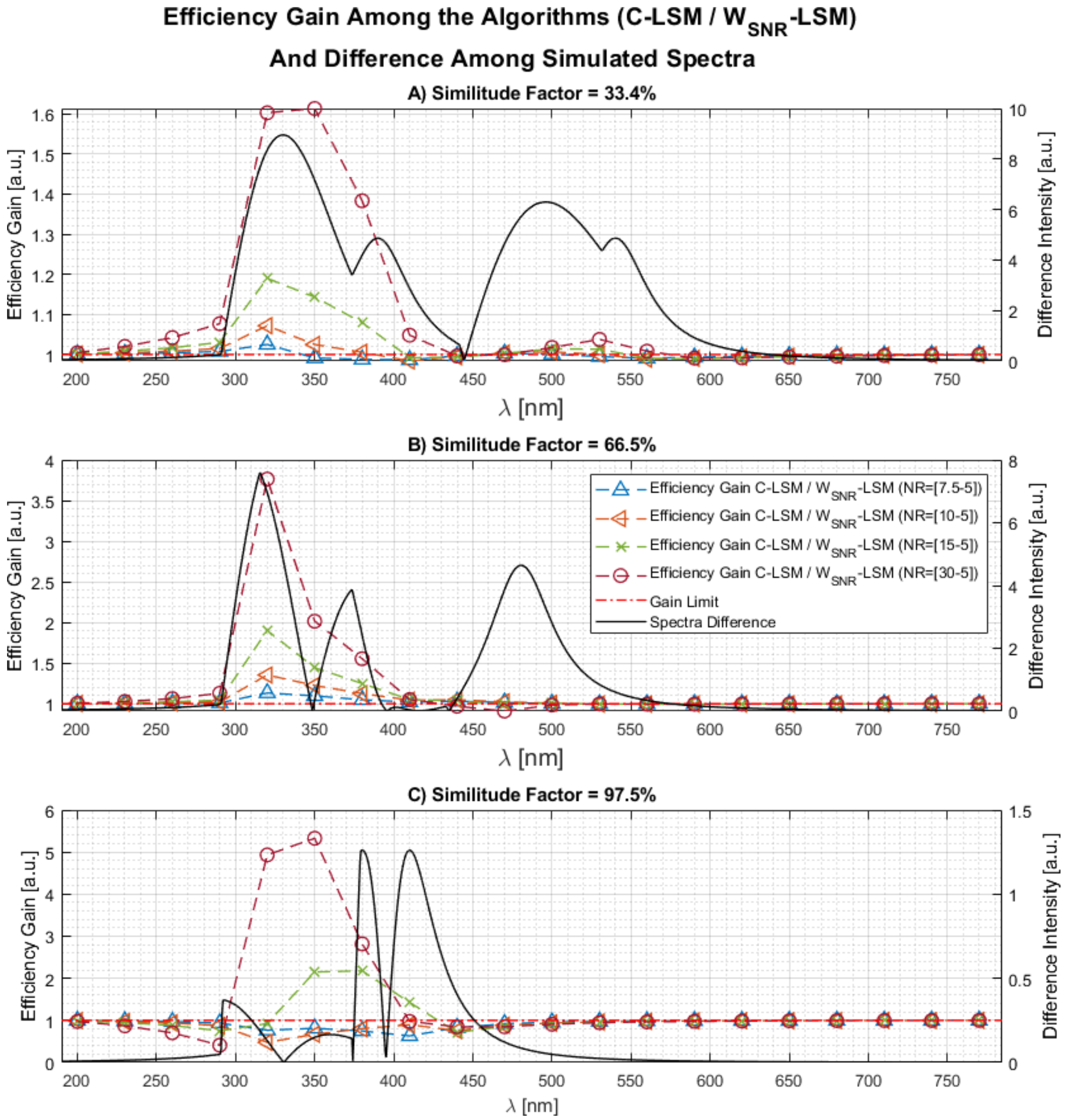

Figure 2. Efficiency Gain (G) resulting from the comparison among the C-LSM and W_{SNR}-LSM algorithms. In each plot are reported the average G values obtained comparing the performances of both algorithms for each noise variation border position and for each Noise Range (NR) Level analysed. In plot A it is shown the results with a database with a f_{sim} = 33.4%, in B the results with a f_{sim} = 66.5%, while in C with a f_{sim} = 97.5%.

2023 JINST 18 C05004

In all three plots (A, B, and C), it can be observed that higher is the NR, higher is the efficiency gain. For example, the red line (NR 30-5) has higher values than the green line (NR 15-5), etc. This result was expected since the W_{SNR}-LSM is thought to alleviate the effect of noise on concentration estimation.

Another interesting and expected result is the trend of the line as a function of the noise border. In fact, high performances are observed in regions before 450 nm, when there are two well separated regions high difference intensity, while once the first high-difference intensity region has been passed, using the W_{SNR}-LSM does not involve relevant gains because the C-LSM has enough high-SNR features to well evaluate the concentration.

The last interesting comment is about the trend of the efficiency gain as a function of the similitude factor. In very different spectra (A), the maximum gain efficiency is around 1.6, while for very similar ones (C), the gain reaches values larger than 5. This was also expected since higher similitude factors means that the correct selection of the features (performed through the Weight matrix) is crucial to have an efficient evaluation of the concentration.

Figure 3, which can be found in the appendix, reports the comparison between W_{DIF}-LSM and W_{SNR}-LSM. What has been described for the previous case mostly applies also for this case, apart from the fact that in for noise border away from the high-difference regions, the choice of the W_{DIF}-LSM better evaluates the concentration respect with the W_{SNR}-LSM. This was expected too, since the noise-based weight matrix, as previously mentioned, is through to allows a better feature selection in regions with high-noise and high-difference.

4 Conclusions

In this work has been explored by a numerical point of view an alternative version about the C-LSM and W_{DIF}-LSM algorithm, both reported in bibliography as possible algorithm to classify and measure the biological agents in a mixture sample through the analysis of LIF spectra. Specifically, the proposed algorithm, named W_{SNR}-LSM, is always a weighted-LSM algorithm that use, to highlight the most important spectra features, a weight matrix based on the Signal-to-Noise Ratio (SNR).

The proposed algorithm has been tested in comparison with the other similar method presented in bibliography [1], using a simulation approach with a series of parametric studies, to preliminary investigate its behaviour.

As expected, the results showed that W_{SNR}-LSM had good performances when the level of noise used in the spectra simulations had high values. Indeed, comparing the C-LSM with the W_{SNR}-LSM it was possible to see, how the W_{SNR}-LSM performs better in the case of larger noise. Moreover, it has been observed that performances increases when there is a large range of wavelength that are noisier than the other ones.

Comparing the W_{DIF}-LSM with the W_{SNR}-LSM it was possible to see how at extreme values of similitude factor the proposed algorithm gives good performances only in case of high level of noise. While for mean values of the similitude factor (66.5%) it has been observed how the results indicate good performance by W_{SNR}-LSM especially when considering an increase in noise level.

From the test numbers performed it was possible to conclude that the alternative W-LSM algorithm looks promising. The possible application of this type of algorithm could be useful in case of LIF measurements with large noise levels.

2023 JINST 18 C05004

The possible future developments certainly require further investigation with the use of laboratory samples in different conditions to validate the real performances of the above algorithms, highlighting the conditions at which one algorithm is more suitable than other ones.

A Results about the comparison between W_{DIF}-LSM and W_{SNR}-LSM

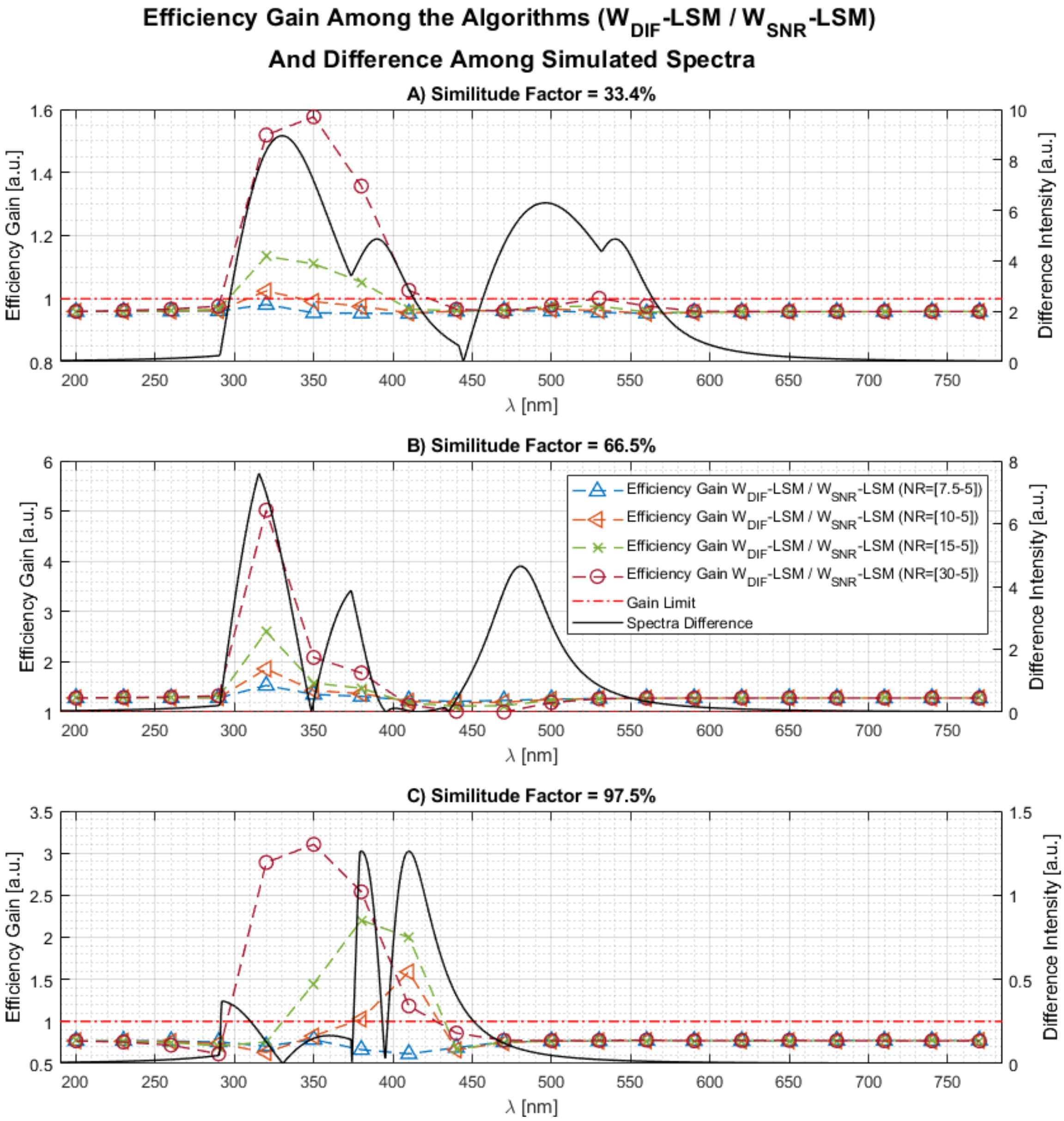

Figure 3. Efficiency Gain (G) resulting from the comparison among the W_{DIF}-LSM and W_{SNR}-LSM algorithms. In each plot are reported the average G values obtained comparing the performances of both algorithms for each noise variation border position and for each Noise Range (NR) Level analysed. In plot A it is shown the results with a database with a f_{sim} = 33.4%, in B the results with a f_{sim} = 66.5%, while in C with a f_{sim} = 97.5%.

2023 JINST 18 C05004

References

[1] V. Gabbarini et al., *A Weighted-LSM Method to Improve Classification and Concentration Evaluation from Laser-Induced Fluorescence Spectra*, *Sensors* **22** (2022) 7721.

[2] M. Kraus, L. Fellner, F. Gebert, P. Carsten, A. Walter and F. Duschek, *Classification of Substances Combining Standoff Laser Induced Fluorescence and Machine Learning*, *J. Light Laser Curr. Trends* **1** (2018) 003.

[3] F. Duschek et al., *Standoff detection and classification of bacteria by multispectral laser-induced fluorescence*, *Adv. Opt. Tech.* **6** (2017) 75.

[4] V. Gabbarini, R. Rossi, J.F. Ciparisse, A. Puleio, A. Malizia and P. Gaudio, *An UltraViolet Laser-Induced Fluorescence (UV-LIF) system to detect, identify and measure the concentration of biological agents in the environment: a preliminary study*, 2019 *JINST* **14** C07009.

[5] J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer (2006) [DOI:10.1007/978-0-387-46312-4].

[6] L. Fellner et al., *Determination of composition of mixed biological samples using laser-induced fluorescence and combined classification/regression models*, *Eur. Phys. J. Plus.* **136** (2021) 1122.