# Rapid Detection and Quantification of Bacteria in Drinking Water Using Raman and Fluorescence Spectroscopies

Thesis Submitted to

The Robert H. Smith Faculty of Agriculture, Food and Environment

Hebrew University of Jerusalem

For the Degree of "Master of Agriculture Science"

By

Amir Nakar

October, 2018

# Rapid Detection and Quantification of Bacteria in Drinking Water Using Raman and Fluorescence Spectroscopies

Thesis Submitted to

The Robert H. Smith Faculty of Agriculture, Food and Environment

Hebrew University of Jerusalem

For the Degree of "Master of Agriculture Science"

By

Amir Nakar

October, 2018

This work was co-supervised by

Dr. Ze'ev Schmilovitch
Institute of Agricultural Engineering,
Department of Sensing, Information and Mechanization Engineering,
Agriculture Research Organization,
The Volcani center, Rishon LeZion

Dr. Mikhail Borisover
Institute of Soil, Water and Environmental Sciences,
Department Soil Chemistry, Plant Nutrition and Microbiology,
Agriculture Research Organization,
The Volcani center, Rishon LeZion

Prof. Shlomo Sela (Saldinger)
Institute of Postharvest and Food Sciences,
Department of Food Quality and Safety,
Agriculture Research Organization,
The Volcani center, Rishon LeZion

**Acknowledgments**

I would like to thank my supervisors: Dr. Ze'ev Schmilovitch, Dr. Mikhail Borisover and Prof. Shlomo Sela for their patient guidance, encouragement and advice. This project included a complex collaboration of researchers from different fields of science, and only through their open-mindedness, curiosity, flexibility and relentlessness we managed to do this research.

I thank Dr. Dalit Vaizel-Ohayon and Mekorot National Water Company for their support of this project, for sampling water, analyzing them and sharing their results.

Also, I would like to thank Dr. Timea Ignat, Emir Haleva and Dr. Giorgi Shtenberg, for many hours trying to figure out how the Raman Effect works in bacteria, or rather, why it does not work.

I thank our lab members for the enjoyable work, counselling and good coffee, to Dr. Yulia Kroupitsky, Lilya Kotliarevski, Dr. Jitendra Keshri, Orian Dayan and Nadia Bukhanovsky. And to my friends next door: Dr. Orr Shapiro, Dr. Ella Kaplan, Sivan Galai and Olubunmi Adejimi for listening to my ideas and helping me form them into real research questions that could later be properly looked into.

Finally, this work could not have been done without the love and support from my wife Atara, who listened patiently, waited when I got home late, and supported me in the highs and lows of doing research.

**Abstract**

Bacterial contamination of drinking water is a considerable concern for public health. Annually, waterborne diseases in developed countries affect thousands of people, and even millions may be affected in developing countries. While water authorities go through many processes to assure the safety of drinking water, the methods for testing water sanitation are too slow, taking between 12-72 hours to produce results and are labor intensive. Some modern molecular and analytical methods, such as DNA hybridization, Enzyme-Linked Immunosorbent Assay (ELISA) and mass spectrometry have been suggested to replace the standard microbiological methods. Yet, these methods are hardly applicable on an everyday basis because they are labor-, money- and time-consuming or not accurate enough. In contrast, Raman and fluorescence spectroscopies might provide a basis for fast and non-expensive detection of bacteria in drinking water.

In this study, low-resolution Raman spectroscopy and fluorescence spectroscopy were used to assess the detection ability of bacteria in a model system and in groundwater intended for drinking. *Escherichia coli, Bacillus subtilis* and *Pseudomonas aeruginosa* were used to assess the detection thresholds of the two methods by measuring the emission spectra using either a Raman spectrometer or a spectrofluorometer. Partial least squares (PLS) regression and partial least squares discriminant analysis (PLS-DA) were used to analyze the complex data. Subsequently, drinking water samples, obtained from groundwater wells in the north of Israel, were examined using fluorescence spectrometry to determine the method's ability to detect real microbial contamination of water.

It was found that low resolution Raman spectroscopy is not sensitive enough, being capable of detecting bacteria only at concentrations of $\geq 10^8$ and $\geq 10^2$ CFU/ml for *E. coli* and *B. subtilis,* respectively. In fluorescence spectroscopy, the sensitivity of detection significantly increased by analyzing the complex 3-dimensional fluorescence contour (excitation-emission matrices, EEM) as compared with using a single excitation-emission wavelength pair. In a controlled experiment, a PLS algorithm applied to the entire EEM obtained in the range of 210-400 nm excitation and 220-450 nm emission was shown to reduce the detection threshold from $10^5$ CFU/ml to as low as 10 CFU/ml.

Since fluorescence spectroscopy showed great potential for detection and quantification of bacteria in water, it was later tested on drinking water. During a year-long survey of groundwater, no significant microbial contamination was found. Only one sample out of 99 was contaminated, containing 2,900 CFU/ml of heterotrophic bacteria (according to heterotrophic plate counts) and 2 CFU/100 ml of *E. coli*. This contamination was easily detected using fluorescence spectroscopy. Water chlorination and filtration reduced the spectral fingerprint completely, indicating the connection between microbial contamination and fluorescence. In addition, both Raman and fluorescence spectroscopy-based classification models accurately differentiated between *E. coli* and *B. subtilis* at high concentrations, indicating differences in spectral fingerprints between species.

To conclude, this is the first demonstration of the capability of fluorescence EEM to straightforwardly detect microorganisms in water at a concentration of 10 CFU/ml. This study illustrates the high potential of fluorescence spectroscopy as a mean for rapid and accurate detection of bacteria in groundwater and drinking water.

**Table of contents**

**List of Abbreviations**

CFU – Colony Forming Units

EEM – Excitation Emission Matrix

HPC – Heterotrophic Plate Counts

OD – Optical Density

PARAFAC analysis– Parallel Factor Analysis

PLS – Partial Least Squares

PLS-DA – Partial Least Squares Discriminant Analysis

RMSE – Root Mean Square Error

SERS – Surface Enhanced Raman Spectroscopy

TLF – Tryptophan-Like Fluorescence

UV – Ultra Violet

VI – Variable Importance

# 1. Introduction

## 1.1 Bacterial contamination of drinking water

In spite of the efforts invested in maintaining drinking water safe, we still face incidence of water contaminations by bacteria, such as *Legionella*, *Salmonella*, *Escherichia* and *Pseudomonas* species that cause gastroenteritis, skin disease and other ailments (Leclerc et al., 2002). In the United States alone, ~40,000 people are hospitalized due to waterborne diseases, resulting in 970,000 dollars of damages annually (Collier et al., 2012). Water contamination may result from contamination of wells, pipelines, home and institutional taps or caused by failures in the sanitation system in bottling and water processing plants.

Drinking-water treatment plants use various methods to remove microbial contamination, including filtration, disinfection by chlorination or ozonation, and reverse osmosis. In the supply chain, water may be re-filtered to eliminate contaminants originating from pipes in the water-distribution system. In spite of all these efforts, contamination of drinking water still occurs (Adam et al., 2017). In cases where contamination is detected, the local authorities would often redirect water distribution for disinfection of until the contamination naturally subside (Israeli Ministry of Health 2013, 2016, 2017; Ashbolt, 2015). It is critical, therefore, to detect contaminations as early as possible to prevent waterborne contaminations reaching consumers.

Due to the risk of water contamination, constant monitoring of the microbial quality of water is required. Timely identification and treatment of contaminations can prevent pathogens from reaching consumers, improving public health and reducing costs. To control microbial contamination, the water industry uses various methods, such as turbidity measurement and culture-based techniques to detect coliforms, fecal coliforms and fecal streptococci. These tests are performed throughout the water treatment process and at various control points along the water supply network. The main drawback of the current methods is that they are time-consuming and may not prevent contamination from reaching consumers. Furthermore, the standard methods provide data regarding indicators microorganisms, such as coliforms, and are not specific to pathogens, such as *Legionella* and *Campylobacter*. Another microbial method is the heterotrophic plate count (HPC), which quantifies the general microbial load of heterotrophic bacteria regardless of their source (Edberg et al., 2000; Rompré et al., 2002).

Common laboratory methods for the detection of bacteria, generally include culturing samples in selective media, under optimal conditions, followed by colony counts (Edberg et al., 2000). Colony counts are an indication of the number of viable bacteria that is measured by the normalized amount of bacterial colonies on an agar plate. This number is usually expressed as colony-forming-units per ml (CFU/ml) of water. The limitation of these methods is that the formation of colonies requires between 12 hours to a week. Turbidity measurements can detect microbial load as high as $10^6$ CFU/ml or more, but cannot distinguish between microorganisms and suspended matter in general. Moreover, current methods require a large team of trained personnel and resource consuming (i.e., requiring growth media, petri dishes etc.) which makes testing expensive (Allen et al., 2004; Rompré et al., 2002). Not only that, but testing is only done on small sample sizes of between 0.1-1 liter out of the actual water volume that pass continuously through the water supply system. The microbial tests are carried out periodically, such that the majority of water is not tested. Since water contamination is unpredictable and sporadic (Cabral, 2010; Frolich et al., 2017), it may be missed and, therefore, undermine public health. Although all these

limitations of the current microbial testing of drinking water are well known, presently no sensitive and affordable alternative exists for controlling high quality drinking water (Rompré et al., 2002; Willemse-Erix et al., 2009, Monzó et al., 2015).

Several modern methods have been suggested to replace the standard water quality methods. Molecular methods, based on specific DNA amplification have been shown to be highly specific to pathogens (Rompré et al., 2002) but they still require a long culturing stage. Nucleic acid microarray or Enzyme-Linked Immunosorbent Assay (ELISA) technology could provide highly specific and fast (within hours) results. However, these methods rely on expensive reagents, instrumentation, highly trained personnel and are thus not applicable to most industrial uses. Mass spectrometry has also been proposed, especially the Matrix-assisted laser desorption/ionization (MALDI-TOF) approach utilizing high energy lasers to ionize samples and analyzing the break-down products using mass spectrometry (Kriegsmann et al., 2018). While this approach requires fewer expensive reagents, it remains an expensive alternative due to the cost of the instruments and need for highly experienced and trained personnel. Further, Mass spectrometry requires culturing (albeit shorter than standard) which takes time and is limited to specific types of bacteria (Stöckel et al., 2015). Therefore, there is a need for a rapid, quantitative and accurate method for the detection of bacteria in drinking water that will be economical and reliable.

## 1.2 Optical Spectroscopy

Optical spectroscopy approaches, such as Raman, infrared and fluorescence spectroscopy have been suggested as alternatives to the standard methods of detection of microbial contamination (Pahlow et al., 2015; Stöckel et al., 2015). Raman scatter and fluorescence emission allow measuring the intensity of the light passed through or emitted by a sample; the intensity depends on the concentration of materials interacting with photons. The spectra, i.e., dependence of the transmittance/emission intensity on the wavelength, provide qualitative and quantitative information on composition of various substances. In order to create a reliable method for detecting bacteria in the food and water industries, a large and robust dataset of spectra must first be collected (Stöckel et al., 2015). Once the dataset is created, a statistical model can be developed and designed for rapid detection of bacteria according to their spectra. The advantages of optical spectroscopy methods are that they require little to no sample preparation, do not use expensive reagents and the instrumentation can be relatively cheap. Furthermore, once a suitable method is designed, no skilled personnel are required and, in many cases, the sample scanning can be done automatically and even on-site, thus saving on operational costs and reducing human error. In this work we endeavor to apply two different spectroscopy approaches, Raman and fluorescence spectroscopies, for the detection of bacteria in water.

### 1.2.1 Raman spectroscopy

#### 1.2.1.1 Raman shift phenomenon

Raman spectroscopy is based on Raman shift, a phenomenon first described by C.V. Raman in 1928 (Bernhard, 1995a). When photons of a specific wavelength interact with a molecule, some of the photons undergo a Raman Shift, where the photons' energy (and wavelength accordingly) are altered. This is because some of the photons wave energy is transformed into vibrational energy in the molecular structure during an absorption-emission reaction (Bernhard, 1995a). The shift depends on specific chemical structure and composition of materials. Consequently, the Raman Effect is specific and a

molecular "fingerprint" can be assembled (Bernhard, 1995b). Raman shift is dependent upon the polarizability of a substance, i.e., the ability of its electron cloud to undergo polarization in the electric field induced by other molecules or externally. Water molecules ($H_2O$) have very low polarizability, and thus a minute Raman signal, making them an excellent background for Raman spectroscopy (Stöckel et al., 2015).

Raman shift is measured using a spectrometer detecting the light's wavelength after it was scattered by a sample and converts it into a digital signal for computer processing (Krafft and Popp, 2015 ; Stöckel et al., 2015). The spectrum displays the intensity of scattered light at each wavelength that is longer than the original transmitted light (which is cutoff from the signal). This light had undergone Raman Shift. In Figure 1, an example of the Raman shift of ethanol is displayed to illustrate how a well resolved Raman spectrum looks (Figure 1). Raman shift is commonly presented by wavenumber in units of $cm^{-1}$.

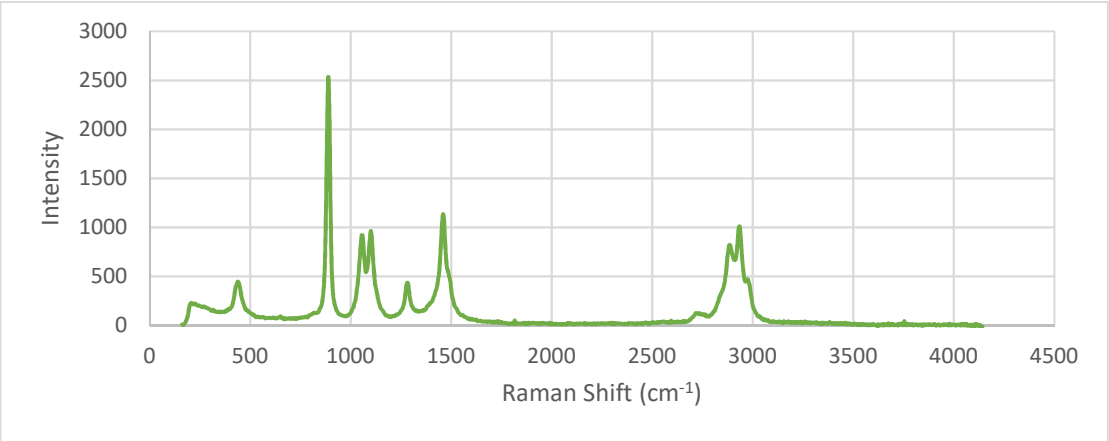


Figure 1 Spectral fingerprint of ethanol. Excitation is by 785 nm laser, for 5 seconds, 3 scans were averaged.

#### 1.2.1.2 Detecting bacteria with Raman spectroscopy

First attempts to describe bacteria and their Raman fingerprints have managed to accurately differentiate different strains of bacteria using long laser exposure times and very high concentrations of bacteria (Fehrmann et al., 1995; Manoharan et al., 1990; Naumann et al., 1995; Williams and Edwards, 1994), such as *Clostridia* spp., *Staphylococcus aureus*, *Escherichia coli*, *Bacillus subtilis*). Later, Zeiri et al. (Zeiri and Efrima, 2005; Zeiri et al., 2004) have shown that most of the spectral fingerprint of bacteria is due to DNA moieties, such as adenine and flavin adenine dinucleotide (FAD). They found that certain phosphorous and carboxylic bonds can also be detected using Raman spectra. In another paper (Premasiri et al., 2017) it was shown that the Raman spectrum of bacteria is also related to guanine, another DNA moiety.

These findings have encouraged research regarding the application of Raman spectroscopy for detection of pathogenic and spoilage causing bacteria in the food industry. Others have demonstrated detection of *Brucella, Escherichia and Yersinia* spp. in milk samples. In that work (Meisel et al., 2012), over 2,000 spectra had been collected, and scanned using a Raman microscope to improve the signal. A machine learning process enabled accurate differentiation between species with >95% accuracy. It is important to note that while the bacteria were grown in milk, which is regarded as a complex background, scanning

was done after washing with water to remove residual milk substances (Meisel et al., 2012). A similar study (Meisel et al., 2014) demonstrated the ability to differentiate meat-associated *Salmonella, Listeria and E. coli* species with >85% specificity. Both studies required overnight culturing to achieve meaningful results and the use of microscopes, which are both expensive and time consuming. Others have also demonstrated the possibility of detecting and differentiating bacteria relevant for the food industry (Boyaci et al., 2015; Li and Church, 2014; Maquelin et al., 2000; Wang et al., 2015; Nicolaou et al., 2011).

A different approach was to quantify bacteria in a sample. The challenge in quantifying bacteria is that the Raman fingerprint of bacteria is very weak (Pahlow et al., 2015; Rösch et al., 2005; Stöckel et al., 2015). This has led to the use of surface enhanced Raman spectroscopy (SERS) techniques. Zhou et al (Zhou et al., 2014) have managed, using gold particles and SERS, to detect *E. coli* cells at less than 100 CFU/ml in a 3-µL sample. The gold particles were added to the sample, causing a localized surface Plasmon resonance reaction, which traps light and increases Raman shift. The same approach was used by Sundaram et al. (2013a; 2013b) to increase the signal of different bacterial strains by a factor of $10^6$-$10^{10}$. While highly attractive, the use of SERS is difficult to scale up to the food and water industries. The high cost of gold and silver required for SERS, the expertise required to construct them and the need to replace them regularly, make this method too expensive for industrial application.

### 1.2.2 Low resolution Raman spectroscopy

While most Raman spectroscopy studies have applied microscopy-based approaches or surface enhancement using gold and silver nanoparticles, few have tried to detect or quantify bacteria using low-resolution Raman. Low-resolution Raman spectroscopy involves obtaining the Raman spectra, without using microscopes nor SERS, but rather with direct application of the laser onto a sample, and spectra collection using a simple Raman spectrometer. Low-resolution Raman tools are relatively inexpensive and easier to operate. These instruments produce a weaker spectrum with lower resolution. In order to overcome the difficulty of low-resolution signals, advanced algorithms are applied to analyze large datasets. These algorithms (such as, PLS, Support Vector Machine and others) clarify the spectra and consequently improve detection limits and resolution, without use of expensive instrumentation or sample preparation. Schmilovitch et al. (2005) detected *Erwinia* and *Clavibacter* species, known plant pathogens, using a simple 785 nm laser and a spectrometer. They managed to detect the bacteria at concentrations as low as 100 cells/ml, using a Partial Least Squares (PLS) chemometric model, with an accuracy >90%. This work was a proof of concept for detection of bacteria using low-resolution Raman spectroscopy that can be applied for the food and water industries. Mizrach et al. (2007) have shown similar sensitivity regarding quantification of yeast. All these findings were obtained using low-resolution Raman. Therefore, low-resolution Raman spectroscopy combined with proper analytical algorithms could be considered as a potential approach to detect microorganisms in the food and water industry.

## 1.3 Fluorescence spectroscopy

### 1.3.1 Fluorescence phenomenon

Fluorescence results from the absorption of a photon by a molecule or ion, which causes electron excitation from the ground state to higher energetic levels (excited states). When excited, the electron may lose some energy due to molecular bond vibrations, collisions and other non-radiative interactions. This causes the electron to reach a lower-energy excited state. Then, the electron may return to the ground state by emitting a photon. However, the emitted photon will have a lower energy as compared with that of the exciting photon due to the loss of energy by internal interaction (Lakowicz, 2006a) (Figure 2).

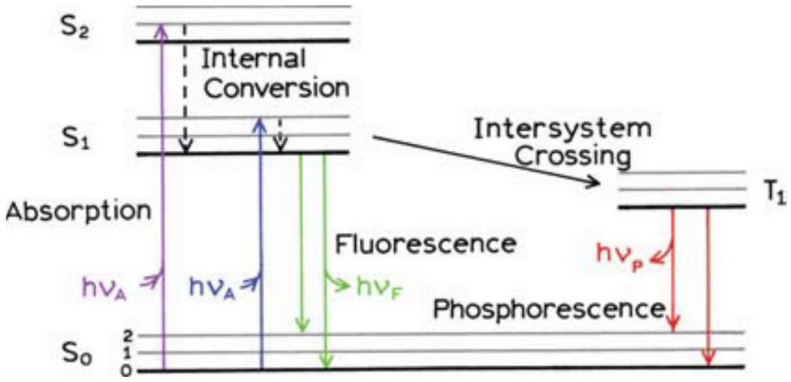


Figure 2. Jablonski diagram illustrating fluorescence (Lakowicz, 2006a).

Examination of fluorescence can be applied for describing the chemical nature of a sample. This approach is widely used for the analysis of soluble organic matter in natural fresh water (Borisover et al., 2009; Ishii and Boyer, 2012), sea water (Stedmon et al., 2007; Yamashita et al., 2011), wastewater (Carstea et al., 2016; Cohen et al., 2014; Yang et al., 2015) and drinking water (Baghoth et al., 2011; Bieroza et al., 2009; Sorensen et al., 2018b). Certain organic compounds, particularly those with aromatic ring structures, will fluoresce when a specific light source is applied to them. This has been widely described in the literature, specifically as a method for the identification and quantification of humic and proteinaceous substances. In proteinaceous substances, the major fluorescence effect is derived from the aromatic amino acids: tyrosine, tryptophan and phenylalanine, whose chemical structure is illustrated in Figure 3. The proteinaceous substances may include free amino acids, proteins (folded or denatured), partially digested proteins, short peptides and even some different organic molecules, which contain the aromatic amino acids, or similar structures, such as indole groups. The fluorescence wavelength of these structures may be shifted and the emission intensity changed depending on specific chemical environment and structures in which the amino acids are organized (Lakowicz, 2006b).

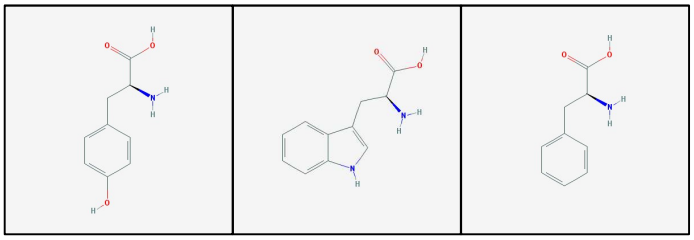

Figure 3. Molecular structures of (from left to right) tyrosine, tryptophan and phenylalanine (Pubchem, 2018).

### 1.3.2 Uses of fluorescence for the detection of bacteria

The presence of proteinaceous substances in water has been suggested as a proxy to the presence of bacteria. Previous studies have shown that bacteria have spectral fingerprints within the proteinaceous substance region of fluorescence, roughly in the range of 200-400 nm excitation and emission (Determann et al., 1998; Elliott et al., 2006). Furthermore, several studies have shown the ability of fluorescence to differentiate between bacterial species, at a high concentration, in colonies (Belal et al., 2011) and in enriched blood samples (Walsh et al., 2013).

This has led to the development of a method to detect bacteria and to quantify their presence in various types of water and specifically in drinking water. The full fluorescence map of *E. coli* suspended in distilled water is shown in Figure 4, where emission intensity was measured in range of excitation and emission wavelengths. By measuring the emission of light at multiple pairs of excitation/emission wavelengths, these data can provide a clear picture of the different organic elements in the sample (Figure 4). However, for measuring proteinaceous-like fluorescence in water, very often, illuminating by monochromatic light at a specific wavelength and measuring emission at a certain wavelength, are used. For example, in order to measure tryptophan fluorescence, Tedetti et al. (2013) used excitation at 280 nm and measured emission at 340 nm.

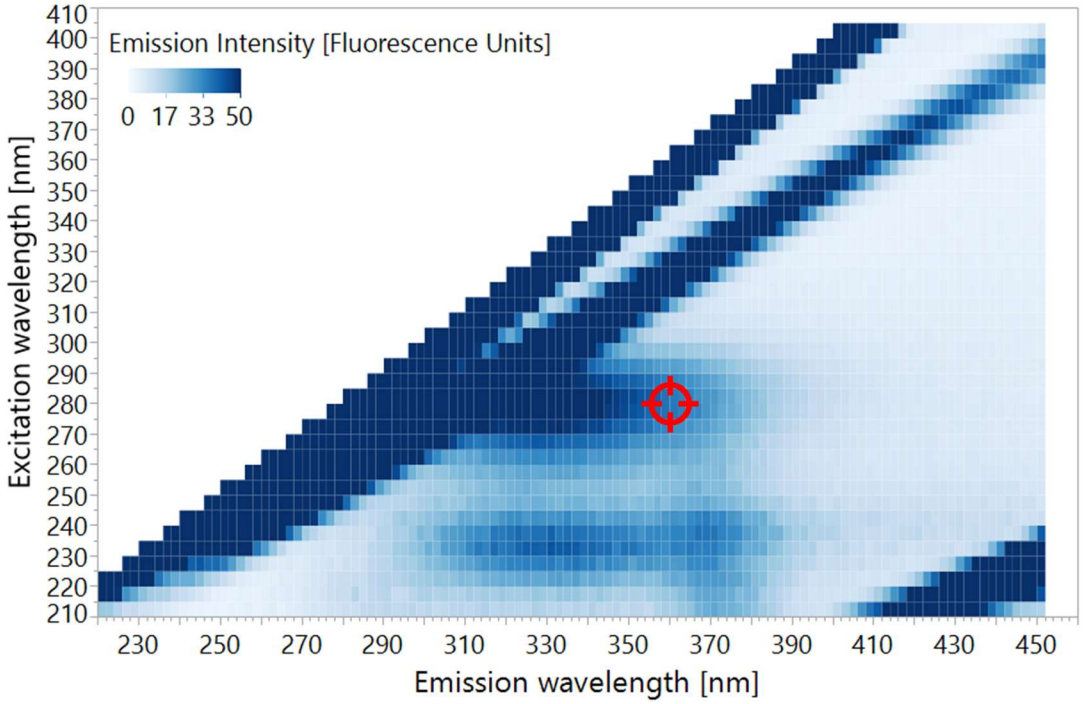


Figure 4. EEM of *E. coli* suspended in distilled water at a concentration of $10^5$ CFU/ml. The intensity has been cutoff at 50 fluorescence units to enhance visibility. The intensity upon excitation by 280 nm and emission by 360 nm is indicated in red.

Sorensen et al. (2015), suggested the application of simple single-wavelength measurements of tryptophan fluorescence for the detection of microbial contamination of water. They used a portable fluorometer with excitation at 280 nm, which measures emission at 360 nm on water samples taken from boreholes and shallow wells in Zambia. They have shown the ability of this simple instrument to detect *E. coli* contamination of >1 CFU/100 ml of water. The emission intensity measured was equivalent to a solution containing ~3.5 ppb of tryptophan. A follow up study by Baker et al. (2015) used a similar instrument to detect *E. coli* contaminations in low-quality catchment water in South African rural areas.

The researchers managed to accurately detect *E. coli* contamination at a concentration of >100 CFU/100 ml with a tryptophan equivalence of 3 ppb (reflecting the emission associated with the whole microbial and non-microbial background at this excitation/emission wavelength pair). Further studies by Sorensen et al. show similar results, displaying an ability to detect low levels of *E. coli* contamination in low quality drinking water in India, Malawi and Zambia, with a tryptophan equivalence ranging between 1.5-3.5 ppb (Cumberland et al., 2012; Sorensen et al., 2018a). These studies suggest that the application of single-wavelength fluorescence spectroscopy, at the tryptophan peak of excitation-emission, e.g., 280 - 360 nm (±10 nm) can be used for the detection of bacteria in low quality water. It is important to note that water with >1 CFU/100 ml of *E. coli* is considered low quality water in most western countries (Allen et al., 2004). It should be also noted that *E. coli* is used as an indicator microorganism for fecal contamination, and the measured fluorescence, does not represent only *E. coli*, but is attributed also to other microorganisms. In low-quality water, *E. coli* may be indicative and proportional to the total number of viable bacteria in water samples and therefore, the measured fluorescence reflected, in fact, the entire microbial population in the water, as well as other tryptophan-containing substances. Nevertheless, in cases of less severe contamination, *E. coli* may not necessarily correlate with the total number of bacteria and therefore fluorescence measurement might not be useful for quantification of *E. coli*.

A similar study was performed in the UK, where the tested water quality was high, and used a laboratory-constructed fluorometer with a LED excitation light source at 280 nm and emission detection at 350 nm. Measurements of both fluorescence and standard microbial indicators, such as *E. coli* and HPC, were carried out throughout the supply chain, at the source, in checkpoints, in the pipeline and at home taps. Good correlation was found between the microbial indicators and the fluorescence intensity at the tryptophan peak (Sorensen et al., 2018b). These findings imply that measuring emission at a single excitation/emission wavelength pair may be sufficiently sensitive to monitor indicative microorganisms in high quality water.

In 2018, Heibati et al. investigated a multispectral approach to link fluorescence emission and presence of bacteria in drinking water in Sweden. While in earlier studies (Baker et al., 2015; Bridgeman et al., 2015; Sorensen et al., 2018b), emission associated with tryptophan was measured at 360 nm upon exciting at 280 nm, Heibati et al. (2018) used the EEM data covering the entire 220-600 nm excitation wavelength range and the 240-800 nm emission wavelength range. The EEMs were analyzed by multivariate statistical methods, firstly Parallel Factor Analysis (PARAFAC) and then Partial Least Squares (PLS). Throughout the measuring period, the water samples did not show substantial microbial contamination, and no *E. coli* or coliforms were detected. The only applicable microbial indicator was HPC, and only weak correlations were observed between the EEM data and HPC. Moreover, the correlations were unusually found to be negative, meaning the more bacteria were in the water, the less intense was the fluorescence effect. The researchers reasoned that perhaps the bacteria are taking up available tryptophan and this is the cause of the negative correlation. This study may serve as an indication that at very low concentrations of heterotrophic bacteria (<150 CFU/ml) it is difficult to detect bacteria using the fluorescence measurement approach (Heibati et al., 2017). Altogether, these studies demonstrated the potential and some limitations of using fluorescence spectroscopy for detection of bacteria in water. It appears that fluorescence emission measurements at a single excitation/emission wavelength pair may be a suitable proxy for microbial contamination in low quality water, however it may not be sensitive

enough to enable bacterial detection in high quality drinking water with low levels of bacteria, as is the case in most developed countries including Israel.

While these studies exemplify the ability of fluorescence spectroscopy to detect bacteria in water, the use of drinking water microbial indicators may be problematic. All studies suggest that the reason bacteria are detectable using fluorescence spectroscopy is the bacterial concentration of tryptophan and its derivatives (primarily proteins). However, tryptophan concentration is affected by all bacterial species (and other organisms) and not only by the specific microbial indicators such as *E. coli*. In order to explain the direct correlation between tryptophan-related fluorescence and bacterial concentration, one may require a laboratory model system which can correctly quantify all bacteria and not only indicative species. Previous works done on this subject (Determann et al. 1998), to the best of our knowledge, did not consider the use of multispectral analysis.

### 1.4 Multivariate data analysis

Because of the complex nature of EEMs and Raman spectra, they are always analysed using multivariate statistics, and specifically chemometric methods. PLS regression or PLS discriminant analysis (PLS-DA) perform well when the number of predictor variables is high and some variables correlate with each other (Heibati et al., 2017; Schmilovitch et al., 2005). Since it is a case when examining EEMs or Raman spectra comprised of signals ($X_n$) measured at hundreds of wavelengths, PLS was a suitable method for predicting bacterial concentration using spectral indicators. PLS algorithms find the correlations between these different variables ($X_n$), and then correlate those to the dependant variable (Y) – microbial concentration measures. Using an iterative process, this enables optimization of the importance or weight given to each variable and a prediction formula can be calculated (Equation 1) (Geladi and Kowalski, 1986).

Equation 1 - PLS prediction formula

$$Y = \sum a_i * X_i + e$$

Y = Predicted Bacterial concentration

$i$ = the wavelength vector index (all wavelengths analysed)

$a_i$ = coefficient at wavelength $i$

X = light intensity at wavelength $i$

e = the calculated error

Chemometric prediction models might have a bias called overfitting, where your model is only relevant to your dataset, but performs poorly when tested on new data (Leinweber 2007). In order to overcome this, the data should be split randomly into calibration and validation sets. The model is trained first on the calibration or training set, and is then tested for accuracy and validity on the validation set for several iterations. This approach is called cross-validation and enables calibration of robust prediction models (Geladi and Kowalski, 1986; Gholizadeh et al., 2015).

## 1.5 Research Problem

The need for a fast, accurate method to enumerate bacteria in water is of high priority for public health (Collier et al., 2012). To serve this goal, spectroscopy-based methods have been suggested (Stockel et al. 2015, Krafft and Popp 2015, Determan et al. 1998). However, different studies have shown conflicting data regarding minimal detectable bacterial concentrations. Fluorescence spectroscopy performs in variable real-life scenarios in which a wide variety of microorganisms are present and most of them are unculturable, whereas all the microorganisms contribute to the spectral fingerprints. Towards these scientific questions, this study examined two spectroscopy methods: low-resolution Raman spectroscopy and fluorescence spectroscopy.

## 2 Research objectives

The aims of this study are to:

a) Test the application of low-resolution Raman spectroscopy combined with PLS for detecting bacteria in water.
b) Compare the ability of fluorescence spectroscopy to detect bacteria in water, using single excitation emission wavelength pair or full EEM analysis.
c) Assess the ability of low-resolution Raman and fluorescence spectroscopy for detecting bacteria in drinking water samples
d) Evaluate the ability of both low-resolution Raman and fluorescence spectroscopy to differentiate between species.

# 3 Materials and methods

## 3.1 Chemical and materials

The materials used in this study are listed in Table 1.

Table 1. Material used in the study

| Materials and chemicals | Manufacturer | Origin (City, Country) |
|---|---|---|
| Ethanol (>99%) | Gadot | Netanya, Israel |
| HPLC-grade Water | Biolab / Merck | Jerusalem, Israel / Darmstadt, Germany |
| Bacto Agar | Difco | Sparks, USA |
| LB Broth (Lennox) | Difco | Sparks, USA |
| Glycerol | Biolab | Jerusalem, Israel |
| L-tryptophan (>98%) | Sigma-Aldrich | St. Louis, USA |
| Delicate Task Wipers | Kimberly-Clark | Irving, USA |
| 0.45 µm Durapore Filters | Millex (Merck) | Darmstadt, Germany |

## 3.2 Bacterial Strains

The following bacterial strains were used in this study: *Escherichia coli* DH5α, *Bacillus subtilis* 3610 and *Pseudomonas aeruginosa* PA14, all strains were taken from the laboratory stocks at the Volcani Center.

Bacteria were transferred from pure cultures stored at -20°C (20% glycerol stock) and were grown on LB agar medium by incubating at 37°C overnight. Several well-isolated colonies were transferred to 10 or 15 ml LB broth and incubated at 37°C with shaking overnight to prepare a starter culture. The starter cultures were used in the sample preparation procedures, as described below.

## 3.3 Optical density to bacterial concentration calibration

In order to associate optical density with microbial concentration, bacterial cultures grown in LB were diluted (with LB) to an optical density ($OD_{590}$) of 1.0 using a benchtop photometer (Biochrom, Cambridge, UK). The cultures were then serially diluted up to $1:10^{10}$ and 100µL were spread onto LB agar plates and incubated overnight in 37°C. Each dilution was plated 3 times for replicates. After incubation, colonies were counted and calibration curves were plotted against bacterial concentrations for each bacterium (*E. coli*, *B. subtilis* and *P. aeruginosa*).

### 3.4 Sample preparation for lab-grown bacteria

Fresh bacterial cultures were washed by centrifugation for 10 minutes at 4991 g (Heraeus Primo R, Thermo Scientific, Waltham, MA, USA) and the bacterial pellet was re-suspended in 15 ml of distilled HPLC-grade water. This washing stage was repeated 3 times, and the final pellet was re-suspended in distilled HPLC-grade water to the volume necessary to reach the required bacterial concentration based on the calibration curves. The actual number of CFU was confirmed by plating 10-fold serial dilutions on agar plates.

### 3.5 Preparation of L- tryptophan solution

A stock solution of $10^7$ ppb was prepared by dissolving 1 g of L-tryptophan powder (CAS number 73-22-3) into 100 ml of HPLC-grade water at 80±10°C using an electric stirrer and hot-plate (Freed Electric, Haifa, Israel). The tryptophan stock solution was serially diluted to prepare a series of concentrations: 1, 2, 3, 5 and 10 ppb.

### 3.6 Groundwater samples

#### 3.6.1 Groundwater sampling

Groundwater samples were obtained from 6 drinking water wells in the north of Israel (Figure 5) between July 2017 and June 2018, prior to and after standard chlorination treatment using sodium hypochlorite solution. The list of wells, the frequency of sampling and the number of the samples collected are listed in Table 2. Water was collected in cleaned dark glass bottles and transported cooled (4-8°C) to the ARO laboratory where it was stored for 12-48 hours at 4°C until analyzed.

Groundwater pH was measured using a pH meter (SI Analytics, Weilheim, Germany). All samples were in the range of 6.9-8.7, indicating there is little quenching effect (Reynolds 2003).

Figure 5. Sampled groundwater wells on an Israeli map (map adjusted from Fine *et al.* 2007)

#### 3.6.2 Groundwater analysis

Water samples were tested for the presence of coliforms, fecal coliforms, fecal streptococci, heterotrophic bacterial counts and turbidity in the certified service laboratory of Mekorot (Mekorot central laboratory, Eshkol Site, Hevel Hayarden, Israel). In chlorinated samples, chlorine concentration was also measured.

Table 2. Groundwater sampling: wells, frequency of sampling and the number of samples collected

| Groundwater well (code) | Sampling Frequency | Number of collected samples (prior chlorination) |
|---|---|---|
| Alonei Ha'Bashan 5 (AAI05) | Monthly | 11 |
| Alonei Ha'Bashan 8 (AAI08) | Monthly | 10 |
| Kidmat Tzvi 1 (KDZ) | Intermittently between October 2017 and June 2018<br><br>(22/10/17, 26/12/17, 15/01/18, 20/03/18, 23/04/18, 23/05/18, 18/06/18) | 7 |
| Shimron 7 (SH7) | Monthly | 11 |
| Einan 3 (QP003) | Summer (July-December 2017, March-June 2018): 2 per month<br><br>Winter (January-February 2018): 2 per week | 26 |
| Einan 6 (QP006) | Summer (July-December 2017, March-June 2018): 2 per month<br><br>Winter (January-February 2018): 2 per week | 34 |
| Total number of samples | | 99 |

### 3.7 Low-resolution Raman spectroscopy measurements

A 785 nm excitation laser was used (Ocean Optics laser module I0785MM0350MS, Ocean Optics, Largo, USA), coupled with a spectrometer (Ocean Optics QE65 Pro, Ocean Optics) covering a wavenumber range 160.7-4142.2 $cm^{-1}$ (or the 766 – 1100 nm wavelength range). Readings were done using a stainless-steel fiber optic probe (RamanProbe II 785/12-5, InPhotonics, Norwood, USA) attached to both laser and spectrometer. All data was collected using the OceanView software (version 1.52, Ocean Optics) and exported as CSV files for further analysis. The experimental setting is shown in Figure 6.

Scans were done at 350 mW for 10 seconds, with 3 immediate scans averaged (for a total of 30 seconds exposure time). Prior to scanning, the height of the probe was adjusted to the best resolution of the peaks of 100% ethanol. 'Dark' scans were measured with the laser off and auto-subtracted from the data to remove machine noise. After obtaining 'dark' scans and adjusting probe height, the scanning procedure involved cleaning the loading aluminum cup (Figure 6) with 100% ethanol and appropriate wipers (Kimberly-Clark, Irving, USA) and loading 150µL of aqueous bacterial suspensions onto the cup. The cup was then placed under the probe and brought to the correct height with all external light sealed off. At this stage, the laser was turned on and the scan commence. After each scan the cup was cleaned and

reloaded with a new sample. Each sample was separately loaded 3-5 times (depending on the experiment) and scanned. The washed bacterial samples were kept at room temperature during 1-3 hours and then spectra were obtained.

In order to enhance the bacterial Raman signal, the additional preparation had been made which included (1) boiling diluted bacterial suspensions for 15 min in 100°C in hot plate (Freed Electric, Haifa, Israel), (2) cooling bacteria in ice for an hour (Premasiri et al., 2017).

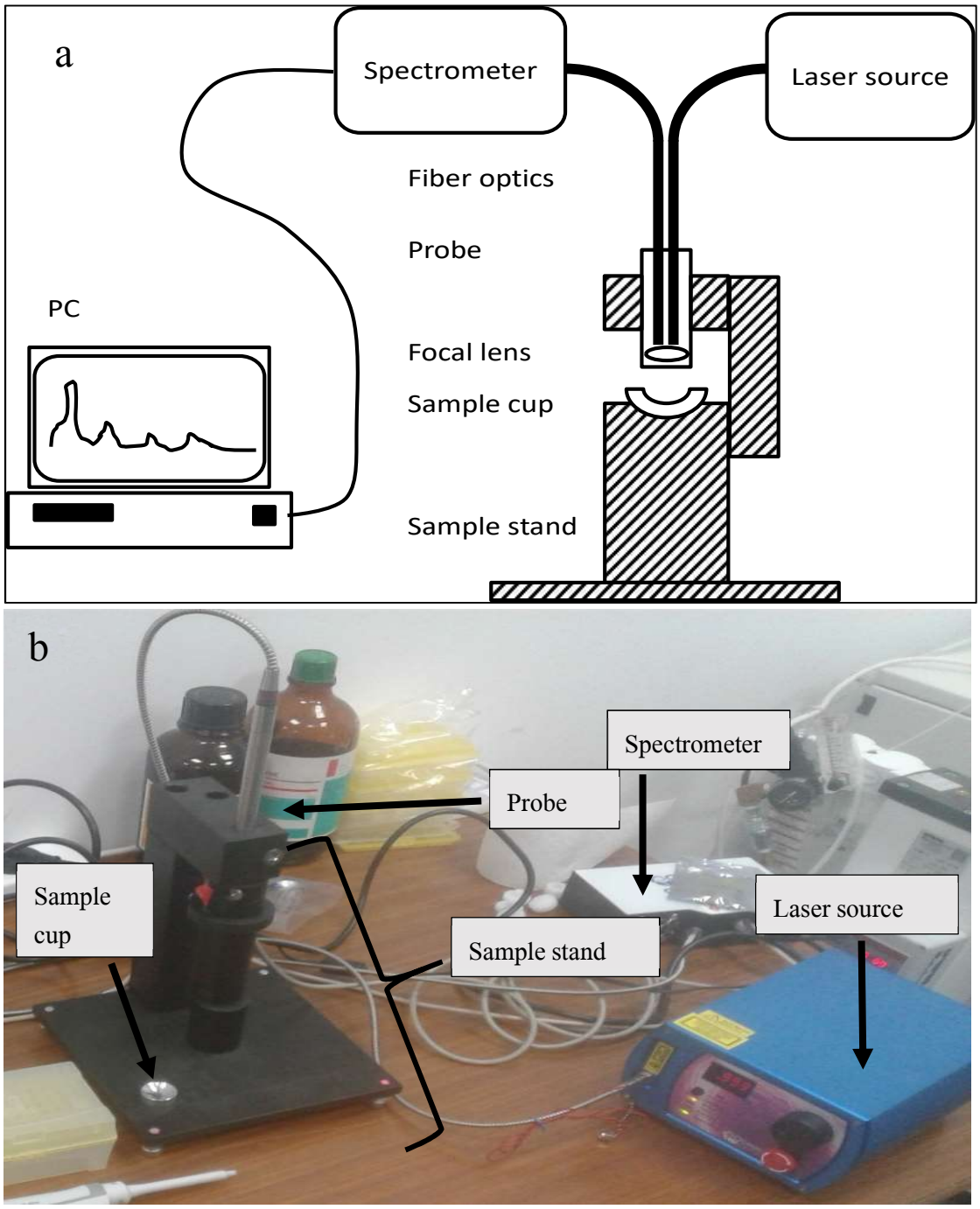


Figure 6. (a) A scheme of the Raman setup including excitation laser, dual fiber optics connected to a probe with focal lens and to a spectrometer, holder for the probe and sample cup. A PC controls scanning and collecting the measured spectra. (b) The Raman setup, including laser source, spectrometer, probe, sample cup and sample stand.

### 3.8 Survival of bacteria during sample preparation, transfer and radiation

In order to verify that irradiation by the 785 nm laser does not compromise cell viability, breaking down signal-producing chemicals, a sample of $10^8$ CFU/ml water of *E. coli* were placed in a sample cup and irradiated by laser for 30 seconds. After that, the samples were serially diluted and plated on LB agar plates. The plates were incubated for 24 hours at 37°C and colonies were counted.

### 3.9 Fluorescence spectroscopy measurements

Fluorescence spectra of water samples were measured with an RF-5301PC spectrofluorometer (Shimadzu, Kyoto, Japan) equipped with 150-W Xenon lamp (Ushio Inc., Tokyo, Japan) as described by Borisover et al. (2009, 2011). Prior to scanning, the water samples were allowed to equilibrate to room-temperature (22±3°C) for ~3 hours. The cuvette used was the standard quartz cell (Hellma, Müllheim, Germany) having a path length of 10 mm and chamber volume of 3,500 µL. Scanning was done at 22±2°C. Fluorescence emission spectra between 220 and 450 nm were collected at 2-nm increments, with excitation wavelengths ranging from 210 to 400 nm at 5-nm increments. Scanning rate per map was about 2.5 minutes and the excitation and emission optical slits were set to 10 nm. Instrument stability was controlled by measuring the intensity of the Raman peak of water at excitation wavelength at 275 nm and emission wavelength at 305 nm. Cuvettes were washed 3 times with HPLC-grade water before the beginning of scanning, and washed by 6 ml of the sample water between each scan. Three ml of sample water were loaded for scanning. The cuvette was always aligned to the same side for scanning. Water samples were scanned as is and after filtration using a 0.45 µm syringe filter (Durapore©, Millex (Merck), Darmstadt, Germany) to remove bacteria and other suspended particles. Prior to filtration, 15 ml of the sample water were used to wash the filter. Each water sample was scanned twice, i.e., 2 independent samples of 3ml).

### 3.10 Statistical analysis

#### 3.10.1 Mathematical sample preparation – preprocessing

Data preparation for the analysis needs preprocessing. We first optimized our protocol to find the best preprocessing procedure by the following procedures:

1. Signals recorded were normalized. Normalization, i.e., a ratio of signal to a set point, minimizes the interferences, which may originate from variation of ambient temperature, lamp intensity and other factors supposed to affect similarly the signal and the reference point. In the Raman experiments, this was done by dividing all intensities in the spectra by the maximum point in the spectra, thus achieving a range of 0-1 for all data points, while retaining the relations between different peaks. In the fluorescence measurements, this was done by dividing the signal to the Raman signal of water. Normalization was done using Excel software (Excel 2016, Microsoft, Redmond, WA, USA).

2. Centering and scaling processes were done automatically using the JMP software (JMP®, Version 13 Pro, SAS Institute Inc., Cary, US) prior to running the PLS analysis. These procedures are commonly used in the PLS analysis and reduce noise to enable better fitting of the model (Geladi and Kowalski 1986).

3. Signal weighing was performed by using the 1/signal and $\log_{10}$(1/signal values. This procedure serves to increase weak signals and uncover more information about a sample (Mizrah 2007). Weighed signals were calculated using Excel software.

4. The 1st derivative of spectra was obtained in order to amplify the trends in the data. It was done by calculating signal $_{Em(n)}$-signal $_{Em(n-1)}$. The first data point in each spectrum was assigned "0", as for this point signal $_{em(n-1)}$ does not exist. The first derivative of spectra was calculated using Excel software.

In order to analyze only fluorescence data, all data-points in the EEMs where the excitation wavelength is longer than the emission wavelength were removed.

The preprocessing method with the highest $R^2$ was chosen for further analysis. In case of equal $R^2$ values, the method with the lowest RMSE was chosen.

### 3.10.2 Partial Least Squares treatment of the collected spectral data

All statistical analysis was done using JMP®, Version 13 Pro (SAS Institute Inc., Cary, US). In this work Partial Least Squares (PLS) NIPALS algorithm (Nonlinear Iterative Partial Least Squares) was used to both quantify and classify bacteria. Classification was done using a Partial Least Squares Discriminant Analysis (PLS-DA) option. Each PLS model consists of several latent variables (LV) or factors. The number of latent variables represents the complexity of the model, a general rule of thumb is that LVs of a good model should be between 3-10 (Geladi and Kowalski, 1986). Variable importance (VI) is a measure of the importance of each variable in the model for the predictive ability. It is calculated by sum of the decrease in error when split by a variable (Wold 2001). For the analysis of EEMs, it was required to reduce the dimensionality of the data from 3-dimentional- to 2-dimentional-martrices, since PLS can only work with 2-dimentional data. For this purpose, each excitation-emission wavelength pair was given an index number and the EEM was transformed into a table of indices (excitation-emission wavelength pair) and emission intensities prior to modelling.

Further, square of regression coefficients ($R^2$); a measure of model correlation with actual values and Root Mean Square Errors (RMSE); a measurement of the model's error rate in the units of measurement were calculated using JMP software. RMSE was calculated according to Equation 2 (Fearn, 2002).

Equation 2 - calculation of Root Mean Square of Error (RMSE)

$$RMSE = \sqrt{\frac{\sum\left(Yi_{real} - Yi_{predicted}\right)^2}{n}}$$

Statistical significance was computed using JMP software. Statistical significance was analyzed by using the Student's t-test and Wilcoxon rank-sum test (also known as the Mann-Whitney test). In order to test whether the variance of two groups is equal, Levene's test was used. Throughout the study, statistical significance was required to be $p<0.01$. Cohen's kappa coefficient was also calculated using JMP, as a measure of the agreement between two methods for measurement.

# 4 Results and discussion

## 4.1 Use of Raman Spectroscopy for quantification of bacteria and species differentiation

Initially, aqueous suspensions of *E. coli, B. subtilis* and pure water were scanned using a low-resolution Raman spectrometer, with laser excitation of 785 nm, in an attempt to identify indicative peaks for bacteria. The data did not show any significant peaks or areas of major deviation between the bacteria suspensions and the water samples, even at a concentration of $10^8$ and $10^7$ CFU/ml for *E. coli* and *B. subtilis*, respectively (Figure 7). This is probably because the Raman signal of bacteria is very weak, and

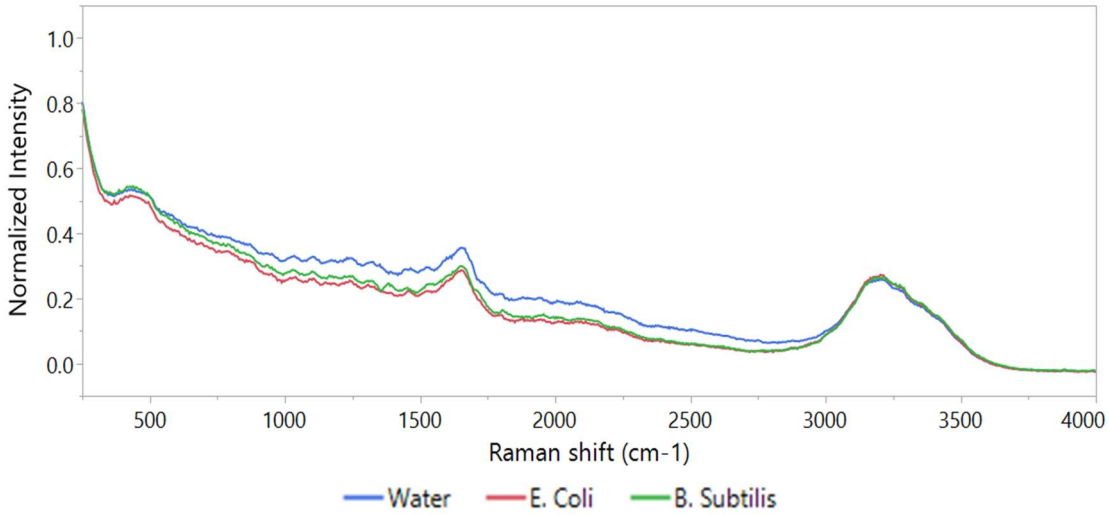


Figure 7. An example of the Raman spectra of clean double-distilled water, *E. coli* at a concentration of $10^8$ and *B. subtilis* at a concentration of $10^7$ CFU/ml. Excitation is by 785 nm laser, for 5 seconds, 3 scans were averaged.

even at high bacterial concentrations the chance of a Raman scatter occurring at a frequency which can be clearly visible using a low-resolution instrument is low (Jarvis et al., 2006; Schmilovitch et al., 2005).

Attempts have been made to improve the spectrum, by increasing integration time, cooling the samples to 4°C, or boiling the samples to lyse the cells and expose the intracellular content, which might better respond to Raman (Premasiri et al. 2017). None of these treatments showed significant improvement. Since no clear difference could be found visually, a PLS model was designed to expose underlying information.

### 4.1.1 PLS model

A PLS model, based on the *B. subtilis* dataset (n=184), was optimized since a slightly better resolved spectrum could be recognized visually for this this microorganism. In order to build a set protocol, different preprocessing approaches were tested and their statistical characteristics are detailed in Table 3. These include latent variables (LVs), RMSE and $R^2$. In all methods, the same attribution of calibration-validation sets was used, with 60% of samples used for calibration and 40% for validation.

Table 3. Comparison of different preprocessing approaches: RMSE and $R^2$ values are calculated on validation set, n=74. LVs = Latent variables. The best method is the one that shows the highest $R^2$ with the lowest possible RMSE, and is in **bold**.

| # | **Preprocessing 1** | **Preprocessing 2** | **LVs** | **RMSE** | **$R^2$** |
|---|---|---|---|---|---|
| **1** | None | None | 7 | 1.04 | 0.56 |
| **2** | None | Centering | 7 | 1.14 | 0.58 |
| **3** | None | Scaling | 3 | 0.4 | 0.08 |
| **4** | None | Centering + Scaling | 5 | 1.35 | 0.51 |
| **5** | Normalized to Max | None | 7 | 1.07 | 0.56 |
| **6** | **Normalized to Max** | **Centering** | **6** | **1.07** | **0.58** |
| **7** | Normalized to Max | Scaling | 11 | 1.43 | 0.29 |
| **8** | Normalized to Max | Centering + Scaling | 6 | 1.46 | 0.4 |
| **9** | Normalized + 1/signal | None | 6 | 0.78 | 0.06 |
| **10** | Normalized + 1/signal | Centering | 8 | 1.03 | 0.07 |
| **11** | Normalized + 1/signal | Scaling | 8 | 1.42 | 0.22 |
| **12** | Normalized + 1/signal | Centering + Scaling | 6 | 1.41 | 0.41 |
| **13** | Normalized + LOG10(1/signal) | None | 2 | 0.74 | 0.17 |
| **14** | Normalized + LOG10(1/signal) | Centering | 2 | 0.72 | 0.23 |
| **15** | Normalized + LOG10(1/signal) | Scaling | 3 | 0.75 | 0.23 |
| **16** | Normalized + LOG10(1/signal) | Centering + Scaling | 5 | 1.2 | 0.5 |
| **17** | Normalized + 1$^{st}$ derivative | None | 2 | 1.22 | 0.46 |
| **18** | Normalized + 1$^{st}$ derivative | Centering | 1 | 1.22 | 0.45 |
| **19** | Normalized + 1$^{st}$ derivative | Scaling | 2 | 1.14 | 0.37 |
| **20** | Normalized + 1$^{st}$ derivative | Centering + Scaling | 1 | 1.16 | 0.38 |

Since normalization and centering was found to best describe the data, the following analysis was done using these preprocessing steps.

### 4.1.2 Establishing detection thresholds

Constructing a PLS model for quantifying both *E. coli* and *B. subtilis* has shown significant difference between species detection thresholds. While *E. coli* could only be detected at $10^6$ CFU/ml, *B. subtilis* prediction models enabled confident detection at $10^2$ CFU/ml (Figure 8).

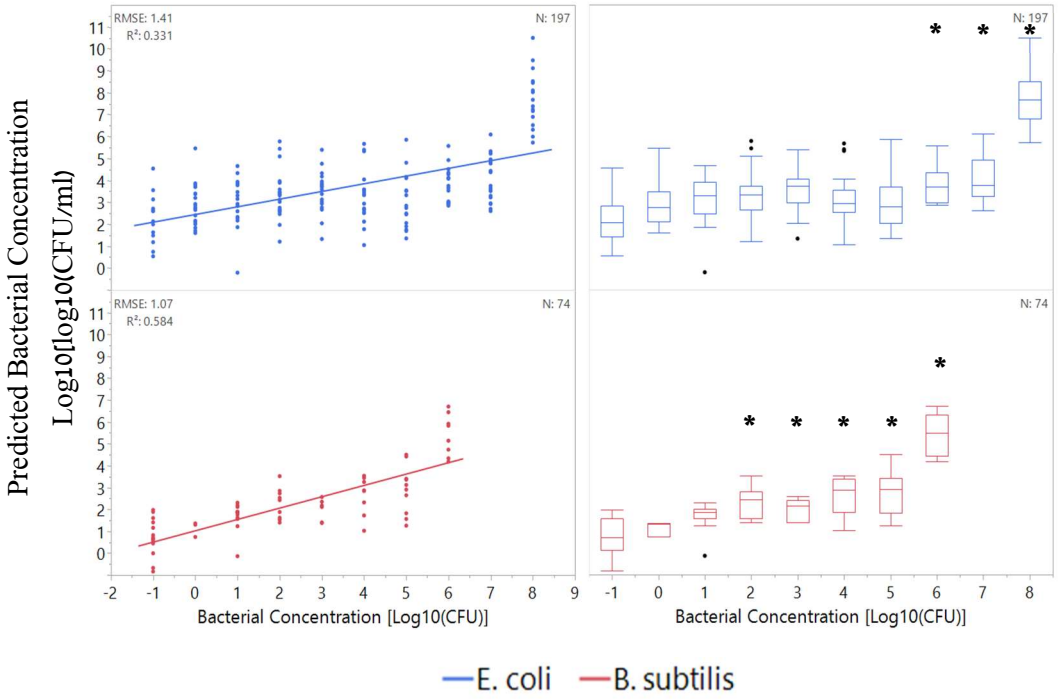


Figure 8. Left: regression of predicted bacterial concentration according to PLS model against real bacterial concentration of *E. coli* and *B. subtilis*. Right: box plot of the data as used for threshold analysis. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown as black dots. Asterisks signify diffrence from control (concentration of 0 CFU/ml) according to Student's t-test, $p<0.01$ .N of *E. coli* = 197, N of *B. subtilis* = 74. Only validation set is shown.

Since detectable *E. coli* concentrations are very high, the analytical performance of this method for this species is generally not sufficient. However, in the case of *B. subtilis,* a significant improvement of detection capability is expected when using low resolution Raman spectroscopy as compared with use of absorbance at 590 nm. However, these results do not seem promising for the rapid and accurate detection of bacteria in water.

### 4.1.3 Differentiation between species

In order to explain the differences between the detection abilities of *E. coli* and *B. subtilis*, variable importance (VI) was calculated and plotted (Figure 9). VI is calculated by sum of the decrease in error when split by a variable. Generally, values over 0.8 are considered significant in models with many variables (Eriksson et al., 2006). It appears that the different bacteria have different spectral fingerprints. The spectral fingerprint of *B. subtilis* includes an additional peak in the functional group region between 3,000-4,000 $cm^{-1}$ (Figure 9). This region signifies that the signal was likely from biological molecules, such as sugars and lipid functional groups (Premasiri et al., 2005; Zeiri et al., 2004; Zeiri and Efrima, 2010). It is difficult however to assign a specific structure to the fingerprint, since it is not a sharp, well resolved peak. This is probably due to many different molecules contributing to the Raman effect. The

fingerprint of *E. coli* includes two sharp peaks at ~1,340 $cm^{-1}$ and 1,700 $cm^{-1}$. These seem to coincide with the spectral fingerprint of pyruvate and pyruvate analogs (such as acetoacetate) that are derived from the symmetric and asymmetric $vCO_2^-$ stretches (De Gelder et al., 2007), but since the biological sample is complex and no reference substance was measured, we cannot assign any significant bands to molecules.

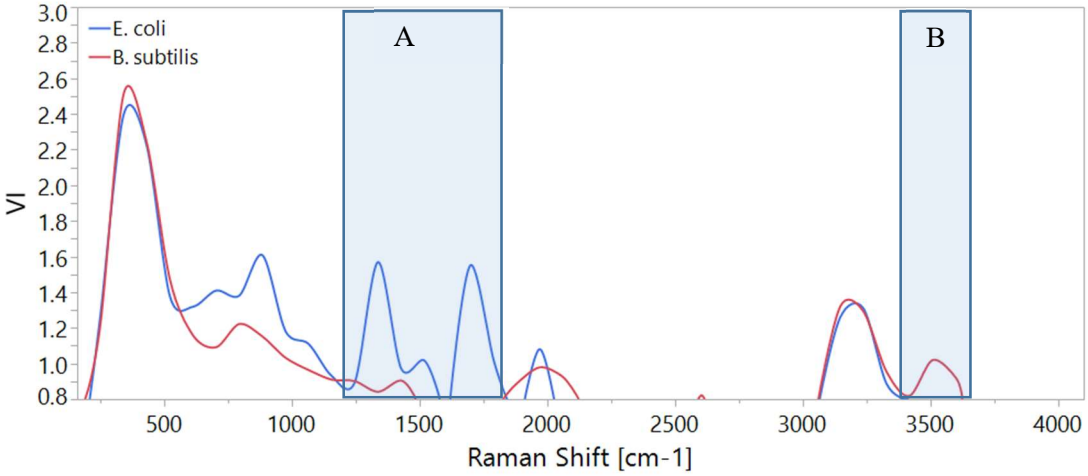


Figure 9. Variable importance (VI) plot of PLS models for quantification of bacteira according to Raman spectra. The plots show only values >0.8. Shades A and B indicate areas of signals which may be indicative of specific bacterial species.

These findings suggest the possibility of differentiation between different bacterial species using Raman spectroscopy. In order to examine the ability of Raman spectroscopy to identify the bacteria in water, control experiments were carried out with the samples divided into 3 categories:

1. *E. coli* at $10^8$ CFU/ml (n=42)
2. *B. subtilis* at $10^6$ CFU/ml (n=51)
3. Pure water samples (n=67)

Then, a PLS-DA (Discriminant Analysis) model was calculated, in an attempt to classify these samples correctly, with the same training-validation ratio of 60% calibration and 40% validation. Table 4 illustrates the model's ability to accurately differentiate between two different bacteria and water, using a confusion matrix. This allows visualization of the performance of the model, where each row of the matrix represents the instances in an actual class while each column represents the instances in the predicted class. Wherever the actual and prediction coincide is highlighted. As can be seen in Table 4 the model accurately differentiates between *E. coli* and *B. subtilis* and also detects their presence in water. This model has almost prefect agreement according to its Cohen's kappa coefficient of 0.93 (Landis and Koch, 1977), indicating the significant predictive abilities of the low-resolution Raman spectroscopy data when modelled.

Table 4. Confusion matrix of classification of bacteria by PLS-DA, based on Raman spectra, at concentrations $10^8$ and $10^6$ CFU/ml for *E. coli* and *B. subtilis* respectively. Only validation set is displayed. Cohen's kappa coefficient displayed at the top (κ).

| | | Prediction κ=0.93 | | | |
|---|---|---|---|---|---|
| Actual | | *B. subtilis* | *E. coli* | Water | Total |
| | *B. subtilis* | 18 | 0 | 1 | 19 |
| | *E. coli* | 0 | 18 | 0 | 18 |
| | Water | 2 | 0 | 28 | 30 |
| | Total | 20 | 18 | 29 | 67 |

Taken together, these results indicate a poor ability to detect *E. coli* and *B. subtilis* in water using low resolution Raman spectroscopy. This finding was surprising since several earlier studies have already demonstrated a potential of Raman spectroscopy for detecting bacteria (Mizrach et al., 2007; Schmilovitch et al., 2005; Stöckel et al., 2015). It was suggested that suspending the bacteria in distilled water would weaken the intensity of their Raman spectra, due to osmotic-stress-related lysis, or that the laser used might damage the cells and change the signal. However, a viability test has shown that the cell viability was not compromised under the experimental setting used in this study. Moreover, neither the use of lower laser energy nor suspending the cells in a saline solution, instead of water, improved Raman signal (data not shown). One of the possible explanations for the discrepancy is that the current instrument may not be appropriate for this purpose, and a more sensitive, yet still low cost, instrument may yield better results.

## 4.2 Use of fluorescence spectroscopy for quantification of bacteria and species differentiation

### 4.2.1 Fluorescence emission of bacteria examined at a single excitation/emission wavelength pair

#### 4.2.1.1 Calibration curve for tryptophan concentration

In fluorescence spectroscopy, one approach in sample analysis is the 'peak picking' or measuring emission intensity at a certain wavelength, after excitation by another wavelength, lesser by magnitude. This method has been applied in many studies for the detection of microbial contamination of water, based on fluorescence of proteinaceous substances in the tryptophan-like-fluorescence (TLF) area at the of 275±5 nm excitation and 360±10 nm emission (Baker et al., 2015; Bridgeman et al., 2015; Cumberland et al., 2012; Sorensen et al., 2018a; Sorensen et al., 2015; Sorensen et al., 2018b). In these studies, the minimal detectable microbial concentration was equivalent to the 1-3 ppb concentration of free tryptophan.

Dilutions of L-tryptophan at concentrations of 0.1, 1, 2, 3, 5 and 10 ppb were prepared. Each dilution was scanned 3 times and 2 scanning sessions were conducted (6 scans per each concentration). The maximum

fluorescence peak of tryptophan was visually found to be at 362 nm after excitation of 275 nm (data not shown).

A calibration curve was plotted to enable quantification of tryptophan according to the emission intensity at 362 nm upon excitation by 275 nm light (Figure 10). Accordingly, a linear regression was calculated and a tryptophan equivalence formula extracted from the regression line.

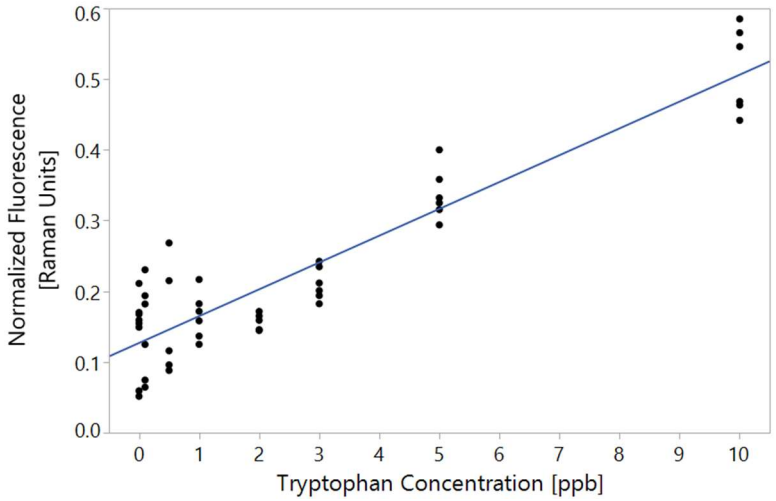


Figure 10. Correlation between tryptophan concentration and fluorescence at 362 nm after at excitation of 275 nm. N=48 (51-3 excluded). p<0.01 according to Student's t-test.

Equation 3. Claculation of tryptophan equivalence from normalized fluorescence

$$Tryptophan\ Equivalence_{[ppb]} = \frac{Normalized\ fluorescence - 0.127}{0.038}$$

$$\text{Where the normalized fluorescence } = \frac{(\lambda_{ex}275/\lambda_{em}362)}{(\lambda_{ex}275/\lambda_{em}305)}$$

#### 4.2.1.2 Quantification of individual laboratory-grown bacteria suspended in double distilled Non-fluorescent water

The tryptophan-like fluorescence (at excitation/emission of 275 and 362 nm respectively) of *E. coli, B. subtilis* and *P. aeruginosa* in high-purity non-fluorescent water (HPLC-grade water) was plotted against their concentration (in CFU/ml) in Figure 11. The log concentration given as -1 is assigned to zero CFU/ml value. The detection thresholds of the emission measurements at single excitation/emission wavelength pair in detecting different bacterial species are marked in Figure 11. Significance was tested using a Wilcoxon non-parametric test, since the variances are not equal according to Levene's test. The differences in fluorescence can only be observed when bacterial concentration reached $10^3$-$10^4$, which is roughly 0.15 Raman units or the equivalent of ~ 3 ppb of tryptophan calculated using Equation 3.

This suggests that the studies using microbial detection at a single excitation/emission wavelength and

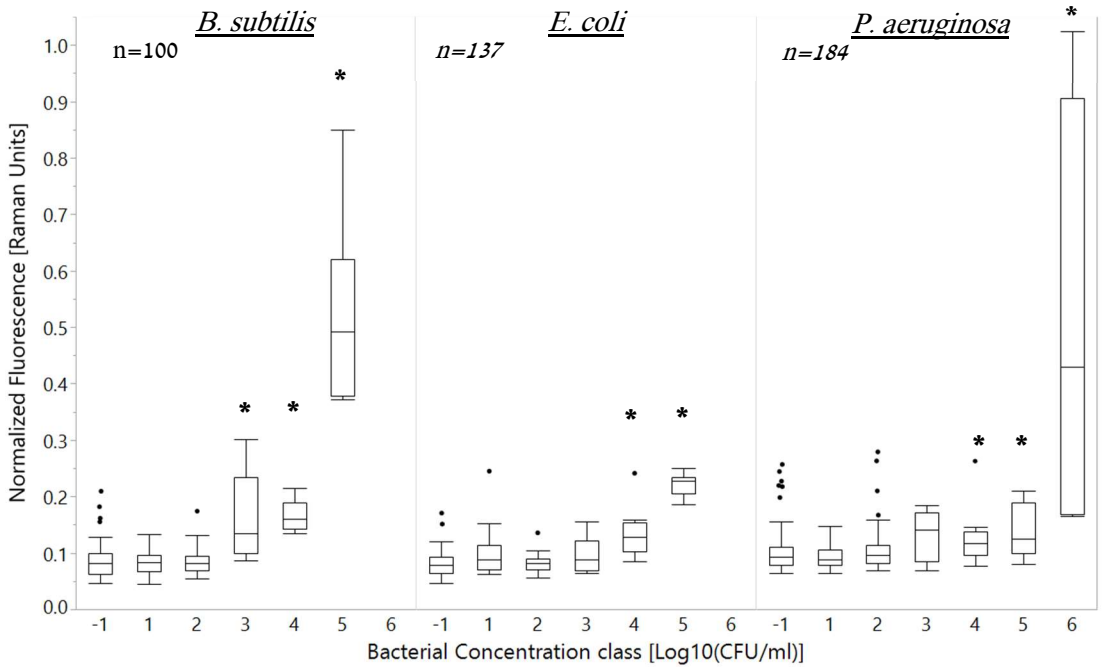


Figure 11. Fluorescence measurements at excitation-emission of 275-362 nm correlated with specific bacterial species. Bacterial concentration of log10 (-1) means 0 CFU/ml. Asterisks mean significant differenct according to a Wilcoxon test, p<0.01. n indicated at the top. Bacterial concentration class means the all measurements were rounded to nearest class (i.e., 1,2,5 CFU/ml regarded as 1) for convenience. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown as black dots.

claiming the detection limit equivalent to 1-3 ppb of tryptophan are capable of detecting bacteria at the ~$10^4$ CFU/ml level. This detection limit seems to be insufficient to the requirements of Israeli Ministry of Health regulations (2013), as well as to others' (Allen et al., 2004), but it could have implications, e.g., in water, food, pharmaceutical and beverage industry for monitoring microbial contamination at a relatively early stage, and much earlier than most methods enable, using a quick, cheap fluorimeter.

#### 4.2.1.3 Detection of heterotrophic bacteria in groundwater

Groundwater samples from six wells in Israel were examined in the period between July 2017 and June 2018 by means of fluorescence spectroscopy. Over the entire sampling period, only 1 sample contained HPC of >1,000 CFU/ml. That sample was also the only one to contain *E. coli* >1 CFU/100 ml. Therefore, the only available microbial indicator was HPC. Fluorescence emission intensity was plotted against the HPC of groundwater samples and are presented in Figure 12. Figure 13 demonstrates that measuring fluorescence emission at a single excitation/emission wavelength was sufficient to differentiate between water samples of different concentrations (according to Wilcoxon ranked sum test, $p<0.01$).

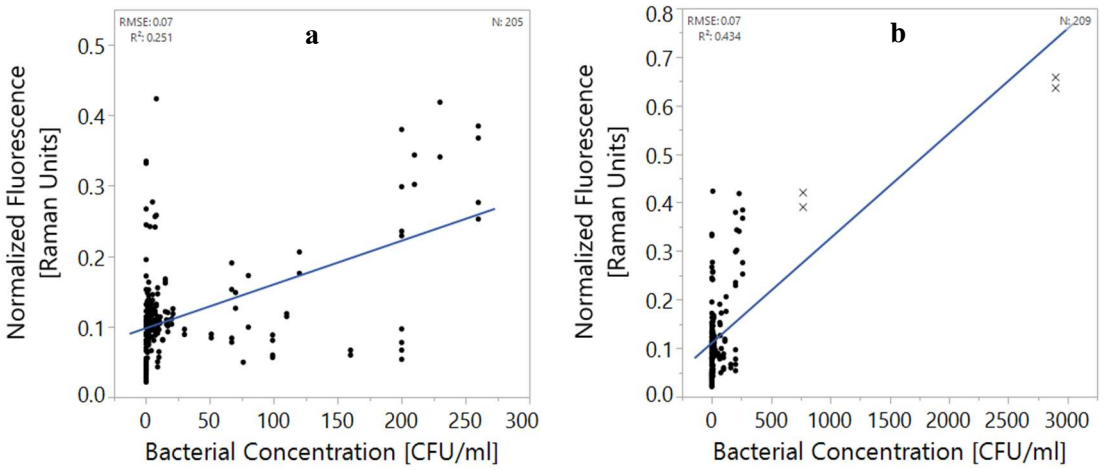


Figure 12. Correlation between Fluorescence measurements at excitation-emission of 275-362 nm and HPCs in groundwater samples. (a) includes only the data with CFU <300 and (b) shows the entire data set (n=209). Outliers with >300 CFU/ml are marked as X. All regressions are statistically significant, with p-value < 0.0001 using a Student's t test.

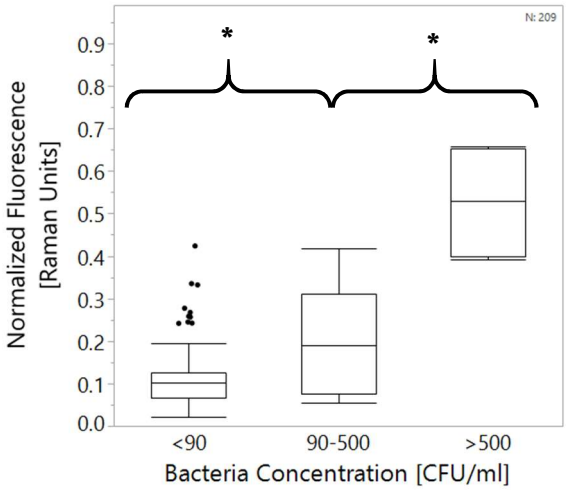


Figure 13. Box plots of tryptophan-like fluorescence measurements vs. heterotrophic bacteria counts in groundwater. All groups are significantly differenct from each other according to a Wilcoxon test, $p<0.01$. n=209. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown as black dots.

#### 4.2.1.4 Case study – Shimron well

The water sample from Shimron well obtained at 22/5/2018 was the only case of severe contamination i.e., HPC of 2,900 CFU/ml and 2 CFU/100 ml of *E. coli* which indicate fecal contamination. The sample had very strong fluorescence (Figure 14A), easily detected visually, however, chlorination at the treatment site (B) and/or laboratory filtration with 0.45 µm filter (C, D) reduced significantly the TLF at the excitation range of 210-280, and emission at 320-380).

Fluorescence of all other groundwater samples was practically non-detectable visually as exemplified in Figure 15. All samples showed strong optical signals along the Rayleigh scatter line, where the excitation wavelength equals to emission wavelength, and lesser one, but significant, along the Raman scatter line of water (marked in Figure 15) however, no distinct signals of proteinaceous (or any other fluorescence) could be easily detected.

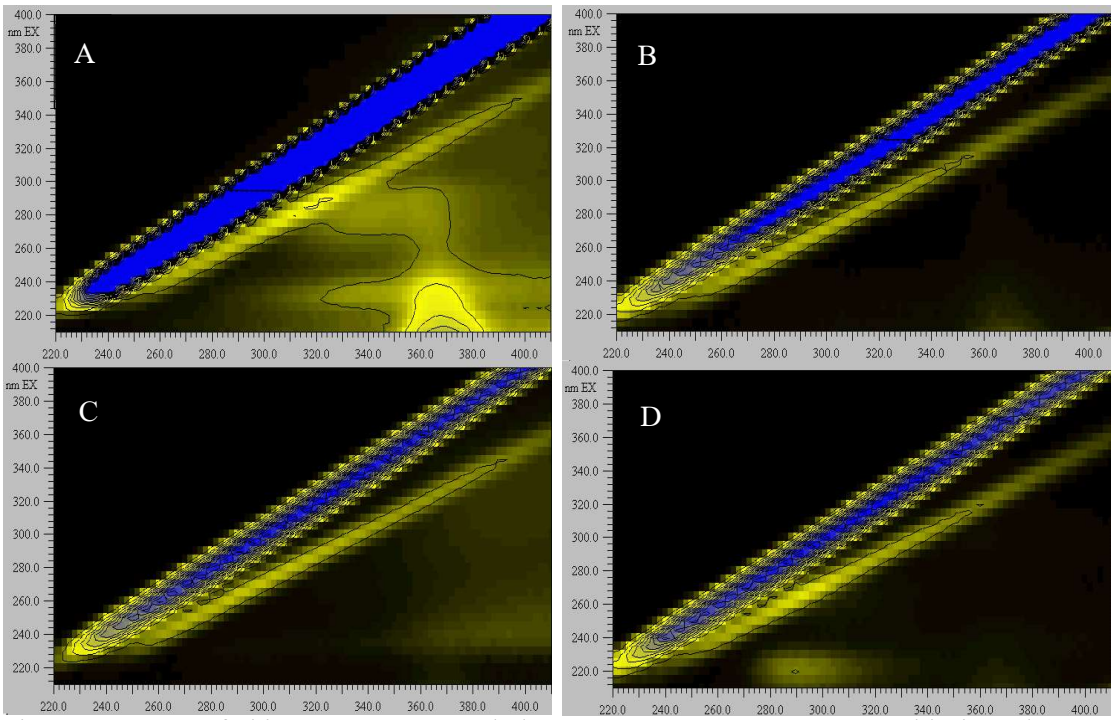

Figure 14. EEMs of Shimron 7 water sampled on 22/5/2018. (A) raw water, (B) chlorinated water, (C) filtered raw water, (D) filtered chlorinated water.

By examining the highly contaminated samples, it was clear that the emission was not limited to the TLF region but was actually spread over a large area of the EEM. This, including the entire proteinaceous region (the 220-300 nm excitation wavelength range, the 300-380 nm emission wavelength range (Yang et al., 2015). Furthermore, the edge of the humic substance region, excitation in the 210-300 nm range and the emission above 400 nm (Yang et al., 2015), also included a strong peak with excitation around 210 nm and emission at 370±10 nm emission, as described by Simelane (2013). This has led us to hypothesize that examining the entire EEM could improve the detection abilities. This was particularly relevant since samples with <500 CFU/ml showed very little overall fluorescence (Figure 15). Such weak signals could be properly analyzed using PLS multivariate analysis.

Indeed, a common approach for analysis of fluorescence data uses entire excitation-emission matrices (EEMs) (Borisover et al., 2009; Carstea et al., 2016; Heibati et al., 2017). PLS regression enables the formulation of a model for predicting the concentration of bacteria in water, based on the entire EEM data.

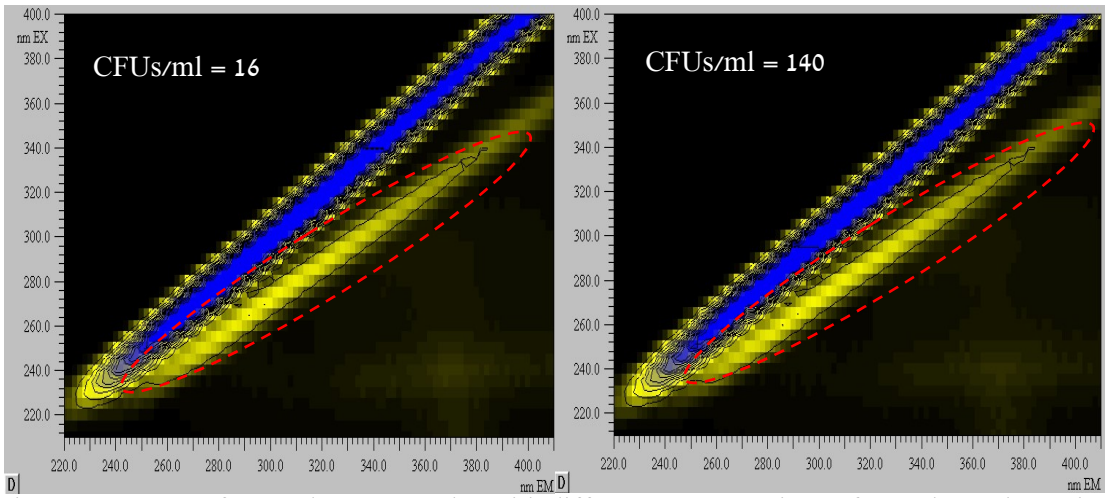


Figure 15. EEMs of groundwater samples with different concentrations of waterborne bacteria, indicated in the top of each panel. The zone marked in blue is the Rayleigh scatter region, and the area surrounded by a red, dashed line is the Raman scatter region.

### 4.2.2 Multispectral fluorescence spectroscopy and PLS analysis

In order to analyze the entire EEM, a series of preprocessing approaches were initially tested on the *E. coli* data set (n= 137, Table 5). In all methods, 60% of the samples were used to calibrate a model, and 40% of the samples served to validate it.

Table 5. Comparison of different preprocessing approaches: RMSE and $R^2$ values are calculated on validation set, n=60. LVs = Latent variables. The best method is the one that shows the highest $R^2$ with the lowest possible RMSE, and is in **bold**.

| # | **Preprocessing 1** | **Preprocessing 2** | **LVs** | **RMSE** | **$R^2$** |
|---|---|---|---|---|---|
| **1** | None | None | 12 | 0.96 | 0.62 |
| **2** | None | Centering | 11 | 0.97 | 0.60 |
| **3** | None | Scaling | 2 | 0.61 | 0.59 |
| **4** | None | Centering + scaling | 4 | 0.77 | 0.68 |
| **5** | Normalized | None | 10 | 1.12 | 0.54 |
| **6** | Normalized | Centering | 9 | 1.15 | 0.55 |
| **7** | Normalized | Scaling | 2 | 0.9 | 0.49 |
| **8** | Normalized | Centering + scaling | 8 | 0.98 | 0.65 |
| **9** | Normalized + 1/signal | None | 7 | 0.7 | 0.74 |
| **10** | Normalized + 1/signal | Centering | 7 | 0.78 | 0.72 |
| **11** | Normalized + 1/signal | Scaling | 13 | 0.74 | 0.78 |
| **12** | **Normalized + 1/signal** | **Centering + scaling** | **7** | **0.7** | **0.78** |
| **13** | Normalized + LOG10(1/signal) | None | 7 | 0.85 | 0.71 |
| **14** | Normalized + LOG10(1/signal) | Centering | 7 | 0.83 | 0.73 |
| **15** | Normalized + LOG10(1/signal) | Scaling | 8 | 0.79 | 0.72 |
| **16** | Normalized + LOG10(1/signal) | Centering + scaling | 8 | 0.73 | 0.77 |
| **17** | Normalized + 1st derivative | None | 1 | 0.13 | 0.1 |
| **18** | Normalized + 1st derivative | Centering | 5 | 0.96 | 0.45 |
| **19** | Normalized + 1st derivative | Scaling | 3 | 0.81 | 0.48 |
| **20** | Normalized + 1st derivative | Centering + scaling | 4 | 0.83 | 0.45 |

Since the combination of normalization, 1/signal centering and scaling was found to best describe the data, i.e., having the highest $R^2$ value with the lowest RMSE, this protocol was used for the rest of the analysis.

#### 4.2.2.1 Quantification of laboratory-grown *E. coli* suspended in double distilled Non-fluorescent water: a PLS-enhanced efficacy

The PLS algorithm was used to quantify *E. coli* cells suspended in in double distilled non-fluorescent water by using emission spectra obtained at a certain excitation wavelength and by analyzing the full EEM. The rational of introducing emission spectrum obtained at a single excitation wavelength into the consideration was that if it would be found successful and competitive with the full EEM analysis, operational and computational costs can be reduced. Both quantification methods were compared also with use of a single excitation/emission wavelength pair, as was described in detail in section 4.2.1.

Therefore, the detection thresholds and regression coefficients ($R^2$) were compared for the following detection methods:

1) Single wavelength method – the emission intensity at 362 nm upon excitation at 275 nm
2) Emission spectrum recorded in the 220-450 nm wavelength interval upon excitation at 275 nm.
3) The entire EEM obtained for excitation and emission wavelengths between 210-400 and 220-450 nm, respectively.

In order to measure predictive abilities for the single excitation/emission wavelength pair, a calibration curve was plotted, linear regression was calculated and predictions were made based on the regression equation (Figure 16).

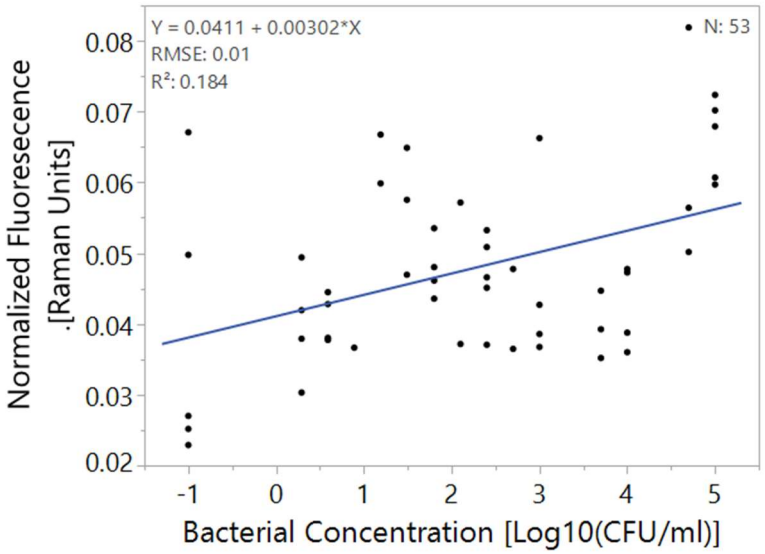


Figure 16. Linear regression of *E. coli* concentration and normalized fluorescence at excitation/emission pair 275/362. N=53.

Detection threshold was defined as the lowest concentration level differing from <10 CFU/ml. Tukey-Kramer tests were used. (Table 6, Figure 17).

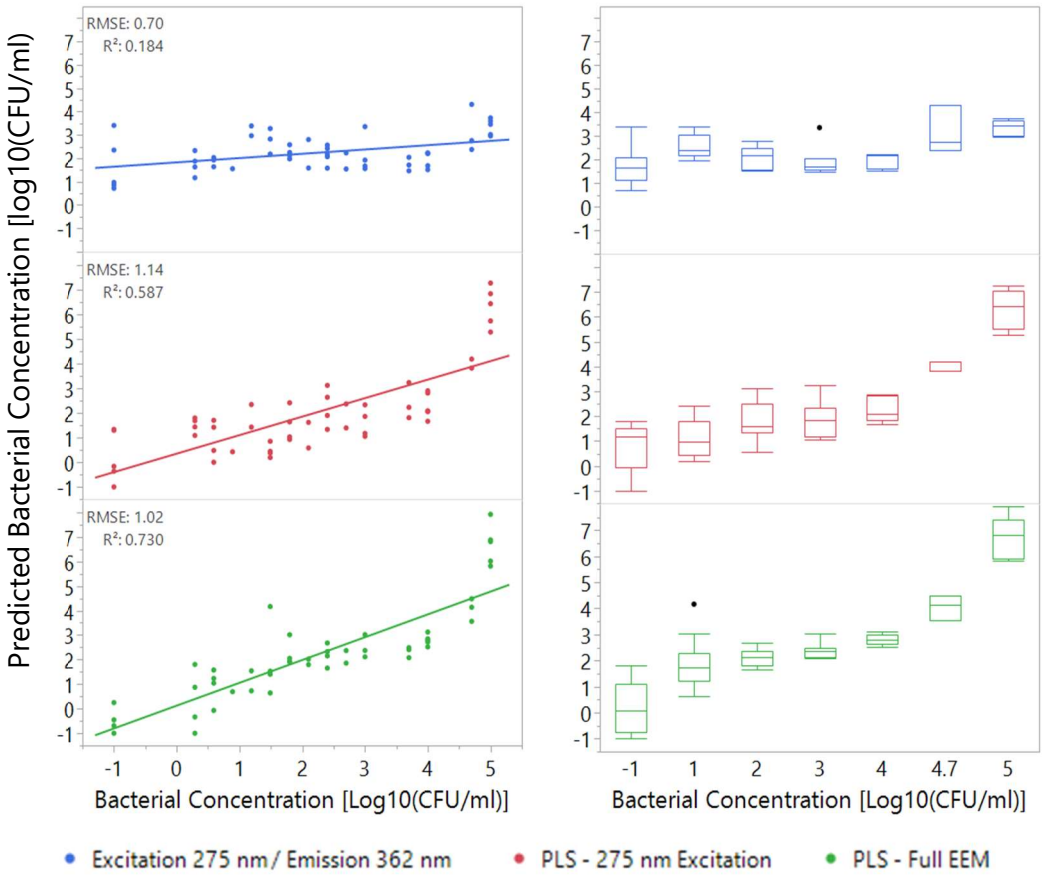


Figure 17. Left: regression of *E. coli* concentration with different detection methods, Right: box plot of the data as used for threshold analysis. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown as black dots. N=53.

Table 6 Comparison of fluorescence-based methods for the detection of bacteria. Analysis was performed for the validation sets of the *E. coli* dataset. Detection threshold defined as a difference between the indicated concentration and <10 CFU/ml is different significantly (p<0.01) according to a Tukey-Kramer test. n=53.

| Method | $R^2$ | RMSE [$Log_{10}$(CFU/ml)] | Detection Threshold [CFU/ml] |
|---|---|---|---|
| Single excitation/emission wavelength pair [275/362] | 0.09 | 0.22 | $5*10^4$ |
| PLS-analyzed emission spectrum obtained upon excitation at 275 nm | 0.59 | 1.14 | $10^3$ |
| PLS- analyzed full EEM | 0.73 | 1.02 | 10 |

#### 4.2.2.2 Detection of *E. coli, B. subtilis* and *P. aeruginosa* suspended in water using PLS-analyzed emission spectrum obtained upon excitation at 280 nm or full EEMs

At this stage, PLS models were created to try and predict bacterial concentrations of *E. coli, B. subtilis* and *P. aeruginosa* as representatives of common pathogens. These bacterial species were chosen because a) they represent the gram-positive and gram-negative groups, expected to have a different chemical composition (Jean et al., 2016) and b) *P. aeruginosa* is a very common environmental pathogen which may be acquired from drinking water (Costa et al., 2015). Each strain was first modeled alone, and later all samples were modeled together. Detection threshold defined as a difference between the indicated concentration and <10 CFU/ml is different significantly ($p<0.01$), using Wilcoxon ranked-sum test or Tukey-Kramer when available. Results are shown in Table 7.

Table 7. Detection of bacteria in water: comparison of fluorescence-based PLS-supported approaches. All $R^2$ values are statistically significant at $p<0.01$.

| **Bacteria** | | *E. coli* (N=53) | *B. subtilis* (N=46) | *P. aeruginosa* (N=86) | All bacteria (N=185) |
|---|---|---|---|---|---|
| **Emission spectrum upon excitation at 275 nm** | **RMSE** [$Log_{10}$(CFU/ml)] | 1.14 | 0.74 | 0.9 | 0.83 |
| | **$R^2$** | 0.59 | 0.42 | 0.19 | 0.25 |
| | **Detection threshold** (Wilcoxson) [CFU/ml] | $5*10^4$ | $10^5$ | $10^5$ | $10^4$ |
| | **Detection threshold** (Tukey-Kramer) [CFU/ml] | $10^3$ | --- | $10^6$ | --- |
| **Full EEM** | **RMSE** [$Log_{10}$(CFU/ml)] | 0.76 | 1.01 | 1.04 | 1.06 |
| | **$R^2$** | 0.74 | 0.65 | 0.55 | 0.48 |
| | **Detection threshold** (Wilcoxson) [CFU/ml] | 10 | $10^4$ | 10 | 100 |
| | **Detection threshold** (Tukey-Kramer) [CFU/ml] | 10 | $10^3$ | 10 | 100 |

The detection thresholds of the three bacterial species, for both approaches, ranged from 10 to $10^5$ CFU/ml. However, all bacteria could be detected using the full EEM PLS model at $10^3$ CFU or lower, even when different bacteria were modelled together, indicating the common spectral fingerprint of all the species was sharp enough to enable detection at very low concentrations. This finding is promising regarding the use of this method for industrial purposes since all species were detected below the $10^3$ CFU/ml required threshold for HPC (Israeli Ministry of Health 2017 ,2016 ,2013 ; Ashbolt, 2015).

The detection thresholds of these species differed when using the entire EEM or only the single emission spectrum; the differences reached 1-5 orders of magnitude (Table 7). This indicates that the single emission spectrum does not capture the entire spectral fingerprint of the bacteria. In addition, for all bacteria, $R^2$ and RMSE values improved when using the entire EEM. Also, for *B. subtilis* and the model, which combined all three species, it was not possible to use the Tukey-Kramer test, since the variance of the different groups was too diverse, indicating a less coherent model.

This is possible because a large fraction of the proteinaceous region in a fluorescence EEM is not analysed when only a single 2-dimentional segment is analysed. The differences in protein contents, chemical composition and structures, may not be properly captured by examining a single emission spectrum. These subtle differences are further overlooked when measuring fluorescence emission at a single excitation-emission wavelength pair. These differences include also the spatial distributions of aromatic amino acids within proteins and the different aromatic moieties in bacterial cells as well as the interactions of fluorophores with the surrounding medium. Interestingly, when application of PLS-analysed emission spectrum improves detection ability by ~3 orders of magnitude when compared to the peak-picking approach. Further improvement is reached when 3D EEM data is analysed with PLS. In the analysis of 2D or 3D data, there is an amplification effect caused by the distribution of signals. Since a signal is calculated not only by its peak, but by the whole distribution of the signal over several wavelength pairs, the same signal receives more intensity units. This is especially correct for full EEM analysis, since the distribution is 3D. This further explains the improvement of detection using the full EEMs.

In addition, Gram negative bacteria (*E. coli, P. aeruginosa*) were detected at a concentration of 10 CFU/ml, while the gram-positive bacterium (*B. subtilis*) was detected only at $10^3$ CFU/ml (Table 7). This may be due to the different nature of the cell wall of these bacteria; since gram-positive bacteria have a thick layer of peptidoglycan that does not contain aromatic amino acids (Jean et al., 2016). Gram-negative bacteria are known to have an outer membrane, which is abundant with proteins, some of which might contain aromatic side-chains (Madigan et al., 2006), thus increasing the whole cellular emission. It is possible that the different detection threshold resulted from different spectral fingerprint of each bacterium. Further studies with other gram-positive and gram-negative species should be done in order to test thus notion.

#### 4.2.2.3 Detection of heterotrophic bacteria in groundwater using PLS-analyzed emission spectrum obtained upon excitation at 280 nm or full EEMs

In this analysis, EEMs of groundwater samples, collected throughout the year, were modelled using a PLS algorithm. Two samples were significantly different from all others, and could easily detected visually. Both samples were from the Shimron-7 well, having 770 and 2900 CFU/ml. Therefore, prior to modelling, those outliers were excluded from the analysis. The results of the PLS models are displayed in Figure 18.

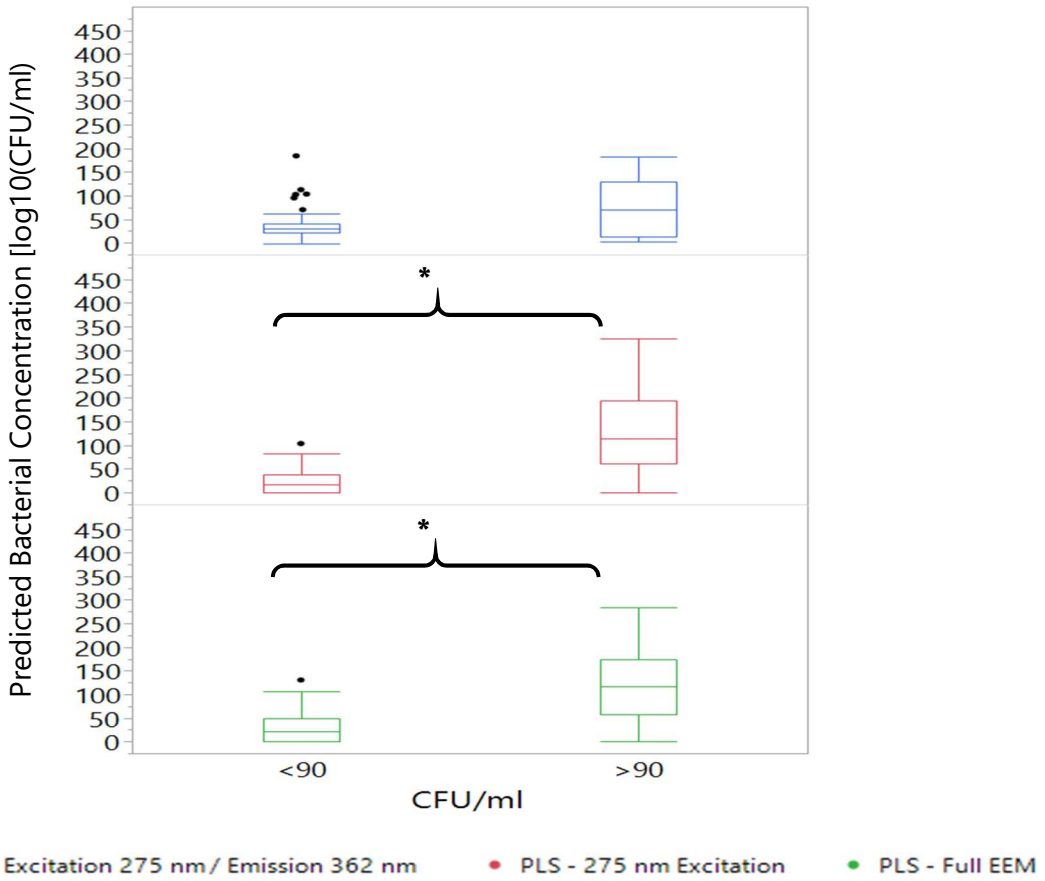


Figure 18. Box plots of predicted heterotrophic bacteria concentrations using different fluorescence spectroscopy analysis approaches versus real heterotrophic bacteria counts in groundwater. Asterisks mark significant difference between groups according to a Wilcoxon test, $p<0.01$. n=57. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown as black dots.

Figure 18 illustrates the difference between using a PLS model for the emission spectrum upon excitation at 275 nm, PLS model for the entire EEM and single wavelength pair analysis (at 275 nm excitation and 362 nm emission). While the differences are not immediately apparent to the eye, statistical testing shows that by using a PLS model based on the full EEM, or even the emission spectrum upon excitation at 275 nm, differentiation between samples with over or under 90 CFU/ml can be achieved.

Currently, based on this study, it is difficult to conclude which approach, the use of EEMs, single emission spectrum or single excitation/emission wavelength pair fluorescence measurements, will be best applied for detection of low concentrations of heterotrophic bacteria in drinking water. However, based on the experimental data regarding the detection and quantification of single species suspensions, there is an advantage for using the whole EEM data for successful detection of microbial contamination of water.

### 4.2.3 Spectral fingerprints based on PLS model variable importance

In order to see which spectral regions of EEM are important for detecting bacteria in water, variable importance (VI) was calculated for each excitation-emission wavelength pair.

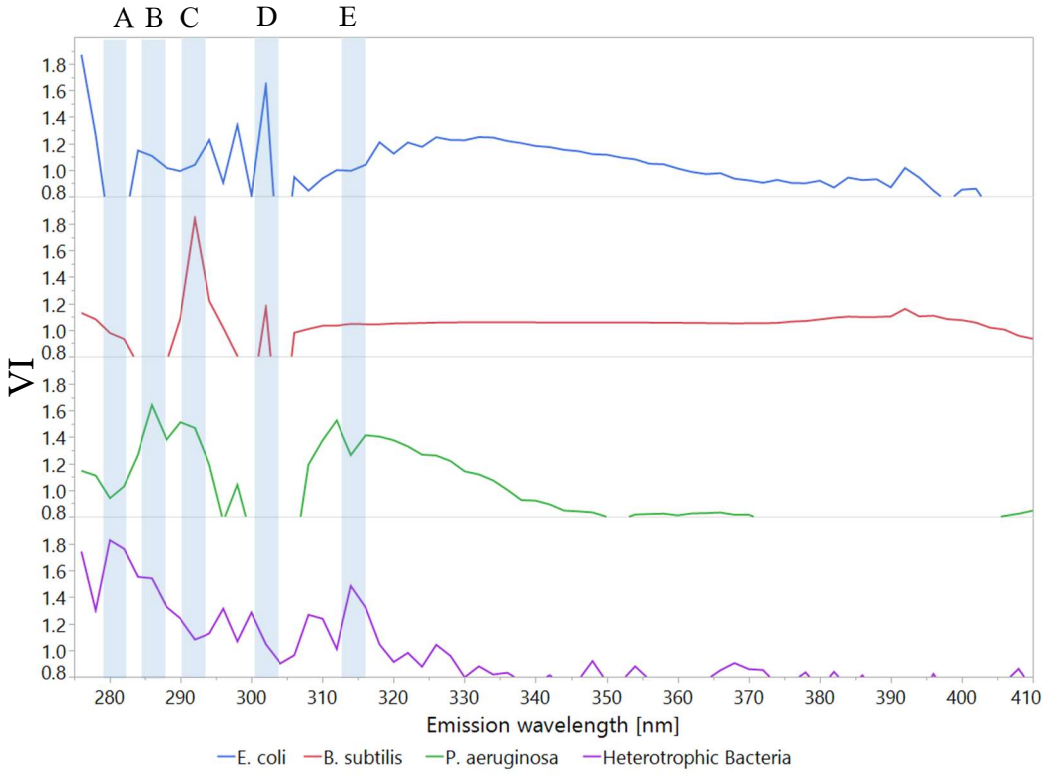


Figure 19. Variable importance (VI) plots of PLS models made from the emission spectrum upon excitation at 275 nm spectra. The plots show only values >0.8. Shades A-F indicate areas of significan signals which may be indicative bacterial concentration.

For this analysis, the VI values were plotted from the PLS model based on the emission spectrum upon excitation at 275 nm (Figure 19). The upper wavelength limit in Figure 19 is 410 nm because the scans of groundwater samples was set at to value, while the pathogen scans was set to 450 nm. The shaded regions in Figure 19 illustrate that there are slight differences between different bacteria's spectral fingerprints, even in cases where the model is based only on the emission spectrum of a single excitation wavelength.

When the VI values of entire EEM, characteristic for the excitation wavelength range of 210-400 nm and the emission wavelength of 220-450 nm, were plotted (Figure 20), a difference between the spectral fingerprints of the different bacterial suspensions could be observed. While most of the signal came from the proteinaceous fluorescence region (excitation wavelengths in the 220-230, 270-300 nm range, the emission in the 300-360 nm range, (Yang et al., 2015), certain differences in this region, between the different bacteria, are visible. Furthermore, it appears that the scatter region, where the excitation and emission wavelengths do not differ more than by 10 nm of the map, is also of high importance to the model.

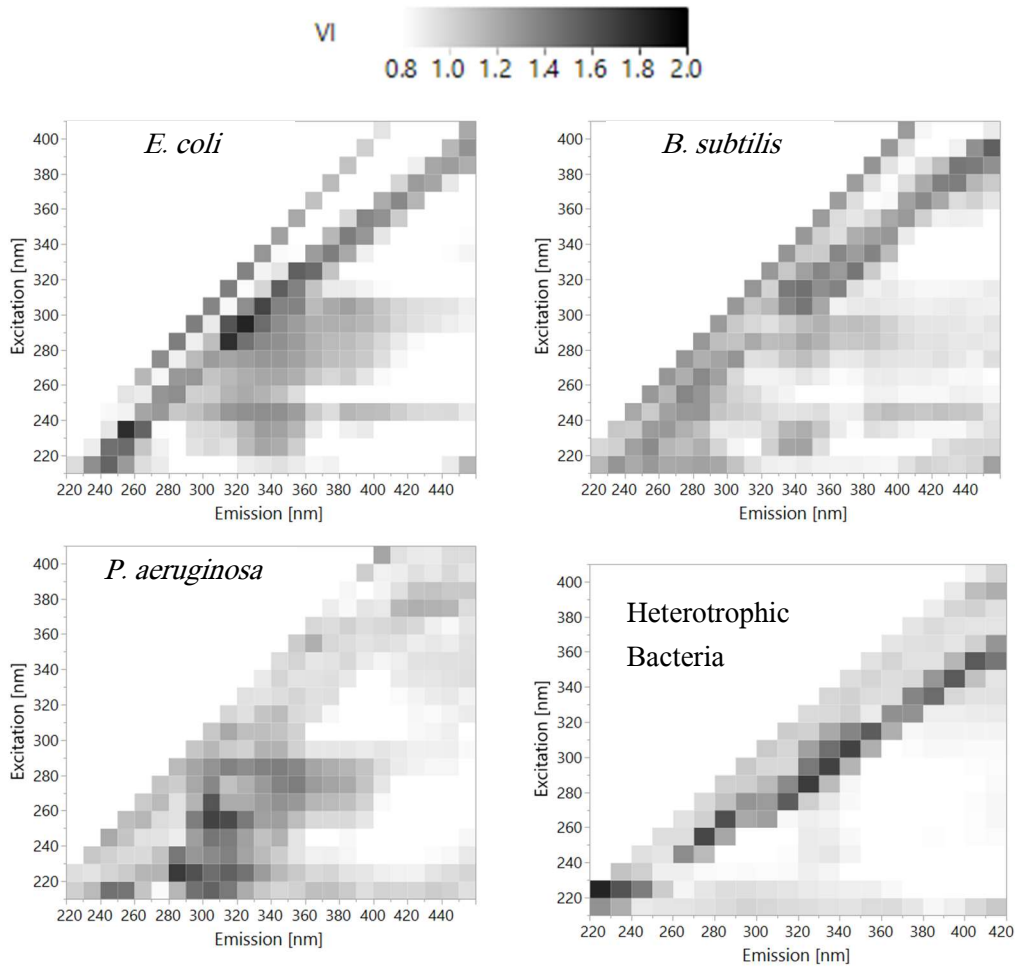


Figure 20. Variable importance (VI) plots of PLS models made from entire EEMs. The plots show only values >0.8.

It is interesting to see the importance of the Rayleigh scatter region and the Raman scattering of water for the quantification of bacteria. It appears that all bacteria have some effect on light scattering, both Raman (of water) and Rayleigh, which is surprising, since Raman spectrometry itself was not sensitive enough to detect these bacteria at concentrations of $10^5$ CFU/ml.

One explanation for this is that while the low-resolution Raman spectrum was measured with excitation by near infrared light (785 nm), the EEM were obtained using UV light (200-400 nm), which has much higher energy levels. It is clear that proteinaceous substances demonstrating tryptophan-like fluorescence form a large part of the spectral fingerprint. These proteinaceous substances may represent entire cells, live or dead, cell wall components or dissolved proteins and peptides that were excreted by the bacteria. Further studies are needed to elucidate the range of chemical components that determine the observed fluorescence fingerprints. It is clear, however, that the intensity of these fingerprints correlates with the concentration of bacteria.

### 4.2.4 Differentiation between bacterial species

In order to examine the ability of fluorescence spectroscopy coupled with modelling to differentiate bacterial species, samples were divided into 4 categories:

1. *E. coli* >$10^4$ CFU/ml
2. *B. subtilis* > $10^4$ CFU/ml
3. *P. aeruginosa* > $10^4$ CFU/ml
4. All samples at ≤1 CFU/ml – classified as 'water'

A PLS-DA model was calculated, in an attempt to classify these samples correctly, with the same training-validation ratio of 6:4. Table 8 illustrates the model's ability to differentiate between *E. coli*, *B. subtilis* and *P. aeruginosa*. However, the differentiation between *P. aeruginosa* and water alone is weak. This model has substantial agreement according to its Cohen's kappa coefficient of 0.78; which means it can correctly classify most of the samples.

Table 8. Confusion matrix of classification of microbes by PLS-DA, based on entire EEMs, at concentrations >$10^4$ CFU/ml. Only validation set is displayed. Cohen's kappa coefficient displayed at the top (κ).

| | Prediction κ =0.78 | | | | | |
|---|---|---|---|---|---|---|
| Actual | | *B. subtilis* | *E. coli* | *P. aeruginosa* | Water | Total |
| | *B. subtilis* | 8 | 0 | 0 | 0 | 8 |
| | *E. coli* | 0 | 8 | 0 | 1 | 9 |
| | *P. aeruginosa* | 0 | 0 | 5 | 7 | 12 |
| | Water | 0 | 0 | 1 | 50 | 51 |
| | Total | 8 | 8 | 6 | 58 | 80 |

It is not clear why *P. aeruginosa* was poorly distinguished from water. Nonetheless, the data analysis suggests that fluorescence spectroscopy of EEMs can be used to differentiate the three species when present alone at concentration of >$10^4$ CFU/ml.

## 5 Summary and Conclusions

The aim of this study was to examine the ability of detecting bacteria in drinking water low-resolution Raman spectroscopy and fluorescence spectroscopy. For this purpose, both methods were used to detect model bacterial species in pure cultures. Low-resolution Raman spectroscopy yielded low detection sensitivity ($\geq 10^6$ CFU/ml for *E. coli* and $\geq 10^2$ CFU/ml for *B. subtilis*). For the purpose of discriminating between different species of bacteria in dense suspensions ($\geq 10^7$ CFU/ml), analysis of Raman shift spectra enabled significant discrimination between *E. coli* and *B. subtilis* as well as pure water. These findings suggest prospective applications, though the method needs to be refined.

Fluorescence spectroscopy can detect bacterial presence in water quickly and accurately. Extending fluorescence analysis from measuring emission at a single excitation-emission wavelength pair to emission spectrum at a certain excitation wavelength and further to use of EEM, resulted in a significant improvement of the limit of detection. A PLS-based chemometric method improved the sensitivity by several orders of magnitudes, from $10^5$ CFU/ml to a detection threshold as low as 10 CFU/ml for *E. coli* and *P. aeruginosa* and $10^3$ CFU/ml for *B. subtilis*. To the best of our knowledge, this is the first study to measure the effect of extending the analysis from single wavelength pair to full EEM on microbial detection threshold.

When compared, we found the use of Full EEM fluorescence spectroscopy analysis to have the best detection ability for lab-grown bacteria, as shown in Table 9.

Table 9 – Summary of detection thresholds different spectroscopy and approaches

| | **Method** | **Low Resolution Raman** | **Single wavelength pair Fluorescence** | **Full EEM Fluorescence** |
|---|---|---|---|---|
| **Detection Threshold [CFU/ml]** | ***E. coli*** | $10^6$ | $10^4$ | 10 |
| | ***B. subtilis*** | $10^2$ | $10^3$ | $10^3$ |

In order to examine if fluorescence spectroscopy can detect bacteria also in natural water sources, groundwater from several wells in the north of Israel were tested over 10 months. Fluorescence was capable of detecting ≥1 CFU/100 ml of *E. coli* contamination in one sample out of 99. All other samples (98) were negative for *E. coli*. A PLS model based on EEM data enabled detection of HPC in concentration of ≥100 CFU/ml (HPC). In light of previous studies that have demonstrated the application of fluorescence spectroscopy for detecting bacteria in low-quality water (Bridgeman *et al.* 2015; Sorensen *et al.* 2015; Sorensen *et al.* 2018a), the detection of bacteria in high quality water, with low microbial abundances, requires the use of chemometric methods on large EEMs.

Furthermore, full EEM analysis of fluorescence data enabled differentiation between *E. coli, B. subtilis* and *P. aeruginosa*, though not as well between *P. aeruginosa* and pure water. These results suggest that both Raman and fluorescence spectroscopies, when analyzed with PLS models, can be used to discriminate between bacterial species in dense suspensions. These findings have industrial and medical significance for rapid identification of bacteria in samples containing high numbers of bacteria in pure

culture. Future studies should examine the ability of spectroscopy to identify bacteria in composite samples containing more than one species.

To conclude, this study has demonstrated a high sensitivity of fluorescence spectroscopy for detection of bacteria in both laboratory model system and in natural groundwater. Collection of a larger set of data, that will be more diverse and include many contamination events, could contribute to a more accurate model, that will enable quick and efficient detection of bacteria in different types of water.

## 6 References


Adam EA, Collier SA, Fullerton KE, Gargano JW, Beach MJ. (2017). Prevalence and direct costs of emergency department visits and hospitalizations for selected diseases that can be transmitted by water, United States. Journal of Water and Health 15, 673-683.

Allen, M.J., Edberg, S.C., and Reasoner, D.J. (2004). Heterotrophic plate count bacteria—what is their significance in drinking water? International Journal of Food Microbiology 92, 265-274.

Ashbolt, N.J. (2015). Microbial Contamination of Drinking Water and Human Health from Community Water Systems. Current Environmental Health Reports 2, 95-106.

Baghoth, S.A., Sharma, S.K., and Amy, G.L. (2011). Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation–emission matrices and PARAFAC. Water Research 45, 797-809.

Baker, A., Cumberland, S.A., Bradley, C., Buckley, C., and Bridgeman, J. (2015). To what extent can portable fluorescence spectroscopy be used in the real-time assessment of microbial water quality? Science of The Total Environment 532, 14-19.

Belal, T., Romdhane, K., Jean-Louis, B., Tahar, B., Eric, D., and Françoise, L. (2011). Optical fiber-based synchronous fluorescence spectroscopy for bacterial discrimination directly from colonies on agar plates. Analytical Methods 3, 133-143.

Bernhard, S. (1995a). Special techniques and applications. In Infrared and Raman Spectroscopy - Methods and Application, S. Bernhard. (Weinheim, Federal Republic of Germany: VCH Verlagsgesellschaft).

Bernhard, S. (1995b). General survey of vibrational spectroscopy. In Infrared and Raman Spectroscopy - Methods and Application, S. Bernhard. (Weinheim, Federal Republic of Germany: VCH Verlagsgesellschaft).

Bieroza, M., Baker, A., and Bridgeman, J. (2009). Exploratory analysis of excitation-emission matrix fluorescence spectra with self-organizing maps as a basis for determination of organic matter removal efficiency at water treatment works. Journal of Geophysical Research-Biogeosciences 114.

Borisover, M., Laor, Y., Parparov, A., Bukhanovsky, N., and Lado, M. (2009). Spatial and seasonal patterns of fluorescent organic matter in Lake Kinneret (Sea of Galilee) and its catchment basin. Water Research 43, 3104-3116.

Boyaci, I.H., Temiz, H.T., Geniş, H.E., Acar Soykut, E., Yazgan, N.N., Güven, B., Uysal, R.S., Bozkurt, A.G., İlaslan, K., Torun, O., et al. (2015). Dispersive and FT-Raman spectroscopic methods in food analysis. RSC Advances 5, 56606-56624.

Bridgeman, J., Baker, A., Brown, D., and Boxall, J.B. (2015). Portable LED fluorescence instrumentation for the rapid assessment of potable water quality. Science of The Total Environment 524-525, 338-346.

Cabral, J.P. (2010). Water microbiology. Bacterial pathogens and water. International Journal of Environmental Research and Public Health 7, 3657-3703.

Carstea, E.M., Bridgeman, J., Baker, A., and Reynolds, D.M. (2016). Fluorescence spectroscopy for wastewater monitoring: A review. Water Research 95, 205-219.

Cohen, E., Levy, G.J., and Borisover, M. (2014). Fluorescent components of organic matter in wastewater: efficacy and selectivity of the water treatment. Water Research 55, 323-334.

Collier, S.A., Stockman, L.J., Hicks, L.A., Garrison, L.E., Zhou, F.J., and Beach, M.J. (2012). Direct healthcare costs of selected diseases primarily or partially transmitted by water. Epidemiology and Infection 140, 2003-2013.

Costa, D., Bousseau, A., Thevenot, S., Dufour, X., Laland, C., Burucoa, C., and Castel, O. (2015). Nosocomial outbreak of *Pseudomonas aeruginosa* associated with a drinking water fountain. Journal of Hospital Infection 91, 271-274.

Cumberland, S., Bridgeman, J., Baker, A., Sterling, M., and Ward, D. (2012). Fluorescence spectroscopy as a tool for determining microbial quality in potable water applications. Environmental Technology 33, 687-693.

De Gelder, J., De Gussem, K., Vandenabeele, P., and Moens, L. (2007). Reference database of Raman spectra of biological molecules. Journal of Raman Spectroscopy 38, 1133-1147

Determann, S., Lobbes, J.M., Reuter, R., and Rullkötter, J. (1998). Ultraviolet fluorescence excitation and emission spectroscopy of marine algae and bacteria. Marine Chemistry 62, 137-156.

Eaton, A., Clesceri, L.S., Rice, E.W., Greenberg, A.E., and Franson, M. (2005). APHA: standard methods for the examination of water and wastewater. Centennial Edition, APHA, AWWA, WEF, Washington, DC.

Eberhardt, K., Stiebing C., Matthaus, C., Matthaus C., Schmitt, M., and Popp, J. (2015). Advantages and limitations of Raman spectroscopy for molecular diagnostics: an update. Expert Review of Molecular Diagnostics 15, 773-787

Edberg, S.C., Rice, E.W., Karlin, R.J., and Allen, M.J. (2000). *Escherichia coli*: the best biological drinking water indicator for public health protection. Symposium series (Society for Applied Microbiology), 106S-116S.

Elliott, S., Lead, J.R., and Baker, A. (2006). Characterisation of the fluorescence from freshwater, planktonic bacteria. Water Research 40, 2075-2083.

Eriksson, L., Andersson, P.L., Johansson, E., and Tysklind, M. (2006). Megavariate analysis of environmental QSAR data. Part I – A basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). Molecular Diversity 10, 169-186.

Fearn, T. (2002). Assessing Calibrations: SEP, RPD, RER and R2. NIR news 13, 12-13.

Fehrmann, A., Franz, M., Hoffmann, A., Rudzik, L., and Wüst, E. (1995). Dairy product analysis: identification of microorganisms by mid-infrared spectroscopy and determination of constituents by Raman spectroscopy. Journal of AOAC International 78, 1537-1542.

Fine, Z. Segev, M. and Lavie, R. (2007) Israel and its neighbors, 2002. In Israel People and Space: selected topics in Geography, Nahum-Levi N. (CET – Center for Educational Technology). (in Hebrew)

Frolich, L., Vaizel-Ohayon, D., and Fishbain, B. (2017). Prediction of Bacterial Contamination Outbursts in Water Wells through Sparse Coding. Scientific Reports 7, 799.

Geladi, P., and Kowalski, B.R. (1986). Partial least-squares regression: a tutorial. Analytica Chimica Acta 185, 1-17.

Gholizadeh, A., Boruvka, L., Saberioon, M., Kozák, J., Vašát, R., and Němeček, K. (2015). Comparing Different Data Preprocessing Methods for Monitoring Soil Heavy Metals Based on Soil Spectral Features. Soil and Water Research 4, 218–227

Heibati, M., Stedmon, C.A., Stenroth, K., Rauch, S., Toljander, J., Säve-Söderbergh, M., and Murphy, K.R. (2017). Assessment of drinking water quality at the tap using fluorescence spectroscopy. Water Research 125, 1-10.

Hennekinne, J.A., De Buyser, M.L., and Dragacci, S. (2012). Staphylococcus aureus and its food poisoning toxins: characterization and outbreak investigation. FEMS Microbiology Reviews 36, 815-836.

Ishii, S.K.L., and Boyer, T.H. (2012). Behavior of Reoccurring PARAFAC Components in Fluorescent Dissolved Organic Matter in Natural and Engineered Systems: A Critical Review. Environmental Science & Technology 46, 2006-2017.

Israeli Ministry of Health. (2013). Administrative directives for plan submissions to operate and monitor and drinking water treatment plant.

Israeli Ministry of Health. (2016). Directives for water sampling.

Israeli Ministry of Health. (2017). Public Health Regulations – Sanitary quality of drinking water and drinking water facilities.

Jarvis, R.M., Brooker, A., and Goodacre, R. (2006). Surface-enhanced Raman scattering for the rapid discrimination of bacteria. Faraday Discuss 132, 281-292; discussion 309-219.

Jean, N.B., Bougault C., and Simorre, J.P. (2016). The Structure of Bacterial Cell Wall. https://glycopedia.eu/The-Bacterial-Cell-Wall

Krafft, C., and Popp, J. (2015) The many facets of Raman spectroscopy for biomedical analysis. Analytical and Bioanalytical Chemistry 407, 699–717

Kriegsmann, J., Casadonte, R., Kriegsmann, K., Longuespée, R., and Kriegsmann, M. (2018). Mass spectrometry in pathology – Vision for a future workflow. Pathology - Research and Practice 214, 1057-1063

Lakowicz, J.R. (2006a). Introduction to Fluorescence. In Principles of Fluorescence Spectroscopy, J.R. Lakowicz, ed. (Maryland, USA: Springer). pp. 1-27

Lakowicz, J.R. (2006b). Protein Fluorescence. In Principles of Fluorescence Spectroscopy, J.R. Lakowicz, ed. (Maryland, USA: Springer). pp. 529-570

Landis, J.R., and Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159-174.

Leclerc, H., Schwartzbrod, L., and Dei-Cas, E. (2002). Microbial Agents Associated with Waterborne Diseases. Critical Reviews in Microbiology 28, 371-409.

Leinweber, D. (2007). Stupid Data Miner Tricks: Overfitting the S&P 500. The Journal of Investing 16, 15-22

Li, Y.S., and Church, J.S. (2014). Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. Journal of Food and Drug Analysis 22, 29-48.

Madigan, M.T., Martinko, J.M., and Brock, T.D. (2006). Brock biology of microorganisms, 11th edn (Upper Saddle River, N.J.: Pearson/Prentice Hall).

Manoharan, R., Ghiamati, E., Dalterio, R.A., Britton, K.A., Nelson, W.H., and Sperry, J.F. (1990). UV resonance Raman spectra of bacteria, bacterial spores, protoplasts and calcium dipicolinate. Journal of Microbiological Methods 11, 1-15.

Maquelin, K., Choo-Smith, L.P., van Vreeswijk, T., Endtz, H.P., Smith, B., Bennett, R., Bruining, H.A., and Puppels, G.J. (2000). Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. Analytical Chemistry 72, 12-19.

Meisel, S., Stöckel, S., Elschner, M., Melzer, F., Rösch, P., and Popp, J. (2012). Raman spectroscopy as a potential tool for detection of Brucella spp. in milk. Applied Environmental Microbiology 78, 5575-5583.

Meisel, S., Stöckel, S., Rösch, P., and Popp, J. (2014). Identification of meat-associated pathogens via Raman microspectroscopy. Food Microbiology 38, 36-43.

Mello, C., Ribeiro, D., Novaes, F., and Popp, R.J. (2005). Rapid differentiation among bacteria that cause gastroenteritis by use of low-resolution Raman spectroscopy and PLS discriminant analysis. Analytical and Bioanalytical Chemistry 383, 701-706.

Mizrach, A., Schmilovitch, Z., Korotic, R., Irudayaraj, J., and Shapira, R. (2007). Yeast Detection in Apple Juice Using Raman Spectroscopy and Chemometric Methods. Transactions of the American Society of Agricultural and Biological Engineers 6, 2143-2149

Monzó, J., Insua, I., Fernandez-Trillo, F., and Rodriguez, P. (2015). Fundamentals, achievements and challenges in the electrochemical sensing of pathogens. Analyst, 140, 7116-7128

Naumann, D., Keller, S., Helm, D., Schultz, C., and Schrader, B. (1995). FT-IR spectroscopy and FT-Raman spectroscopy are powerful analytical tools for the non-invasive characterization of intact microbial cells. Journal of Molecular Structure 347, 399-405.

Nebbioso, A., and Piccolo, A. (2012). Molecular characterization of dissolved organic matter (DOM): a critical review. Analytical and Bioanalytical Chemistry 405, 109-124

Nicolaou, N., Xu, Y., and Goodacre, R. (2011). Fourier transform infrared and Raman spectroscopies for the rapid detection, enumeration, and growth interaction of the bacteria Staphylococcus aureus and Lactococcus lactis ssp. cremoris in milk. Analytical Chemistry 83, 5681-5687.

Pahlow, S., Meisel, S., Cialla-May, D., Weber, K., Rösch, P., and Popp, J. (2015). Isolation and identification of bacteria by means of Raman spectroscopy. Advanced Drug Delivery Reviews 89, 105-120.

Pazos-Perez, N., Pazos, E., Catala, C., Mir-Simon, B., Gómez-de Pedro, S., Sagales, J., Villanueva, C., Vila, J., Soriano, A., García de Abajo, F.J., et al. (2016). Ultrasensitive multiplex optical quantification of bacteria in large samples of biofluids. Scientific Reports 6, 29014.

Pubchem Database, National Center for Bioinformatics. (2018). Tyrosine CID=6057, Phenylalanine CID=6305, Tryptophan CID=6140 (PubChem Compound Database).

Premasiri, W.R., Chen, Y., Williamson, P.M., Bandarage, D.C., Pyles, C., and Ziegler, L.D. (2017). Rapid urinary tract infection diagnostics by surface-enhanced Raman spectroscopy (SERS): identification and antibiotic susceptibilities. Analytical and Bioanalytical Chemistry 409, 3043-3054.

Premasiri, W.R., Moir, D.T., Klempner, M.S., Krieger, N., Jones, G., and Ziegler, L.D. (2005). Characterization of the Surface Enhanced Raman Scattering (SERS) of Bacteria. The Journal of Physical Chemistry B 109, 312-320.

Reynolds, D.M. (2003). Rapid and direct determination of tryptophan in water using synchronous fluorescence spectroscopy. Water Research 37, 3055-3060

Rodriguez, S.B., Thornton, M.A., and Thornton, R.J. (2013). Raman Spectroscopy and Chemometrics for Identification and Strain Discrimination of the Wine Spoilage Yeasts Saccharomyces cerevisiae, Zygosaccharomyces bailii, and Brettanomyces bruxellensis. Applied and Environmental Microbiology 79, 6264-6270.

Rompré, A., Servais, P., Baudart, J., de-Roubin, M.-R., and Laurent, P. (2002). Detection and enumeration of coliforms in drinking water: current methods and emerging approaches. Journal of Microbiological Methods 49, 31-54.

Rösch, P., Harz, M., Schmitt, M., Peschke, K.D., Ronneberger, O., Burkhardt, H., Motzkus, H.W., Lankers, M., Hofer, S., Thiele, H. (2005). Chemotaxonomic identification of single bacteria by micro-

Raman spectroscopy: application to clean-room-relevant biological contaminations. Applied Environmental Microbiology 71, 1626-1637.

Schmilovitch, Z., Mizrach, A., Alchanatis, V., Kritzman, G., Korotic, R., Irudayaraj, J., and Debroy, C. (2005). Detection of bacteria with low-resolution Raman spectroscopy. Transactions of the American Society of Agricultural Engineers 48, 1843-1850.

Simelane, K.S. (2013). Application of Fluorescence Spectroscopy for Monitoring Microbial Contamination of Drinking Water. MSc Thesis: Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem.

Socrates, G. (2001). Infrared and Raman Characteristic Group Frequencies (Chichester, England: John Wiley & Sons Ltd.).

Sorensen, J.P.R., Baker, A., Cumberland, S.A., Lapworth, D.J., MacDonald, A.M., Pedley, S., Taylor, R.G., and Ward, J.S.T. (2018a). Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values. Science of The Total Environment 622-623, 1250-1257.

Sorensen, J.P.R., Lapworth, D.J., Marchant, B.P., Nkhuwa, D.C.W., Pedley, S., Stuart, M.E., Bell, R.A., Chirwa, M., Kabika, J., Liemisa, M., et al. (2015). *In-situ* tryptophan-like fluorescence: A real-time indicator of faecal contamination in drinking water supplies. Water Research 81, 38-46.

Sorensen, J.P.R., Vivanco, A., Ascott, M.J., Gooddy, D.C., Lapworth, D.J., Read, D.S., Rushworth, C.M., Bucknall, J., Herbert, K., Karapanos, I., et al. (2018b). Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water. Water Research 137, 301-309.

Stedmon, C.A., Thomas, D.N., Granskog, M., Kaartokallio, H., Papadimitriou, S., and Kuosa, H. (2007). Characteristics of Dissolved Organic Matter in Baltic Coastal Sea Ice: Allochthonous or Autochthonous Origins? Environmental Science & Technology 41, 7273-7279.

Stöckel, S., Kirchhoff, J., Neugebauer, U., Röscha, P., and Popp, J. (2015). The application of Raman spectroscopy for the detection and identification of microorganisms Journal of Raman Spectroscopy, 89-109.

Stöckel, S., Schumacher, W., Meisel, S., Elschner, M., Rösch, P., and Popp, J. (2010). Raman spectroscopy-compatible inactivation method for pathogenic endospores. Applied Environmental Microbiology 76, 2895-2907.

Sundaram, J., Park, B., Hinton, A., Lawrence, K.C., and Kwon, Y. (2013a). Detection and differentiation of Salmonella serotypes using surface enhanced Raman scattering (SERS) technique. Journal of Food Measurement and Characterization 7, 1-12.

Sundaram, J., Park, B., Kwon, Y., and Lawrence, K.C. (2013b). Surface enhanced Raman scattering (SERS) with biopolymer encapsulated silver nanosubstrates for rapid detection of foodborne pathogens. International Journal of Food Microbiology 167, 67-73.

Tedetti, M., Joffre, P., and Goutx, M. (2013). Development of a field-portable fluorometer based on deep ultraviolet LEDs for the detection of phenanthrene- and tryptophan-like compounds in natural waters. Sensors and Actuators B: Chemical 182, 416-423.

Wold, S., Sjöström, M. and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58, 109-130.

Walsh, J.D., Hyman, J.M., Borzhemskaya, L., Bowen, A., McKellar, C., Ullery, M., Mathias, E., Ronsick, C., Link, J., Wilson, M. (2013). Rapid intrinsic fluorescence method for direct identification of pathogens in blood cultures. MBio 4, e00865-00813.

Wang, J., Xie, X., Feng, J., Chen, J.C., Du, X.J., Luo, J., Lu, X., and Wang, S. (2015). Rapid detection of Listeria monocytogenes in milk using confocal micro-Raman spectroscopy and chemometric analysis. International Journal of Food Microbiology 204, 66-74.

Willemse-Erix, D.F., Scholtes-Timmerman, M.J., Jachtenberg, J.W., van Leeuwen, W.B., Horst-Kreft, D., Bakker Schut, T.C., Deurenberg, R.H., Puppels, G.J., van Belkum, A., Vos, M.C. (2009). Optical fingerprinting in bacterial epidemiology: Raman spectroscopy as a real-time typing method. Journal of Clinical Microbiology 47, 652-659.

Williams, A.C., and Edwards, H.G.M. (1994). Fourier transform Raman spectroscopy of bacterial cell walls. Journal of Raman Spectroscopy 25, 673-677.

Williams, P., and Norris, K.H. (2001). Near-infrared technology in the agricultural and food industries. 2nd ed. (Minnesota, USA: American Association of Cereal Chemists, Inc.) p. 21.

Yamashita, Y., Panton, A., Mahaffey, C., and Jaffé, R. (2011). Assessing the spatial and temporal variability of dissolved organic matter in Liverpool Bay using excitation–emission matrix fluorescence and parallel factor analysis. Ocean Dynamics 61, 569-579.

Yang, L., Hur, J., and Zhuang, W. (2015). Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: a review. Environmental Science and Pollution Research 22, 6500-6510.

Zeiri, L., Bronk, B.V., Shabtai, Y., Eichler, J., and Efrima, S. (2004). Surface-enhanced Raman spectroscopy as a tool for probing specific biochemical components in bacteria. Appllied Spectroscopy 58, 33-40.

Zeiri, L., and Efrima, S. (2005). Surface-enhanced Raman spectroscopy of bacteria: the effect of excitation wavelength and chemical modification of the colloidal milieu. Journal of Raman Spectroscopy 36, 667-675.

Zeiri, L., and Efrima, S. (2010). Surface-Enhanced Raman Scattering (SERS) of Microorganisms. Israel Journal of Chemistry 46, 337-346.

Zhou, H., Yang, D., Ivleva, N.P., Mircescu, N.E., Niessner, R., and Haisch, C. (2014). SERS detection of bacteria in water by in situ coating with Ag nanoparticles. Analytical Chemistry 86, 1525-1533.

**תקציר**

זיהום חיידקי של מי שתייה מהווה סיכון לבריאות הציבור. מדי שנה, מחלות זיהומיות הנובעות מזיהום מים פוגעות באלפי אנשים במדינות מפותחות, ואפילו מיליונים עשויים להפגע במדינות מתפתחות. בעוד רשויות המים מפעילות מנגנונים רבים כדי להבטיח את בטיחותם של מי השתייה, השיטות לבחינת איכותם התברואתית של מים הן איטיות מדי, דורשות בין 12-72 שעות כדי לייצר תוצאות, ומצריכות שימוש בכוח אדם רב. מספר שיטות מולקולריות מודרניות, כגון היברידיזציית דנ״א, Enzyme-Linked Immunosorbent Assay (ELISA) וספקטרומטריית מסות (MS) הוצעו כחלופה לשיטות המיקרוביולוגיות הסטנדרטיות. עם זאת, שיטות אלה אינן ישימות על בסיס יומיומי בתעשייה היות והן דורשות כוח אדם מיומן, זמן רב להפקת תוצאות, עלויות השימוש בהן גבוהות או כי הן אינן מדויקות דיין. לעומת זאת, ספקטרוסקופיית ראמאן ופלואורסנציה עשויות לספק בסיס לשיטה מהירה ולא יקרה לזיהוי מהיר חיידקים במי שתייה.

במחקר זה, נבדקה יכולן של ספרקטרוסקופיית ראמאן ברזולוציה נמוכה וספקטרוסקופיית פלורסנציה להערכת יכולת הגילוי של חיידקים במערכת מודל ובמי תהום המיועדים לשתייה. חיידקי *Escherichia coli*, *Bacillus subtilis* ו- *Pseudomonas aeruginosa* שימשו להערכת סף הגילוי בשתי השיטות על ידי מדידת ספקטרום ההארה באמצעות ספקטרומטר ראמאן או ספקטרופלואורומטר. הנתונים המורכבים שנאספו נותחו באמצעות רגרסיית Partial Least Squares (PLS) ו-Partial Least Squares Discriminant Analysis (PLS-DA). בהמשך, דגימות מי שתייה, שנאספו מבארות מי תהום בצפון הארץ, נבחנו באמצעות ספקטרוסקופיית פלורסנציה על מנת להעריך את יכולת השיטה לגילוי זיהום חיידקי אמיתי במי שתייה.

נמצא ששיטת ספקטרוסקופיית הראמאן ברזולוציה נמוכה רגישה רק באופן חלקי, בעלת יכולת גילוי חיידקים בריכוז מעל $10^8$ ו- $10^2$ CFU/ml לחיידקי *E. coli* ו- *B. subtilis* בהתאמה. בניסוי ספקטרוסקופיית פלורסנציה נמצא שרגישות הגילוי גדלה באופן משמעותי כאשר מנתחים את כלל נתוני מפת העירור-הארה התלת מימדית (Excitation-Emission Matrix, EEM) לעומת ניתוח של צמד יחיד של אורכי גל עירור-הארה. בניסוי מבוקר, נמצא ששימוש באלגוריתם PLS המנתח את כלל ה-EEM בטווח 210-400 ננומטר עירור ו-220-450 ננומטר הארה אפשר שיפור סף גילוי החיידקים מ-$10^5$ CFU/ml ל-10 CFU/ml.

היות וספקטרוסקופיית פלורסנציה הראתה פוטנציאל גבוה לגילוי וכימות חיידקים במים, השיטה נבדקה על מי שתייה. במהלך דיגום מי תהום לאורך שנה, לא נמצא זיהום מיקרוביאלי מהותי של המים. דגימה יחידה מתוך 99 הדגימות שנאספו הייתה מזוהמת, ונמצאו בה ערך של 2,900 CFU/ml חיידקים הטרוטרופיים (לפי שיטת Heterotrophic plate counts) ו-2 CFU/ml של חיידקי *E. coli*. זיהום זה זוהה בבירור באמצעות ספקטרוסקופיית פלורסנציה. הכלרת ו/או סינון המים הפחיתו את החתימה הספקטראלית של זיהום המים כמעט לחלוטין, עובדה המצביעה על הקשר שבין הזיהום החיידקי לפלורסנציה. בנוסף, מודלים לאבחנה בין החיידקים המבוססים על נתוני ספקטרוסקופיית ראמאן או פלורסנציה, אפשרו אבחנה בין חיידקי *E. coli* ו-*B. subtills* בריכוזים גבוהים, ממצא המרמז על שוני בטביעת האצבע הספקטרלית בין המינים.

לסיכום, מחקר זה הוא הראשון המדגים את היכולת של ספקטרוסקופיית פלורסנציה לגילוי חיידקים בריכוזים של 10 CFU/ml במים. המחקר מצביע על פוטנציאל גבוה לשימוש בספקטרוסקופיית פלורסנציה לגילוי מהיר ואמין של חיידקים במי תהום ומי שתייה.

# זיהוי וכימות מהיר של חיידקים במי שתייה באמצעות ספקטרוסקופיית ראמאן ופלורסנציה

עבודת גמר

מוגשת לפקולטה לחקלאות, מזון וסביבה ע"ש רוברט ה. סמית

האוניברסיטה העברית בירושלים

לשם קבלת תואר 'מוסמך למדעי החקלאות'

מוגש על ידי

אמיר נקר

אוקטובר 2018