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

August, 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

August, 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, Bet Dagan

Dr. Mikhail Borisover

Institute of Soil, Water and Environmental Sciences,

Department Soil Chemistry, Plant Nutrition and Microbiology,

Agriculture Research Organization,

The Volcani center, Bet Dagan

and

Prof. Shlomo Sela (Saldinger)

Institute of Postharvest and Food Sciences,

Department of Food Quality and Safety,

Agriculture Research Organization,

The Volcani center, Bet Dagan

1. Acknowledgment

I would like to thank all 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 people from different fields of science, and only through their open-mindedness, curiosity, flexibility and relentlessness we managed to do this research.

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

I thank our lab members for the enjoyable work, counselling and good coffee, to Yulia Kroupitsky, Lilya Karasic, Jitendra Keshri, Orian Dayan and Nadia Bukhanovsky. And to my friends next door: Orr Shapiro, 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.

2. Abstract

Bacterial contamination of drinking water is a considerable concern for public health. Annually, thousands of people are affected by waterborne diseases in developed countries, and millions in under-developed countries. While water authorities go through many processes to assure the safety of drinking water, the methods for testing for water sanitation are too slow, taking between 12-72 hours to produce results. Many modern molecular methods have been suggested to replace the standards, including DNA hybridization, Enzyme-Linked Immunosorbent Assay (ELISA) and mass spectrometry. These however, remain beyond the reach of industrial scale testing because they are either too costly or too inaccurate. Light spectroscopy approaches, such as Raman and fluorescence spectroscopy are modern, fast analytical methods which may be applied for the detection of bacteria in drinking water. These methods are based on the application of a light source to a sample and the measurements of the effect of the sample on the emitted light.

In this study, low resolution Raman spectroscopy and fluorescence spectroscopy were used to try and detect bacteria, first in a model system and later in real drinking water. *E. coli, B. subtilis* and *P. aeruginosa* were used to assess the different methods' detection thresholds, by measuring the light spectra either using a Raman spectrometer or a spectrofluorometer. Partial least squares analysis was used to analyze the complex spectra, since visually no distinctive peaks were found. Subsequently, real drinking water from wells in the north of Israel was scanned using fluorescence spectrometry to determine the method's ability to detect contamination of real drinking water.

It was found that low resolution Raman spectroscopy is not a very valuable tool for the detection of bacteria in water, only detecting bacteria at 10^8 and 10^4 CFU/ml for *E. coli* and *B. subtilis* respectively. Fluorescence spectroscopy on the other hand, showed great potential for detection and quantification of bacteria in a water, detecting bacteria at a concentration as low as 10 CFU/ml for some species. Furthermore, a classification model has been able to accurately differentiate between *E. coli, B. subtilis* and *P. aeruginosa* at high concentrations, indicating different spectral fingerprints between species. In real drinking water, the dataset was found to have too few contamination events, with only one sample officially contaminated over a year-long sampling period. That sample could, however, be distinguished and detected easily using fluorescence spectroscopy, and the treatment of the sample by chlorination or filtration reduced the spectral fingerprint completely, indicating the connection between microbial contamination and fluorescence.

Both in a set of controlled lab experiments and in real drinking water, fluorescence spectroscopy has shown an ability to detect bacteria at a concentration of 1,000 CFU/ml, which is the required industrial standard. Furthermore, the study demonstrates the increase in sensitivity which can be acquired by analyzing the complex 3-dimensional spectrum of fluorescence spectroscopy rather than single-wavelength "peak picking". In a controlled experiment, a Partial Least Squares (PLS) algorithm applied to an entire excitation-emission map (EEM) in the range of 210-400 nm excitation and 220-450 nm emission can increase sensitivity from 10^5 CFU/ml to as low as 10 CFU/ml. This method also enables the differentiation of different species according to their

spectral fingerprint. This study illustrates the high potential of fluorescence spectroscopy for the fast and accurate detection of bacteria in drinking water.

3. Table of contents

4. List of Abbreviations

CFU – Colony Forming Units

EEM – Excitation Emission Map

HPC – Heterotrophic Plate Counts

OD – Optical Density

PARAFAC – 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 put into maintaining drinking water safe, we still face many bacterial water contaminations, from 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 microbial contamination of water, resulting in 970,000 dollars of damages annually (collier (Collier et al., 2012). The source of these contaminations may be direct contamination of water sources, pipelines or home and institutional taps or failures in the sanitation system in bottling and water processing plants.

Water treatment plants use different methods to prevent microbial contamination of drinking water. The water may be filtered by several types of filters, chemically sanitized (by chlorination or ozonation) and sometimes even reverse osmosis is used to insure water safety. Further down the supply chain the water is sometimes re-filtered to get rid of any contaminants that arrive from piping contaminators and in some homes and institutions an in-house treatment is done using micro-filters, active carbon and Ultra Violet (UV) light sanitation. In recent years we have seen more of these in-house sanitation approaches in home "water bars". In spite of all these efforts, pathogenic bacteria sometimes reach consumers. In these cases, the local authorities would often redirect water distribution until the contamination is either disinfected or naturally subside (2013; 2016; 2017; Ashbolt, 2015).

Due to the dangers 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, saving a lot of money and improving public health. The water industry uses many different methods to detect contaminations, such as turbidity measurement and culture-based selective assays that detect coliforms, fecal coliforms and fecal streptococci. These tests are done throughout the water treatment process and in many control points in pipelines. This main issue with these tests is that the time to get results is long, which means that in practice – contaminated water reaches consumers. Furthermore, the standard tests are usually not specific, or are specific only to indicators (such as coliforms) and cannot distinctly detect pathogens such as *Legionella* and *Campylobacter*. Another standard test is the heterotrophic plate count test, which quantifies the general microbial load in the water without discriminating harmful or harmless bacteria (Edberg et al., 2000; Rompré et al., 2002).

Common microbiological lab methods for the detection of bacteria generally include culturing samples in selective media, in optimal conditions, followed by colony counts or turbidity measurements (Edberg et al., 2000). Colony counts are an indication of the number of bacteria that is measured by the normalized amount of bacterial colonies on an agar plate. The number is usually normalized to colony-forming-units per ml (CFU/ml). These methods are problematic because the formation of colonies requires a long time, between 12 hours and a week. Turbidity measurements can be done without culturing but they allow rough estimates, efficient only in exceptional cases where the microbial load is as high as 10^6 CFU/ml or more. These disadvantages are one of the greatest challenges in the food and water industries, and cause a

threat to public health (Hennekinne et al., 2012; Leclerc et al., 2002). Moreover, current methods require a large team of trained personnel and lots of resources (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 hundreds of thousands that pass through the system. The tests are done periodically and the majority of water is not tested. Since water contamination is unpredictable and sporadic (Cabral, 2010; Frolich et al., 2017), a failure in testing in the water industry can be identified which may undermine public health. This failure is well known but presently no high quality and affordable alternative is commonly used (Rompré et al., 2002; Willemse-Erix et al., 2009).

Several modern methods have been suggested to replace the standard. 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. More recent methods, based on nucleic acid microarray or Enzyme-Linked Immunosorbent Assay (ELISA) technology can provide highly specific results faster (within hours). However, these methods rely on expensive reagents, instrumentation and highly trained personnel and are thus not applicable to most industrial uses. Spectral methods, based on mass spectrometry have also been proposed, especially the Matrix-assisted laser desorption/ionization (MALDI-TOF) approach. These utilize high energy lasers to ionize samples, causing molecular structures to breakdown and then measure the 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. Further, the method requires culturing (albeit shorter than standard) which takes time and is limited to specific types of bacteria (Stöckel et al., 2015). There still remains a need for a rapid, accurate and quantitative method for the detection of bacteria in drinking water that is cheap and reliable.

1.2. Light Spectroscopy

Light spectroscopy approaches, such as Raman, infra-red and fluorescence spectroscopy have been suggested by numerous experts as alternatives to the standard methods (Pahlow et al., 2015; Stöckel et al., 2015). These different approaches are based on the interaction of light with the chemicals that comprise bacteria, and cause the light to change in measureable ways. While these interactions are relatively rare and complex, modern instruments with sensitive spectrometers can detect the subtle changes to the light's wavelength and characterize the light after the interaction with a sample; this is called the sample's spectrum. Furthermore, theoretically there is a correlation between altered light intensity and bacterial concentration, according to Beer's law (Williams and Norris, 2001) page 21, since the more bacteria are in a sample, more photons will interact with them. However, 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 designed for rapid detection of bacteria according to their spectra. The main advantages of light spectroscopy are that it requires very little to no samples preparation, no expensive reagents and the instrumentation can be relatively cheap (depending on the quality of the instruments). Further, once a dependable 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.

1.3. Raman spectroscopy

1.3.1. Scientific background

Raman spectroscopy is a modern analytical tool, which can deliver rapid results at a relatively low cost. The method is based on Raman Shift, a phenomenon first described by CV Raman in 1928. When photons of a specific wavelength interact with a molecule some of the photons go through 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 (Bernhard, 1995). The shift is different between molecular structures and bonds, since each vibrational mode can transfer a different amount of energy. Consequently, the Raman effect is specific and a molecular "fingerprint" can be assembled (Bernhard, 1995). Raman shift is dependent upon the Polarizability of a substance, i.e. the flexibility of its electron cloud. This is because the energy is transferred via the electron cloud, and in a low polarizability substance the photons cannot transfer the energy. 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 light spectrometer which measures the light's wavelength after it has gone through a sample and converts it into a digital signal for computer processing (Krafft and Popp; Stöckel et al., 2015). The digital spectrum displays the intensity of light at each wavelength that is longer than the original transmitted light (which is cutoff from the signal). This light had undergone Raman Shift (Figure 1). The wavelength can be displayed as absolute wavelength in nanometers, but is usually converted into energy shift in units of cm^{-1}.

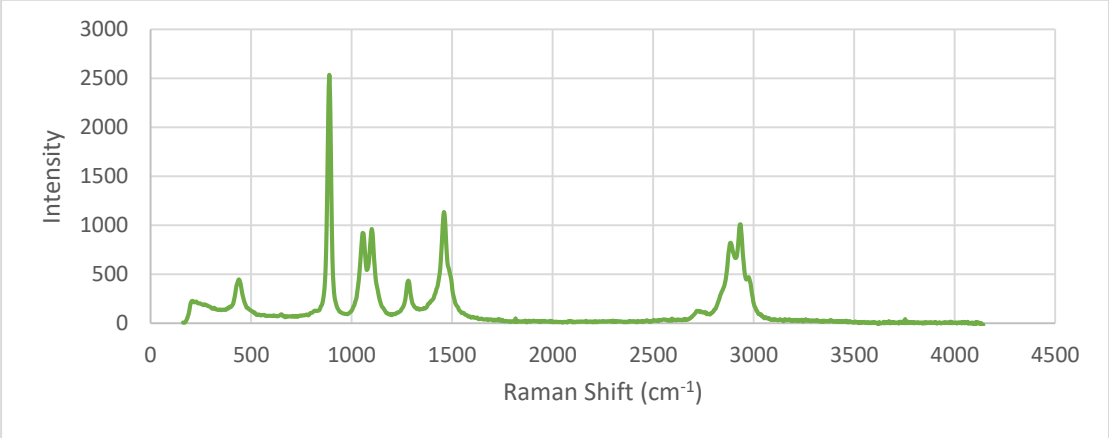

Figure 1 Spectral fingerprint of ethanol.

1.3.2. Uses of Raman spectroscopy for the detection of bacteria

Different studies focused on different aspects of applying Raman spectroscopy for bacteria detection. In the 1990s, some of the first attempts to describe bacteria and their Raman fingerprints have managed to accurately differentiate different strains 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* and others using long laser exposure times and very high concentrations of bacteria. Later, Zeiri et al (L and S, 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. In another paper, Premasiri et al (Premasiri et al., 2017) have shown that the Raman spectra of bacteria is also related to guanine (which is 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 (Meisel et al., 2012) have demonstrated detection of *Brucella, Escherichia and Yersinia* spp. in milk samples. A database of over 2,000 spectra had to be collected, and each sample was 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, scanning was done after washing with water. A similar study (Meisel et al., 2014) demonstrated an 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. Similar works by Wang (Wang et al., 2015), Nicolaou (Nicolaou et al., 2011) and others have 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).

A different approach was to try and quantify the bacteria in a sample. The challenge with detecting and 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 effect. The same approach was used by Sundaram et al (Sundaram et al., 2013a; Sundaram et al., 2013b) to increase their signal 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, the expertise required to construct them and the need to replace them regularly make this method too expensive for industrial scale testing.

1.3.3. Low resolution Raman spectroscopy

While most studies have applied microscopy based approaches or surface enhancement using gold and silver nanoparticles, few have tried to detect bacteria using low resolution tools. Schmilovitch et al (Schmilovitch et al., 2005) detected *Erwinia* and *Clavibacter* species (known

plant pathogens) using a simple 785 nm laser and a spectrometer, without any light enhancement techniques. They managed to detect the bacteria at concentrations as low as 100 cells/ml, using a Partial Least Squares (PLS) prediction model, with an accuracy >90%. This work can be regarded as a proof of concept for detection of bacteria using low resolution Raman spectroscopy that can be applied for the food and water industries. Other works by Mello (Mello et al., 2005) and Mizrach (Mizrach et al., 2007) have shown similar results in enteric bacteria and yeast respectively. All these discoveries were done using low resolution Raman, with advanced algorithms to prepare and analyze large datasets. These algorithms (such as PLS, Support Vector Machine and others) elucidate the spectra and consequently improve detection limits and resolution, without the need for expensive instrumentation or sample preparation. Such an approach, when properly calibrated, could be adopted in the food and water industry.

1.4. Fluorescence spectroscopy

1.4.1. Scientific background

Fluorescence is a natural phenomenon that 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 state). 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. The electron returns to the ground state by naturally emitting a photon. However, the emitted photon has a lower energy level than that of the exciting photon due to the loss of energy by internal interaction (Lakowicz, 2006) (Figure 2).

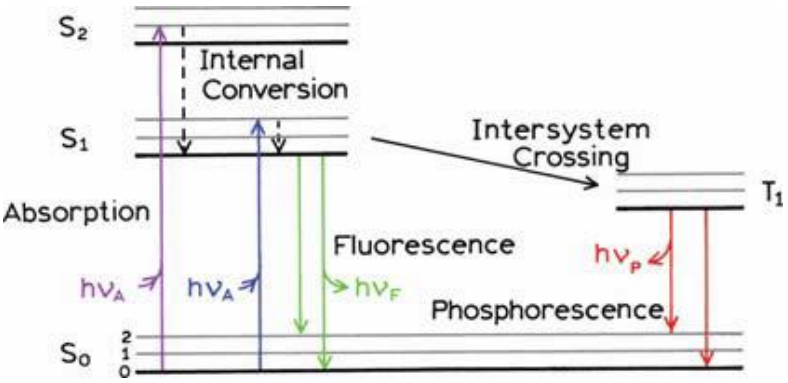

Figure 2. Jablonski diagram illustrating fluorescence

A measurement of the fluorescence effect (i.e. of the difference between the excitation and emission wavelengths of light) can be applied for describing the chemical nature of a sample. This approach is widely used for the analysis of soluble organic matters in natural 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 quantification and identification of humic and proteinaceous substances. In proteinaceous substances, the major fluorescence effect is derived from the aromatic amino acids: Tyrosine,

Tryptophan and Phenyl-alanine (Figure 3). These proteinaceous substances can be 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 of these structures is shifted according to the different combinations and structures in which the amino acids are organized (R., 2006) lako.

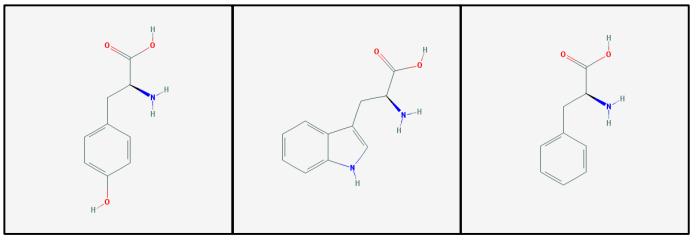

Figure 3. Molecular structures of (from left to right) tyrosine, tryptophan and phenylalanine (Information, 2018)

1.4.2. Uses of fluorescence for the detection of bacteria

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

This has led to studies for the development of a method to detect bacteria in drinking water. Two major approaches are used in the literature for measuring protein-like fluorescence in drinking

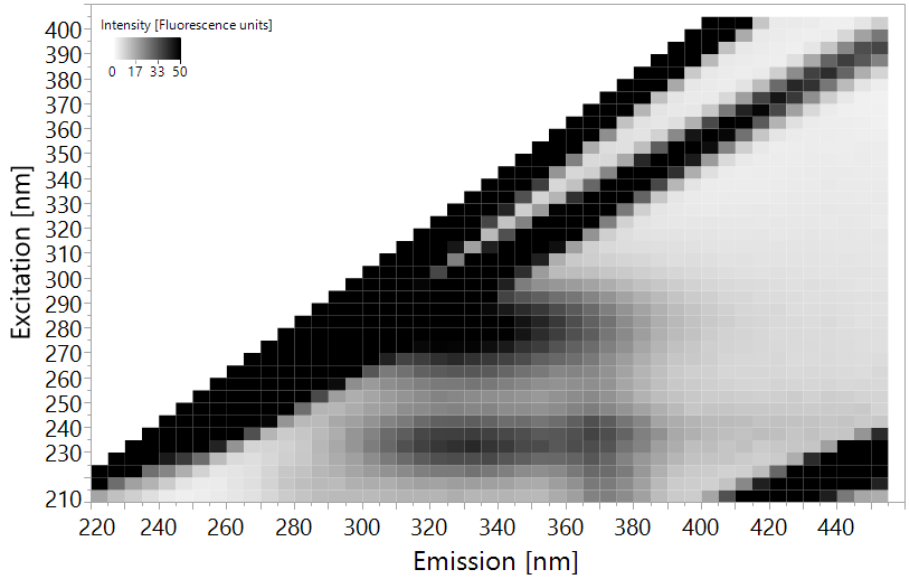

Figure 4. EEM of E. coli suspended in distilled water at a concentration of 10^5 CFU/ml. The intensity has been cutoff at 50 fluorescence units to enhance visibility.

water: The "Peak-Picking" approach and multi-spectral approaches based on Excitation-Emission Maps (EEMs). In the peak picking or single wavelength approach, a sample is illuminated by monochromatic light at a specific wavelength, and the emission of light at a different wavelength is measured. For example, in order to measure tryptophan fluorescence Tedetti et al (Tedetti et al., 2013) used excitation at 280 nm and measured emission at 340 nm. Conversely, the Excitation Emission Map (EEM) based approach uses a range of different excitation and emission wavelengths in order to study a sample. By measuring the emission of light at each pair of wavelengths this method can establish a clear picture of the different organic elements in the sample (Figure 4).

The clear tradeoff between the different approaches is in resolution over costs. While the single wavelength approach is cheaper and easier to analyze, EEMs yield more information at higher operational and computational costs.

Sorensen et al, in 2015 (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 taken from boreholes and shallow wells in Zambia. This study has shown the ability of this fairly simple instrument to detect an *E. coli* contamination of >1 CFU/100ml of water, which was attributed to ~3.5 ppb of tryptophan. A follow up study, by Baker et al (Baker et al., 2015) used a similar instrument to detect *E. coli* contaminations in low quality catchment water in South Africa rural areas. This experiment managed to accurately detect *E. coli* contamination at a concentration of >100 CFU/100ml with a tryptophan equivalence of 3 ppb. 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 3rd world countries such as India, Malawi and Zambia, with the tryptophan equivalence ranging between 1.5-3.5 ppb of tryptophan as a measure of contamination (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 280nm-360nm (+-10nm) can be used for the detection of bacteria in low quality water. It is Important however to note that water with >1 CFU/100ml of *E. coli* is considered very low quality in most western countries. The Israeli, European Union and United State regulations all prohibit consumption of water with this level of contamination, and the phenomena is relatively rare (Allen et al., 2004).

A similar study to the one done in the UK, where the water quality is regarded very good, used a lab-constructed fluorometer with a LED excitation light source at 280 nm and an ability to measure emission at 350 nm. In this study, measurements were taken throughout the supply chain (at the source, in checkpoint in the pipeline and at home taps) of both fluorescence and standard microbial indictors such as *E. coli* contamination and heterotrophic plate counts (HPC). No correlation was found between the microbial indicators and the fluorescence intensity at the tryptophan peak (Sorensen et al., 2018b). This finding implies that the single-wavelength method may not be sensitive enough for high quality water.

In 2018, Heibati et al investigated a multispectral approach to detecting bacteria in drinking water in Sweden. While Sorensen, Baker and Bridgman all used only the tryptophan peak of excitation-emission 280-360, Heibati et al used the full spectra at excitation 220-600 nm and emissions at 240-800 nm (3 nm resolution) to construct high resolution EEMs. The EEMs were analyzed by complex multivariate statistical methods in order to analyze the large dataset of EEMs, this included Partial Least Squares (PLS) and Parallel Factor Analysis (PARAFAC). No severe microbial contamination, and had no *E. coli* or coliforms were found throughout the study's measuring period. The only indicator which was applicable for analysis was HPC, and only weak correlations were observed between the EEM data and HPC bacterial abundances. This may serve as an indication that at very low concentrations (<150 CFU/ml of heterotrophic bacteria) it is difficult to detect bacteria using the fluorescence measurement approach (Heibati et al., 2017).

These studies demonstrate a possibility of using fluorescence spectroscopy for the detection of bacteria in water. This method, like Raman spectroscopy is relatively cheap to apply and requires no sample preparation. It appears though, that simple single wavelength measurements are not sensitive enough to enable detection of low levels of bacteria, as required in modern industrialized countries such as Israel. For this purpose, perhaps the multispectral EEM approach will be better suited. Such an approach, when properly calibrated, could be adopted in the food and water industry.

1.5. Multivariate data analysis

Because of the complex nature of EEMs and Raman spectra, they are always analysed using multivariate statistics. Partial Least Squares (PLS) regression or PLS discriminant analysis is a particularly suitable method for predicting bacterial concentration using spectral data since it performs well even when the number of predictor variables is high and some variables correlate with each other (Heibati et al., 2017; Schmilovitch et al., 2005). In the context of drinking water monitoring, it would be desirable to predict bacteria concentration (Y) from easily-obtained measurements of Raman spectra or fluorescence EEMs, which are comprised of hundreds of wavelengths (X_n). PLS algorithms find the correlation between these different variables (X_n), and then correlates those to the dependant variable (Y). 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

Because of the nature of the analysis it will almost always manage to properly "predict" Y. However, prediction models might have a bias called overfitting, where your model is only relevant to your dataset. In order to overcome this, the data is split into calibration and validation sets. The model is first trained on the calibration or training set, and is then tested for accuracy and validity on the validation set. This approach is called cross validation and enables calibration of robust prediction models (Geladi and Kowalski, 1986; Gholizadeh et al., 2015).

The need for a fast, accurate method to enumerate bacteria in water is of high priority for public health (Collier et al., 2012). Spectroscopy based methods have been suggested by several experts (refs), however, most methods are either too expensive, not sensitive enough or have not been tested in real situations. This study aims to examine two spectroscopy methods: Low resolution Raman spectroscopy and fluorescence spectroscopy.

2. Research objectives

The specific aims of this study are:

a) Test the detection thresholds of low resolution Raman spectroscopy for detecting bacteria in water
b) Test and compare the detection threshold of fluorescence spectroscopy, both single wavelength and full EEM analysis, for detecting bacteria in water
c) Compare both methods and outline a recommendation for further studies
d) Assess the ability of spectroscopy for detecting bacteria in real drinking water samples
e) Evaluate the ability of both low resolution Raman and fluorescence spectroscopy at differentiation between different species of model bacteria

3. Materials and methods

3.1. Materials used

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

Table 1. Material used in the study

Material	Maker	Origin (City, Country)
Ethanol	Gadot	Netanya, Israel
HPLC-grade Water	Biolab	Jerusalem, Israel
HPLC-grade Water	Merck	Darmstadt, Germany
Agar	Difco	Sparks, USA
LB Broth	Difco	Sparks, USA
L-tryptophan	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 lab stocks at the Volcani Center.

Bacteria were transferred from pure cultures stored at -20 0C (with 20% glycerol solution) and were grown on LB agar medium (Difco, Sparks, MD, USA) by incubating at 37°C overnight. Several well isolated colonies were harvested, suspended in a 10 or 15 ml of LB broth (Difco, Sparks, MD, USA) and incubated at 37°C with shaking overnight to prepare a starter culture. The stocks were later used in the sample preparation procedure in different ways for different experiments.

3.3. Optical density to bacterial concentration calibration

In order to associate optical density with microbial concentration, liquid bacterial cultures were diluted to optical density (OD) of 1 using a benchtop photometer (Biochrom, Cambridge, UK) at 590 nm. The cultures were then serially diluted between $1:10$ and $1:10^{10}$ and 100µL were plated on agar plates and incubated overnight in 37°C. Each dilution was plated 3 times for replicates,

and 3 plates were incubated without inoculation as controls. After incubation, colonies were counted on the plates and the bacterial concentration of a 1 OD liquid culture was calculated to be ~10^8 CFUs/ml in *E. coli* and *B. subtilis* and ~10^9 CFUs/ml in *P. aeruginosa*.

3.4. Preparation of L- tryptophan solution

A 10^7 ppb stock solution was prepared by dissolving 1 g of L-tryptophan powder (CAS number 73-22-3, Sigma-Aldrich, St. Louis, MO, USA) into 100 ml of non-fluorescence water (Merck, Darmstadt, Germany) at 100°C using an electric stirrer and hot-plate (Freed Electric, Haifa, Israel). The tryptophan stock solution was serially diluted and 10 ml of tryptophan solution at concentrations of 0.1, 1, 2, 3, 5 and 10 ppb were made.

3.5. Sample preparation for Raman scanning

Prior to scanning with Raman instrument, 15ml of bacteria culture were grown overnight in optimal conditions. The bacterial culture was centrifuged for 10 minutes at 4991 g (Heraeus Primo R, Thermo Scientific, Waltham, MA, USA). The supernatant was thrown away and the bacterial pellet re-suspended in 15ml of distilled HPLC-grade water (Biolab, Jerusalem, Israel). 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 optical density of 1. The sample was later diluted according to each experiment in serial dilutions. The samples were kept at room temperature at all times after this and were measured 1-3 hours after the preparation stage. Different sample preparations were used in some of the experiments as described in section 3.5.1

3.5.1. Other treatments

Other treatments were used in an attempt to increase bacterial Raman signal. These other treatments included: boiling bacteria for 15 min in 100°C in hot plate (Freed Electric, Haifa, Israel) by placing the prepared, diluted suspensions in the hot plate and cooling bacteria in ice for an hour (this has been shown by Premasiri et al to increase certain molecular moieties that are Raman active (Premasiri et al., 2017)).

3.6. Raman Instrumentation and scanning procedure

For all Raman experiments, a 785nm excitation laser was used (Ocean Optics laser module I0785MM0350MS, Ocean Optics, Largo, USA), coupled with a spectrometer (Ocean Optics QE65 Pro, Ocean Optics, Largo, USA) with a range of 160.7-4142.2 cm-1 (or 766 – 1100 nm). 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, Largo, USA) and exported as .CSV files for further analysis (Figure 5).

Unless otherwise stated, scans were done at 350 mW for 10 seconds, with 3 immediate scans averaged (for a total of 30 seconds exposure time). The height of the probe was adjusted to the best resolution of the peaks of 100% ethanol prior to all scans. Dark scans were measured with the laser off and auto-subtracted from all data to remove machine noise.

The scanning procedure was, after taking dark and adjusting probe height, to clean the loading aluminum cup (Figure 5) with 100% ethanol and delicate task wipers (Kimberly-Clark, Irving, USA) and load 150µL of sample onto the cup. The aluminum cup was then placed under the probe and brought to correct height with all light sealed out. At this stage the laser would be turned on and the scan will commence. After each scan the cup would be cleaned and reloaded. Each sample was scanned 3-5 times (depending on the experiment).

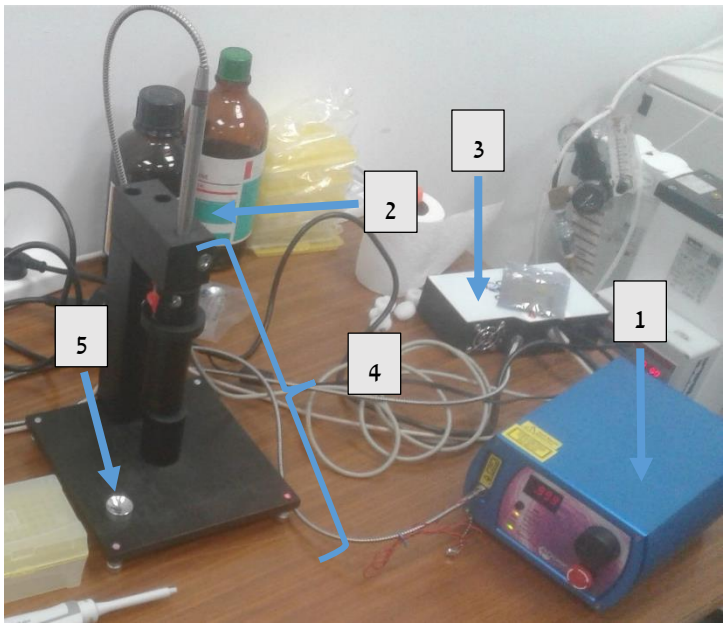

Figure 5. The Raman setup, including (1) excitation laser, (2) Probe, (3) spectrometer, (4) probe holder and (5) aluminum cup.

3.7. Survival test under sample preparation, transfer and radiation

In order to make sure irradiation by the 785nm laser does not compromise cell viability, a sample of water with 10^8 CFU/ml *E. coli* and a sample of clean water were placed in an aluminum cup and irradiated by laser for 30 seconds. After that, the samples were diluted 1: 10^3, 10^4, 10^5, 10^6 and plated on agar plates. The plates were incubated for 24 hours in 37°C and colonies were counted.

3.8. Water sampling

Drinking water samples were taken from 6 drinking water wells in Israel between July 2017 and June 2018 before and after standard chlorination treatment. The list of the drillings, the number of sampling events and the frequency of sampling are listed in Table 2. Water was collected in cleaned dark glass bottles and transported in coolers to the laboratory where it was stored in 4°C until analyzed. Some samplings were cancelled due to unavailability of the drill site reported by the water authorities.

Table 2. Water sampling schedule

Well (code)	Sampling Frequency	Number of 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 22/10 26/12 15/01 20/03 23/04 23/05 18/06	7
Shimron 7 (SH7)	Monthly	11
Einan 3 (QP003)	Summer (July-December 2017, March-June 2018): 2 per month Winter (January-February 2018): 2 per week	26
Einan 6 (QP006)	Summer (July-December 2017, March-June 2018): 2 per month Winter (January-February 2018): 2 per week	34
Total samples		99

3.9. Water samples standard testing

All drinking water samples were sent simultaneously to be analyzed in a service lab according to standard methods for: coliforms, fecal coliforms, fecal streptococci, heterotrophic bacterial counts and turbidity (2016; Eaton et al., 2005). In chlorinated samples chlorine concentrations were also measured. All testing was done by qualified personnel according to ministry of health standards and results were sent online.

3.10. Water sample preparation

250 or 500 ml of water from drillings were sampled by trained personnel according to standard water sampling procedure (2016) and transported in coolers to our lab in the Volcani Center (Rishon L'etzion). The water was put in 4°C until scanning. Prior to scanning the water was brought to room temperature by leaving the bottles in room temp for approx. 3 hours.

3.11. Fluorescence Scanning

Fluorescence spectra of all 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). 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 10mm. Instrument stability was controlled by measuring the intensity of the Raman peak of water at excitation wavelength at 275 nm and emission wavelength at 305nm. Cuvettes were washed 3 times with HPLC-grade water (Merck, Darmstadt, Germany) before the beginning of scanning, and washed by 6 ml of sample water between each scan. 3 ml of sample water were loaded for scanning. The cuvette was wiped with delicate task wipers (Kimberly-Clark, Irving, USA) before each scan and was aligned to the same side for scanning.

Each sample was scanned twice, and was scanned both before and after lab filtration using a 0.45 µm syringe filter. Prior to filtration, 15 ml of sample water were used to wash the filter.

3.12. UV Scanning

Optical density at 210nm was checked in every drinking water sample before fluorescence measurement using a Genesys UV-Visible spectrophotometer (Thermo Scientific; Cat. 335906-02; Madison WI 53711; USA) to ensure it is sufficiently low to prevent an inner filter effect.

3.13. Mathematical sample preparation – preprocessing

All preparation of data for analysis is called preprocessing. We first calibrated our protocol to find the best preprocessing procedure by comparing all the following techniques:

1. Normalization; this process is supposed to reduce noise since the ratio of signal to a set point that is theoretically stable over different measuring periods, temperatures (within 25±5°C) and varying intensities of lamp light will eliminate all the said interferences.

   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 experiment, this was done by dividing the signal to the Raman signal of water; different definitions of the Raman signal are elaborated in Table 7.

   Normalization was done using Excel software (Excel 2016, Microsoft, Redmond, WA, USA).

2. Centering and Scaling processes were done automatically using the JMP software prior to running the PLS analysis. These reduce noise to enable better fitting of the model and are commonplace in PLS analysis (Geladi and Kowalski 1986).
3. 1/signal and $\log_{10}$(1/signal); this is another normalization approach, and is generally used to increase weak signals and uncover more information about a sample (Mizrah 2007). These were calculated using Excel software (Excel 2016, Microsoft, Redmond, WA, USA).
4. 1st derivative; this preprocessing method is used to find trends in the data that are not obvious when looking at the raw data, because they relate to the trend rather than intensity of the spectrum. It was done by calculating $signal_{Em(n)}$-$signal_{Em(n-1)}$. The first data point in each spectrum was assigned "0" because it cannot be calculated, as $signal_{em(n-1)}$ does not exist. 1st derivative was calculated using Excel software (Excel 2016, Microsoft, Redmond, WA, USA).

3.14. Statistical analysis

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. Further, regression coefficients (R^2), Root Mean Square Errors (RMSE) statistical significance was calculated using JMP software by calculating student's t-test and Wilcoxon rank-sum test (also known as the Mann-Whitney test). Throughout the study, significance was defined as $p<0.01$. Cohen's kappa coefficient, which is a measure of the agreement between 2 methods, was also calculated using JMP.

4. Results

4.1. Raman Spectroscopy

Initially, Raman spectra of *E. coli, B. subtilis* and water was 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 (Figure 6).

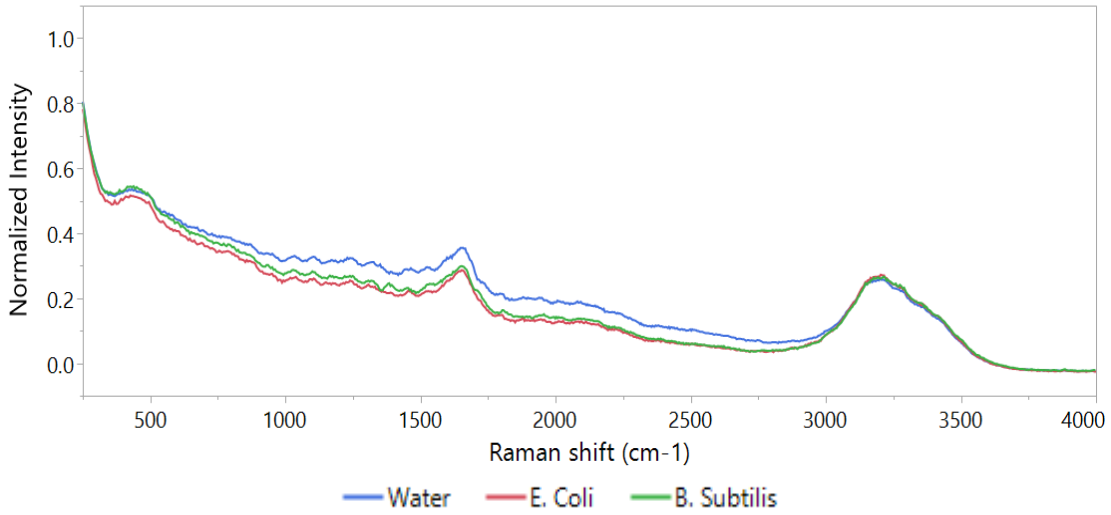

Figure 6. An example of the Raman spectra of clean double-distilled water, E. coli at a concentration of 10^8 and B. subtilis at a concentration of 10^7 CFU/ml. Excitation by 785 nm laser, for 5 seconds, 3 scans were averaged.

Different attempts have been made to improve the spectrum, by increasing time, cooling the samples to 4°C or boiling the samples to lyse the cells. None of these treatments showed significant improvements (data in supplementary). Since no clear difference could be found visually, a PLS model was designed to try and expose underlying information.

4.1.1. PLS model

Before starting to work on a PLS model, it is important to have a set protocol. We optimized the model based on the *B. subtilis* dataset (n=184), since it appeared to the naked eye to have a slightly better resolved spectrum.

In order to build a set protocol, different preprocessing approaches were tested and their measurements of accuracy are detailed in Table 3 as latent variables (LVs); which represent 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), the root mean square error; which is a measurement of the model's error rate in the units of measurement, and is a common and efficient measure of quality (Fearn, 2002) and the R^2; which is an easy to understand measure of model correlation with actual values (Fearn, 2002). We tested 5 types of preprocessing elements:

1) Normalization to maximum
2) Centering
3) Scaling
4) 1/signal, $\log_{10}$(1/signal)

5) 1st derivative

In all methods, the same attribution of calibration-validation sets was used, at a ratio of 6:4.

*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 + 1st derivative	None	2	1.22	0.46
18	Normalized + 1st derivative	Centering	1	1.22	0.45
19	Normalized + 1st derivative	Scaling	2	1.14	0.37
20	Normalized + 1st derivative	Centering + Scaling	1	1.16	0.38

Since no preprocessing was found to best describe the data. This is the protocol that was used for the rest of the experiment.

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 10^8 CFU/ml, *B. subtilis* prediction models enabled confident detection at 10^4 CFU/ml (Figure 7).

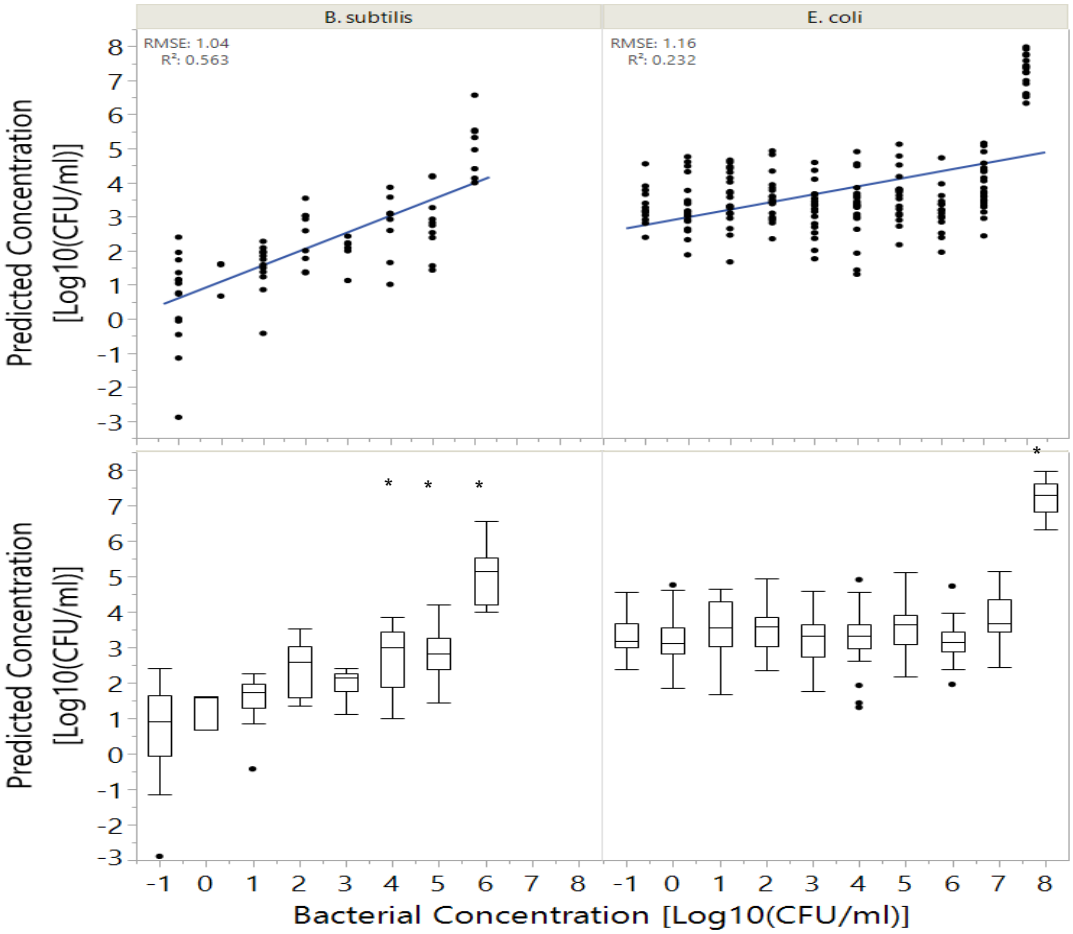

Figure 7. Top: regression of predicted bacterial concentration according to PLS model against real bacterial concentration of E. coli and B. subtilis. Bottom: box plot of the data as used for threshold analysis. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown. Asterixes 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.

4.1.3. Differentiation between species

In order to explain the differences between the prediction abilities of *E. coli* and *B. subtilis*, variable importance was calculated and plotted (Figure 8). It appears that the different bacteria have widely different spectral fingerprints. Furthermore, the major spectral fingerprint of *B. subtilis* is in the functional group region (Region B in Figure 8) and shows a sharp pattern, while the fingerprint of *E. coli* is smeared throughout the spectrum, which may indicate an absorbance signal, since it coincides with the Raman spectra of water (George, 2001).

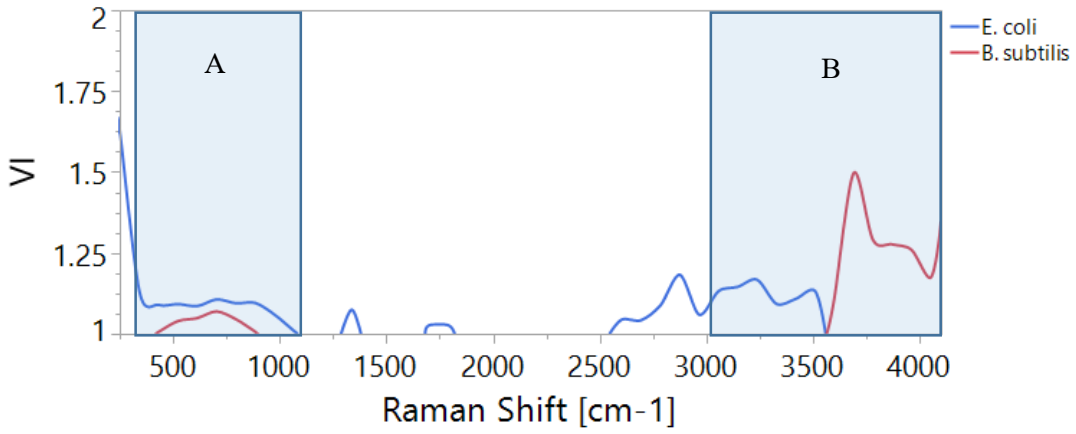

Figure 8. Variable importance (VI) plot of PLS models for quantification of bacteira according to Raman spectra. The plots show only values >1. Region (A) sinifies the fingerprint region , Region (B) signifies the functional group region.

This finding suggests the possibility of differentiation between different bacterial specie using Raman spectroscopy. In order to predict which bacteria is in the water, samples were 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. Clean water samples (n=67)
4. A PLS-DA (Discriminant Analysis) model was calculated, in an attempt to classify these samples correctly, with the same training-validation ratio of 6:4. Table 4 illustrates the model's ability to accurately differentiate between different bacteria. This finding shows that the model accurately differentiates between *E. coli* and *B. subtilis*, and also accurately detects their presence in water. This model has almost prefect agreement according to its Cohen's kappa coefficient of 0.89 (Landis and Koch, 1977).

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.

	Prediction κ=0.89				
Actual		*B. subtilis*	*E. coli*	Water	Total
	B. subtilis	17	1	1	19
	E. coli	0	18	0	18
	Water	3	0	27	30
	Total	20	19	28	67

4.2. Fluorescence

4.2.1. Single wavelength fluorescence

Calibration curve for tryptophan concentration

We used dilutions of L-tryptophan at concentrations of 0.1, 1, 2, 3, 5 and 10 ppb. Each dilution was scanned 3 times and 2 scanning sessions were conducted (meaning 6 scans per concentration).

Different wavelength pairs are used in the literature to calibrate tryptophan concentration (Table 5), and in order to choose the best approach we compared the correlation of different wavelengths with tryptophan concentrations (.). The local maxima of the peak calculated in our experiment using our spectrofluorometer was at excitation of 275 nm and emission at 362 nm. The results are summarized in Table 5, and show that the different methods show almost identical correlation. Because it is most commonly applied, and the differences are negligible, the wavelength combination of excitation at 280 nm and emission at 350 nm was chosen for the calculation of tryptophan equivalence (Equation 2).

Table 5. Different approaches to describe the peak of tryptophan and their correlation with tryptophan concentration. All R^2 values are significant $p<0.01$

	Tryptophan Ex-Em	Raman Ex-Em	R^2
Simelane	275-355	275-305	0.84
Bridgeman 2015,	280-350	280-310	0.82
Baker 2015, Sorensen 2015, Sorensen 2016, Sorensen 2017,	280-360	Not reported, normalized according to Bridgeman	0.84
Sorensen 2018	Ex/Em peak at **280** ± 15 / **365** ± 27.5 nm.	Not reported, normalized according to Bridgeman	0.83
Local Maxima	275-362	275-305	0.85

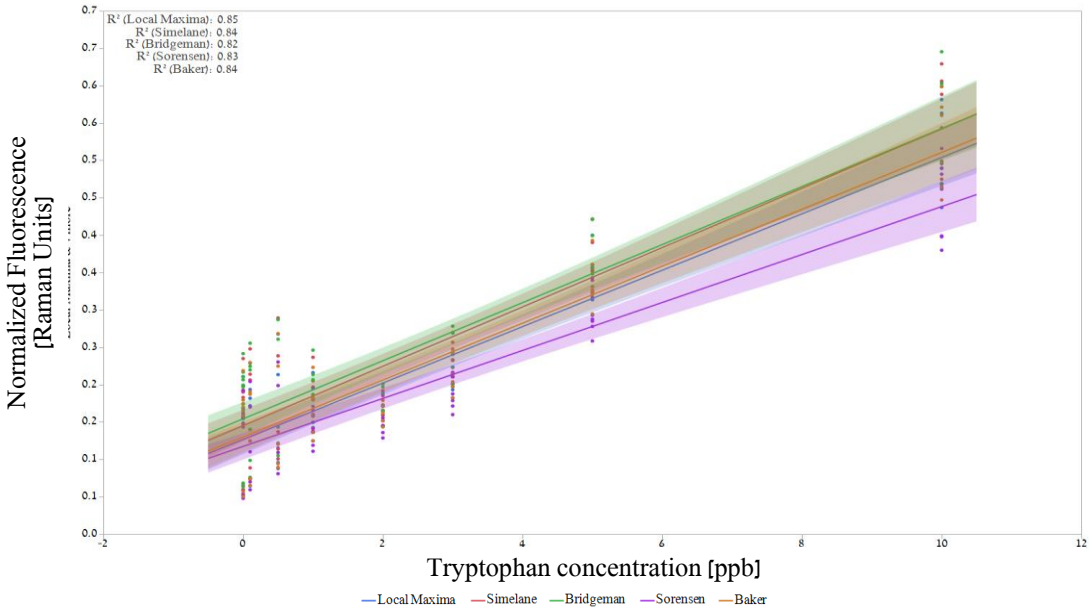

Figure 9. Correlation between different tryptophan fluorescence approaches. N=48 (51-3 excluded). Particular wavelengths of each approach elaborated in Table 5. Halos mark the confidence fit at α=0.05.

Equation 2. Claculation of tryptophan equivalence from normalized fluorescence

$$Tryptophan\ Equivalence_{[ppb]} = \frac{Normalized\ fluorescence - 0.1301063}{0.0380656}$$

$$\text{Where the normalized fluorescence } = \frac{(\lambda_{ex}280/\lambda_{em}360)}{(\lambda_{ex}280/\lambda_{em}310)}$$

Detection of bacteria in clean water using single wavelength fluorescence.

The concentration of *E. coli, B. subtilis* and *P. aeruginosa* in clean distilled water was correlated with tryptophan equivalence in order to assess the ability of single wavelength pair fluorescence measurements to detect bacteria in water. Whenever the concentration is given as "-1" it is a concentration of 0 CFU/ml. This is because there is no $\log_{10}(0)$. Figure 10 illustrates the detection thresholds of the single wavelength method to detect different bacterial species. All significance is tested using a Wilcoxson non-parametric test, since the variances are not equal nor normally distributed. The detection threshold using single wavelength can be as low as 10^3 CFU/ml, however, the major difference in fluorescence can only be observed around 2 ppb of tryptophan (Figure 10).

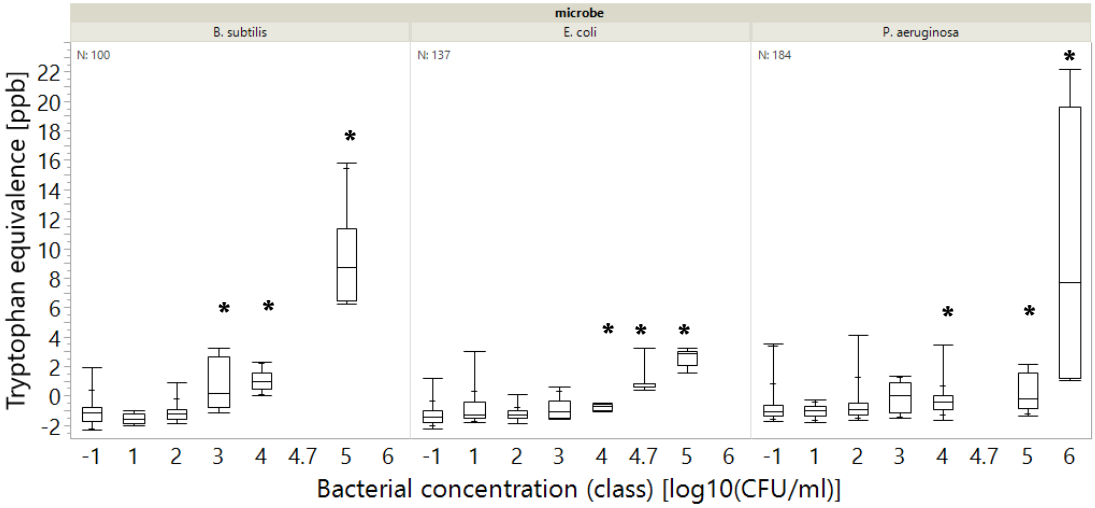

Figure 10 Tryptophan fluorescence measurements correlated with specific bacterial species. Bacterial concentration of (-1) means 0 CFU/ml. Astrixes mean significant differenct according to a Wilcoxon test, p<0.01. n indicated at the top. Bacterial concentration class means the all measurements were rounded to nearest class (ie 1,2,5 CFU/ml regarded as 1) for convenience. Box plots shows the middle quantiles in box, with whiskers for bottom and top values. The crosshairs in the whiskers signify the 90% mark.

Detection of heterotrophic bacteria in drinking water using single wavelength fluorescence.

In order to test the ability of single wavelength fluorescence spectroscopy to detect bacteria in real drinking water, drinking water samples were taken from 6 wells in Israel between July 2017 and June 2018. Over the entire sampling period, only 1 sample contained >1,000 CFU/ml of heterotrophic bacteria according to standard heterotrophic plate counts (HPCs). That sample is also the only one to contain *E. coli* >1 CFU/100ml. Therefore, the only microbial indicator chosen for correlation with fluorescence is HPCs, since it contained some variation.

The correlation between fluorescence measurement in tryptophan equivalence and heterotrophic plate counts (HPCs) in drinking water is shown in Figure 11. The right panel includes outliers with >500 CFU/ml (marked as Xs) and the left does not. Figure 12 shows the single wavelength approach's ability to differentiate between water samples of different concentrations significantly (according to Wilcoxon ranked sum test, $p<0.01$).

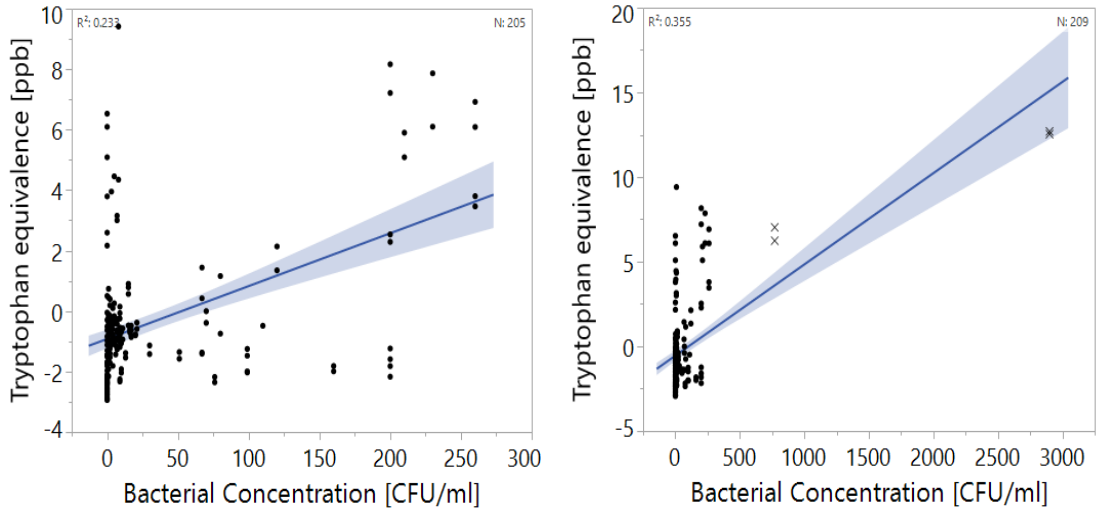

Figure 11. Correlation between HPCs from drinking water samples and tryptophan fluorescence. The right graph shows the entire data set (n=209) and the left one disregards outliers at 770, 2900 CFU/ml, which are marked as Xs. Halo is the confidence fit at α=0.05. All regressions are statistically significant, with p-value < 0.0001 using a student's t test.

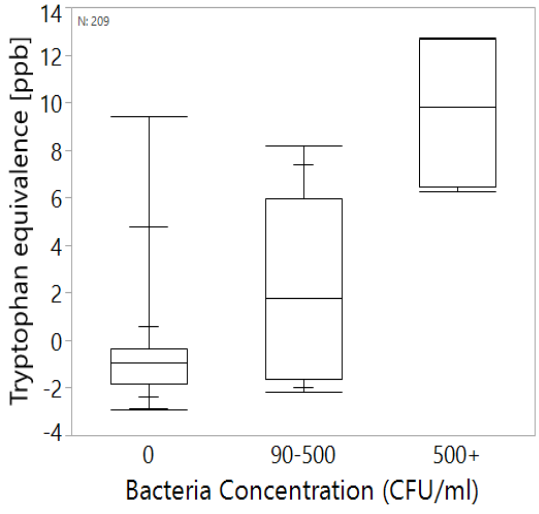

Figure 12. Box plots of Tryptophan fluorescence measurements and heterotrophic bacteria in drinking water. All groups are significantly different from each other according to a Wilcoxon test, p<0.01. n indicated at the top. Box plots shows the middle quantiles in box, with whiskers for bottom and top values. The crosshairs in the whiskers signify the 90% mark.

Case study – Shimron 7, 22/5/2018

The water sample taken at 22/5/2018 is a special case of severe contamination of water. This sample is the only sample in the entire dataset with >1,000 CFU/ml and was found to have 2,900 CFU/ml of heterotrophic bacteria, with 2 CFU/100ml of *E. coli* (an indicator of fecal contamination). The sample had very strong fluorescence (Figure 13) before treatment, however, chlorination at the treatment site and lab filtration with 0.45 µm filter diminishes the signal almost completely.

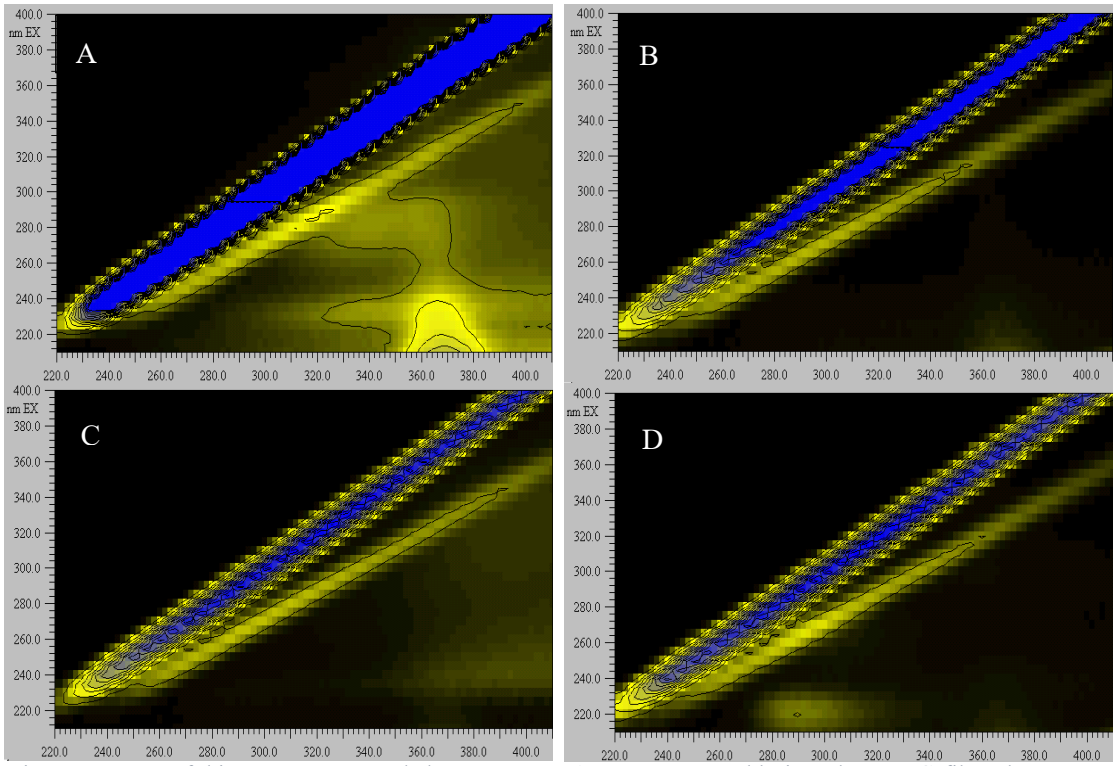

Figure 13. EEMs of shimron 7 water sampled on 22/5/2018. (A) raw water, (B) chlorinated water, (C) filtered raw water, (D) filtered chlorinated water.

This example, however, is a one-of-a-kind example in our data set, as all of our other samples were practically indistinguishable for the naked eye (Figure 14). All samples show strong emissions along the Rayleigh scatter line (where the excitation is roughly equal to emission) and along the Raman line (marked in Figure 14) but no significant signals in the protein or any other regions.

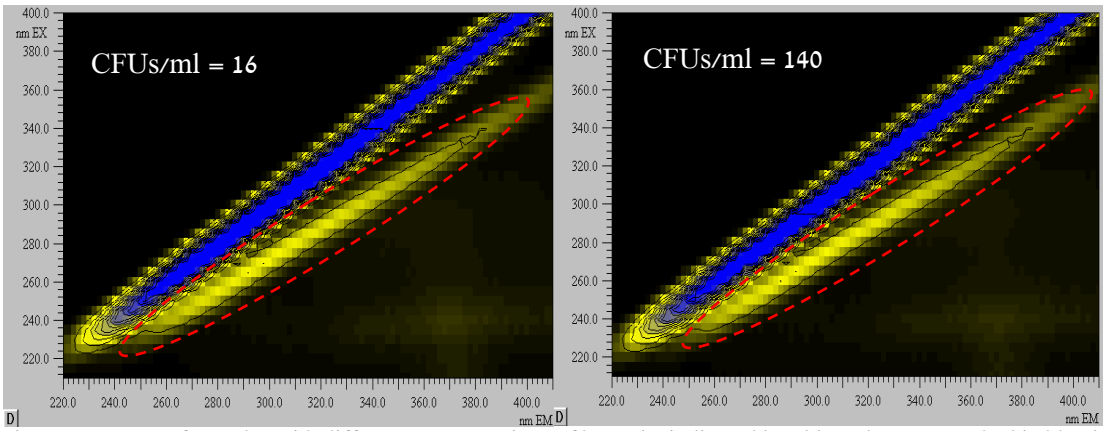

Figure 14. EEMs of samples with different concentrations of bacteria, indicated in white. The zone marked in blue is the Rayleigh scatter region, and the area surrounded by a red, dashed line is the Raman scatter region.

4.2.2. Multispectral fluorescence spectroscopy and PLS analysis

Before starting to work on the PLS models, it is important to have a set protocol. In order to build this protocol, different preprocessing approaches were tested on the *E. coli* data set (n= 137) (Table 6). We tested 5 types of preprocessing elements:

1) Normalization to Raman (280/360)
2) Centering
3) Scaling
4) 1/signal, $\log_{10}$(1/signal)
5) 1st derivative

In all methods, the same attribution of calibration-validation sets was used. At a distribution of 6:4.

*Table 6. 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**. Other format is in the "graphs edited out" section*

#	**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 (the lowest RMSE with the highest R^2 value). This is the protocol that was used for the rest of the experiment.

The effect of PLS on detection; comparison of single wavelength, spectra at excitation of 280 nm and full EEM

The effect of using PLS algorithms on detection of bacteria was first tested on the *E. coli* data set. We compared using the single wavelength approach with the commonly used full EEM analysis, and also tested the detection ability of the spectrum of emission (220-450 nm) at excitation by 280 nm. The idea was that if only a single excitation wavelength can be used, operational and computational costs can be reduced. In this analysis we compared the detection threshold and regression coefficients (R^2) of 4 possible detection methods:

1) Single wavelength method – the intensity at 280/350
2) Single wavelength method – the intensity at 275/362
3) Single wavelength excitation whole spectra – the light emission at all wavelength (220-450 nm) excited by 280 nm light
4) The entire EEM with excitation between 210-400 and emission between 220-450

Detection threshold was tested as the ability to differentiate the bacteria at different concentrations from samples with a concentration of <10 CFU/ml. Both Wilcoxon ranked-sum test and Tukey Kramer tests were used because while the Tukey-Kramer test is more robust, it was not applicable to some methods as they did not have equal nor normal variation (Table 7, Figure 15).

Table 7 Comparison of methods for using fluorescence spectroscopy data for the detection of bacteria. Analysis only includes 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 ($p<0.01$). n=60.

Method	R^2	RMSE I am not sure how to show this… the first 2 are in Raman units, while the second 2 are in log10 of CFU/ml	Detection Threshold [Wilcoxon] Log_{10}(CFU/ml)	Detection Threshold [Tukey Kramer] Log_{10}(CFU/ml)
Single wavelength [280/360]	0.35	0.04	4.7	---
Single wavelength [275/362]	0.302	0.04	4.7	---
PLS - 280 nm excitation spectra	0.646	0.93	2	4
PLS - Full EEM	0.783	0.56	1	1

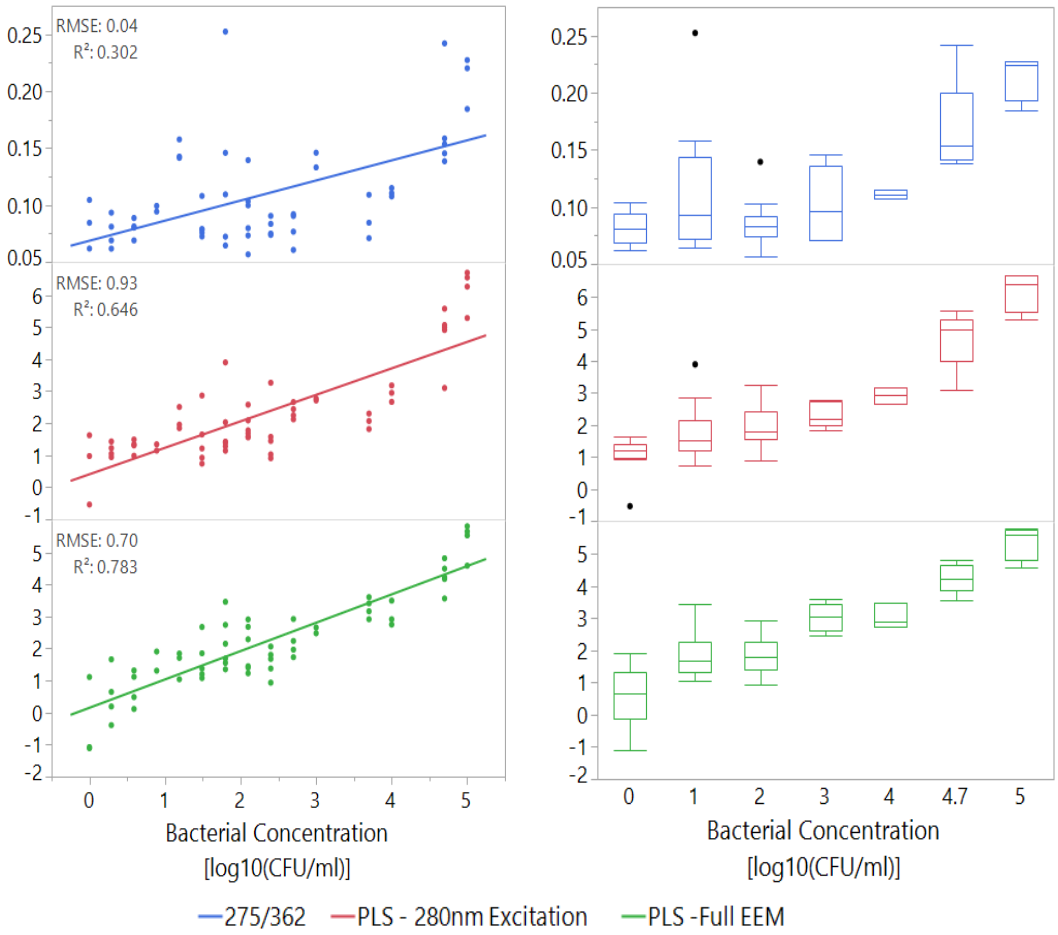

Figure 15. Left: regression of bacterial concentration with different detection methods, Right: box plot of the data as used for threshold analysis. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown. N=60.

Detection of bacteria in clean water using multispectral fluorescence spectroscopy and PLS

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 bacteria were selected as they represent both gram positive (*B. subtilis)* and gram negative (*E. coli and P. aeruginosa*) bacteria and because they have industrial and medical significance. Each strain was first modeled alone, and later all samples were modeled together.

Technical note – because this analysis was done on a different file, with a different randomly selected validation set, different results are shown than the previous *E. coli* model

Detection threshold was defined as significant difference from <10 CFU/ml, using Wilcoxon ranked-sum test or Tukey-Kramer when available. Results are shown in Table 8.

Table 8. Comparison of PLS approaches ability to detect different bacteria. All R^2 values are statistically significant.

Bacteria		E. coli (N=58)	B. subtilis (N=47)	P. aeruginosa (N=80)	All bacteria together (N=185)
Single wavelength (280 nm)	**RMSE** [Log_{10}(CFU/ml)]	0.82	0.84	0.86	0.91
	R^2	0.68	0.58	0.23	0.34
	Detection threshold (Wilcoxson) [CFU/ml]	100	10^5	10^6	10^3
	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

Detection of heterotrophic bacteria in drinking water using multispectral fluorescence spectroscopy and PLS

In this experiment, EEMs of drinking water samples that were collected throughout the year were modelled using a PLS model. 2 samples were significantly different from all others and

could easily be detected visually, both from Shimron 7, having 770 and 2900 CFU/ml. Those outliers were excluded prior to modelling. The results of the PLS models are displayed in Figure 16 and Figure 17.

Error! Reference source not found. Figure 17 illustrates the difference between using PLS models and single wavelength analysis. While the differences are not immediately apparent to the eye, statistical testing shows that by using either the 280nm excitation data or the full EEM, better differentiation between samples with over or under 90 CFU/ml can be achieved. However, when a confusion matrix is calculated and the model is put to the test (Figure 17), it is clear that all methods are not good predictors, with a Cohen's kappa coefficient of <0.6. While the full EEM model is slightly more accurate than the single wavelength approach, it is only the result of 2 more correctly classified samples, which is of little value.

Single Wavelength 275-362	Prediction [RU] * κ=0.529			
		<0.22	>0.22	Total
Actual [CFU/ml]	<90	50	4	54
	>90	7	9	16
	Total	57	13	70

Full EEM	Prediction [CFU/ml] κ=0.590			
		<90	>90	Total
Actual [CFU/ml]	<90	52	2	54
	>90	7	9	16
	Total	59	11	70

280 nm Excitation	Prediction [CFU/ml] κ=0.213			
		<90	>90	Total
Actual [CFU/ml]	<90	50	4	54
	>90	12	4	16
	Total	62	8	70

Figure 16. Confusion matrices of prediction of bacterial concentration, fluorescence intensity in raman units, against actual bacterial concentration. Cohen's kappa coefficient is shows in each table as κ.

4.2.3. Analysis of Spectral Fingerprints

In order to see what spectral regions are important, variable importance (VI) was calculated for each variable (each excitation-emission wavelength pair). 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 a lot of variables ((Eriksson et al., 2006)).

For this analysis, the VI values were plotted from the 280nm model (Figure 18). Note: the graph is cut after 410 nm because the heterotrophic bacteria scans ended at this point, while the pathogen scans ended at 450 nm. The shaded regions illustrate that there are slight differences between different bacteria's spectral fingerprints even in a model based only on the 280 nm excitation spectrum.

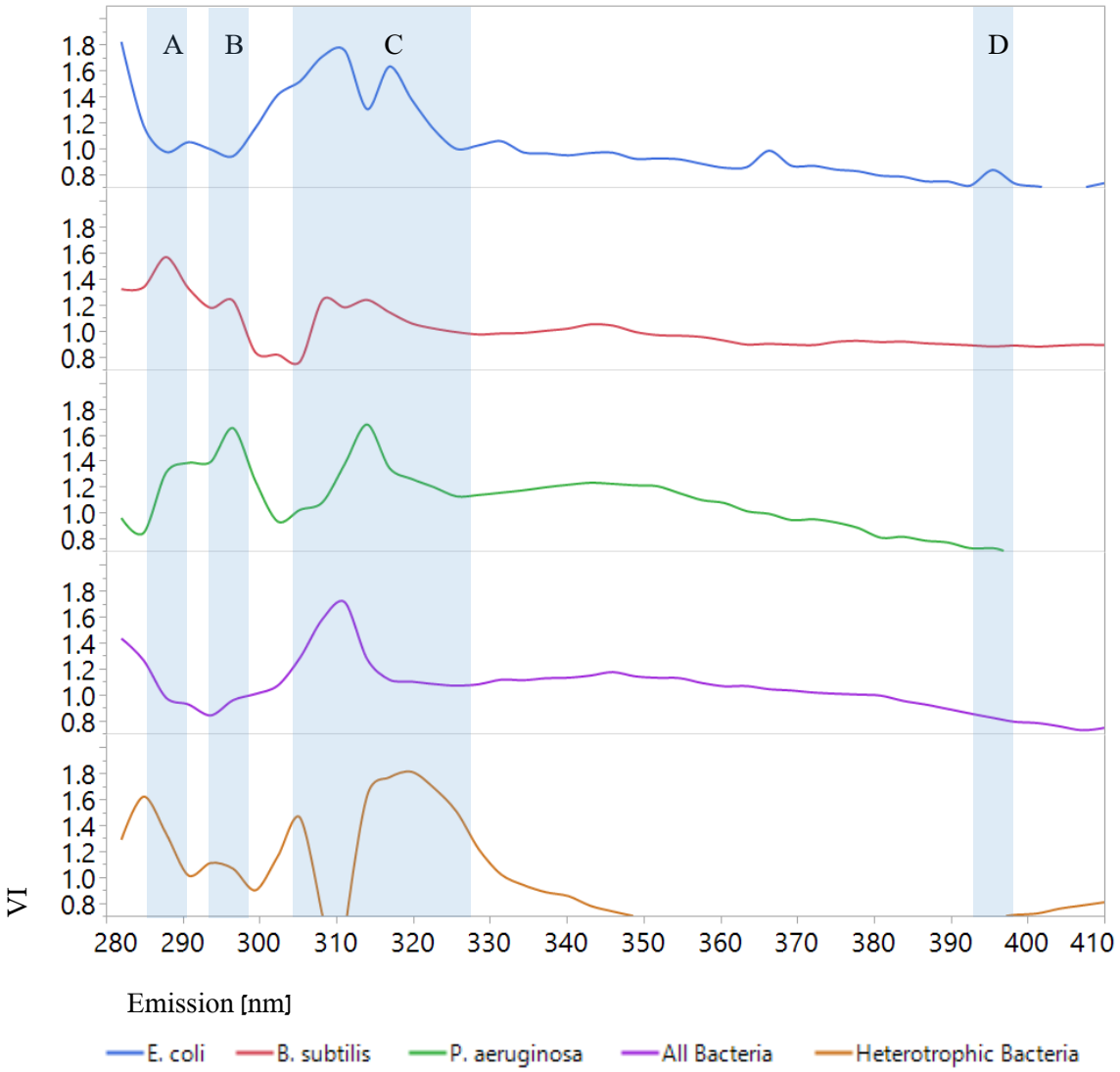

Figure 17. Variable importance (VI) plots of PLS models made from 280 nm excitation spectra. The plots show only values >0.8. Shades a,b,c and d indicate areas of peaks which may be indicative bacterial concentration.

When the VI values of entire EEM, at excitation range of 210-400 nm and emission of 220-410 nm were plotted (Figure 19), a difference between the spectral fingerprints could be observed. While most of the signal came from the protein-like fluorescence region (220-230, 270-300 excitation, 300-360 nm emission, (Yang et al., 2015), certain shifts in this region are visible. Furthermore, it appears that the scatter region (where the excitation and emission are within ~10 nm of each other) of the map is also of high importance to the model.

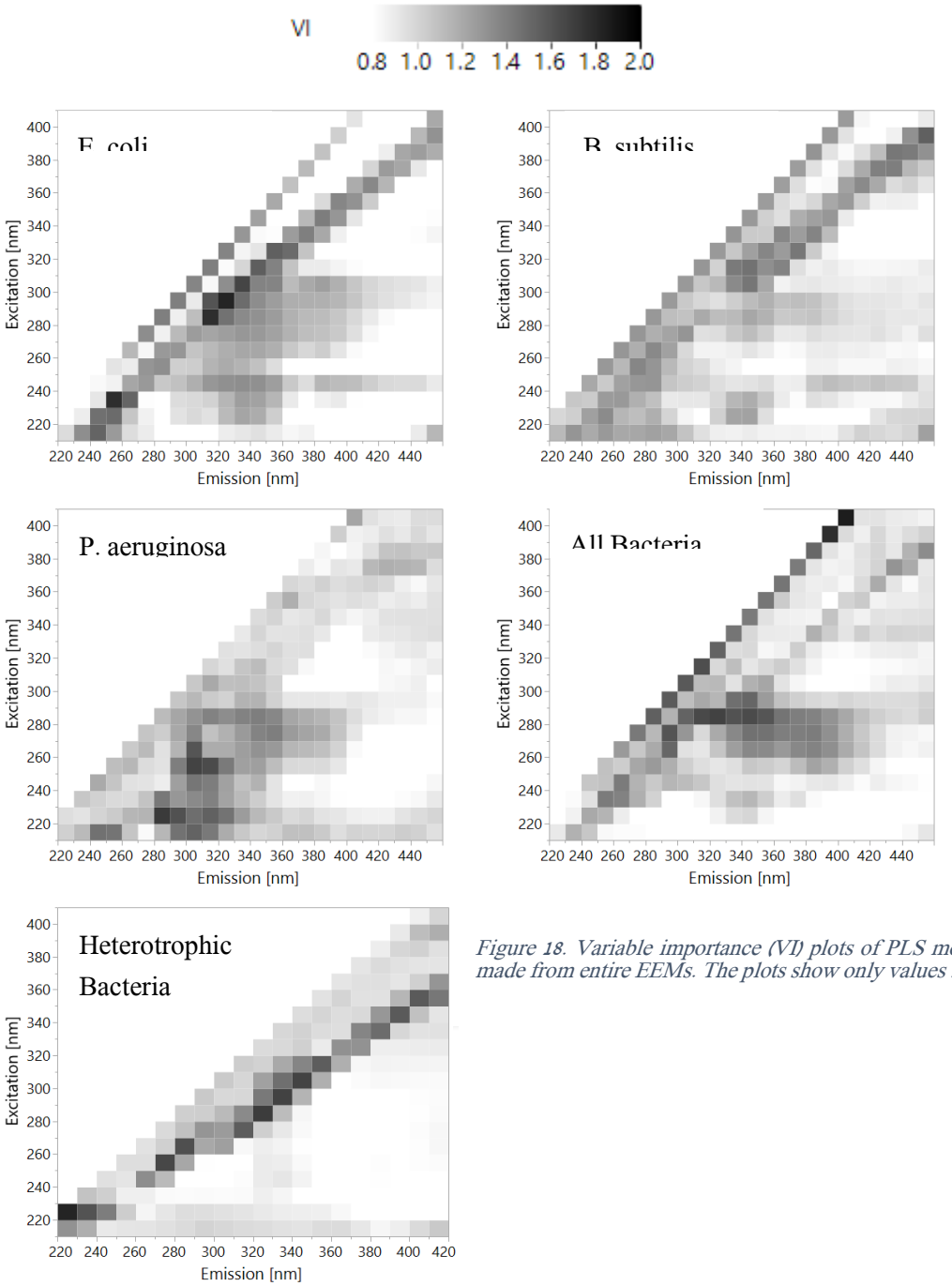

Figure 18. Variable importance (VI) plots of PLS models made from entire EEMs. The plots show only values >0.8.

4.2.4. Differentiation between different species

In order to predict which bacterial species is in the water, 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 9 illustrates the model's ability to accurately differentiate between different bacteria. This finding shows that the model accurately differentiates between *E. coli*, *B. subtilis* and *P. aeruginosa*, however, it does not differentiate between *P. aeruginosa* and water very well. This model has substantial agreement according to its Cohen's kappa coefficient of 0.78

Table 9. Confusion matrix of classification of microbes by PLS-DA, based on entire EEMs, at concentrations >10^4 CFU/ml. Only validation set is displayed. K=0.78

	Prediction					
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

5. Discussion

Microbial contamination of drinking water poses a threat to public health, with thousands of hospitalizations yearly even in devolved countries (Collier et al., 2012; Leclerc et al., 2002). The current standard methods for detection of contamination require a long and labor-intensive culturing stage to get results. Many modern methods have been suggested to replace these standard methods, however most molecular methods are too costly to be applied in the food and water industries (Stöckel et al., 2015). Spectroscopy based methods such as Raman and Fluorescence spectroscopy have also been put forward as alternatives, although these have been mostly applied in the medical fields (Eberhardt et al.; Krafft and Popp) or in highly contaminated water such as boreholes in Indian rural areas and wastewater treatment plants (Sorensen et al., 2015; Yang et al., 2015). The application of low resolution Raman spectroscopy or fluorescence spectroscopy for the detection of bacteria in the water industry requires defining the detection threshold of these approaches, and comparing those to the industry and regulatory requirements. Furthermore, an ability to differentiate between bacteria using these quick analytical methods may have industrial and medical applications. In this study we used low resolution Raman spectroscopy and fluorescence spectroscopy to try and detect bacteria, first in pure cultures suspended distilled water, and later in real drinking water.

5.1. Raman spectroscopy for detection and differentiation of bacteria in water

Low resolution Raman spectroscopy involved applying a laser beam of 785 nm light onto water samples inoculated by different concentration of either *E. coli* or *B. subtilis*. Unsurprisingly, no clear spectra could be obtained by this crude approach. This is because the Raman signal of bacteria is very weak and complex, and even in high concentrations the chance of a Raman reaction occurring at a frequency which can be clearly visible using a low-resolution instrument is remote (Jarvis et al., 2006; Schmilovitch et al., 2005). Nonetheless, with the application of multivariate analysis using PLS, we managed to confidently detect *E. coli* and *B. subtilis* in concentrations of 10^8 and 10^4 CFU/ml respectively (Figure 7). While the finding regarding *E. coli* is problematic, as the concentration detected is very high, an ability to detect *B. subtilis* at a concentration of 10^4 CFU/ml shows a significant improvement on current methods such as absorbance at 590nm. That being said, for industrial purposes, this does not appear to be a promising path for the rapid and accurate detection of bacteria in water.

Different approaches have been applied in an attempt to enhance the Raman signal of the bacteria, without the use of expensive reagents, procedures or instruments. Neither boiling nor cooling the samples did not show any improvement on the signal of *E. coli*.

Interestingly, the spectral fingerprint that was detected for *B. subtilis* did occur in the expected functional group region, meaning 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). In *E. coli,* the spectral fingerprint was smeared, and appeared to be derived from a reduction in overall signal, such as may be expected in an absorbance reaction. This is supplemented by the fact that below the 10^8 CFU/ml, no clear signal is observed and the PLS

model could not find any variation, much like in the use of standard absorbance photometers below the 10^8 CFU/ml threshold no absorbance is observed.

In a different perspective, differentiation of *E. coli* and *B. subtilis* was excellent at a concentration higher or equal to the detection threshold (Table 4). This may suggest a possibility for differentiation between these species in a dense culture quickly and accurately, but may also be the product of the *E. coli* spectral fingerprint being derived from absorbance rather than Raman effect.

It is difficult to explain why exactly *E. coli* does not show a spectral fingerprint while *B. subtilis* does. One hypothesis is that the *B. subtilis* cells create large aggregates which may scatter light more effectively (Duanis-Assaf et al., 2018). Another is that the difference in the cell wall structures of these cells may cause a different scatter effect; since these species are gram positive (*B. subtilis*) and gram negative (*E. coli*), this may affect the transmittance of light into a cell or the scatter off the cell wall. Further research may explain the molecular differences in low resolution Raman spectroscopy.

Taken all together, these results signify a poor ability to detect bacteria in water using low resolution Raman. This finding is surprising since several studies have already demonstrated a promise for the application of this method for detecting bacteria (Mizrach et al., 2007; Schmilovitch et al., 2005; Stöckel et al., 2015). One of the possible explanations for this finding is the use of an instrument that may not be appropriate for the job. The Raman instrument chosen for this study has not been designed for the detection of bacteria, and while it represents an example of a low-resolution instrument, it has not been used in any of the studies cited. This could not have been known in advance since the instrument has been calibrated to read ethanol and appeared to work perfectly. Then again, the Raman effect of ethanol is very strong (Socrates, 2001) and a more sensitive, yet still cheap, instrument may yield better results. It has been suggested that suspending the bacteria in distilled water would ruin their Raman spectra, due to osmotic-stress-related lysis, or that the laser used may damage the cells considerably and will actually reduce the signal, however, a simple viability test has shown that the cell viability has not been compromised after irradiation and suspension in distilled water. Moreover, the use of lower laser energy nor suspending the cells in a saline solution instead of water improved Raman signal (data not shown).

In conclusion, low resolution Raman may not be a good method for the detection of bacteria in water, according to the findings of this study, and the conditions of our experiment.

5.2. Single Wavelength fluorescence spectroscopy for detection of bacteria in water

In fluorescence spectroscopy, two main approaches are commonly applied in order to analyze samples, the first is the "peak picking" or single wavelength approach. This method involves measuring the emission of light at a single wavelength, after excitation by a single wavelength. 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) peak 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 different studies the minimal concentration for detection was 1-3 ppb of tryptophan equivalent, which is the calibrated measure of fluorescence in tryptophan units. Different studies have defined this threshold as signifying >2 CFU/100ml of *E. coli* contamination in drinking water, >100CFU/100ml of *E. coli* in drinking water or 20-150 x 1,000 cells/ml using flow cytometry. This means that the commonly observed detection threshold for bacteria in water, using the TLF peak single wavelength measurement is ~2 ppb of tryptophan equivalent.

We tried to quantify the detection limit of this method using our instrument in known bacterial species suspended in distilled non-fluorescent water, in tryptophan equivalent. We first calibrated our instrument by creating a calibration curve, using different wavelengths found in the literature and our own maximum emission wavelength in the TLF region (.). Using our spectrofluorometer's maxima as the proxy for tryptophan equivalence, we can see a significant increase in TLF in *B. subtilis* and *E. coli* suspensions at a concentration of 10^5 CFU/ml and in *P. aeruginosa* suspensions at 10^6 CFU/ml. The increase is roughly to the known 2 ppb of tryptophan equivalence, signifying that the single wavelength method applied by Sorensen, Baker and Bridgeman is capable of detecting bacteria in ~10^5 CFU/ml. Furthermore, statistical analysis has shown an ability to detect bacteria at a concentration of 10^3, 10^4 and 10^4 CFU/ml for *B. subtilis, E. coli* and *P. aeruginosa* respectively using the single wavelength approach.

This finding signifies that using single wavelength fluorescence spectroscopy, at the TLF region, bacteria can be detected in clean water at a concentration of 10^3-10^4 CFU/ml. While not keeping up with the Ministry of Health regulations (2013), this could have industrial implications, since it enables containment of contaminations at a relatively early stage, and much earlier than most methods enable, using a quick, cheap fluorimeter.

On the grounds of this finding, we collected water samples over a year from different drinking water wells in the north of Israel. The raw drinking water was transferred immediately after sampling to our lab for fluorescence scanning, and to the main microbial lab of the water authorities for standard microbiological testing which includes *E. coli,* coliform and fecal coliform counts as well as heterotrophic plate counts (HPCs). We then tried to apply our method for the enumeration of HPCs in drinking water and while little correlation was measured between fluorescence and bacterial concentration (R^2=0.23, p<0.01, Figure 11**Error! Reference source not found.**), we could statistically differentiate samples with >90 CFU/ml from those with less bacteria (Figure 12). Furthermore, even though a very small sample size (n=2) was found of samples with over 500 CFU/ml, those samples showed a significant increase in fluorescence which could be visually observed (Figure 13).

When observing these highly contaminated samples, it was obvious that the signal is not limited to the TLF peak area of 275±5 nm excitation and 360±10 nm emission but is actually spread over a large area of the EEM, including the entire proteinaceous region (220-300 nm excitation, 300-400 nm emission,(Yang et al., 2015)), the edge of the humic substance region (210-300 nm

excitation, 400+ emission,(Yang et al., 2015)) and also included a strong peak in the region of 210 nm excitation and 370±10 nm emission, as described by Simelane (Simelane, 2013). This has led us to believe that perhaps by observing the entire EEM we could improve significantly our detection abilities. This was particularly relevant since samples with <500 CFU/ml showed very little overall fluorescence (Figure 14), which may be clarified using PLS multivariate analysis.

5.3. Multispectral fluorescence spectroscopy for detection and differentiation of bacteria in water

The second common approach for analysis of fluorescence spectroscopy data is analysis of entire fluorescence excitation emission maps (EEMs) ((Borisover et al., 2009; Carstea et al., 2016; Heibati et al., 2017)). This enables taking into account the large regions previously mentioned, and with advanced multivariate statistics also to weigh their different contribution to the overall spectrum. PLS regression enables the formulation of a prediction model to predict the concentration of bacteria in water according to the entire EEM. In order to measure the benefit of using a PLS model we compared the detection threshold of *E. coli* in distilled water using the single wavelength approach, with a PLS model designed on either the spectrum of emission at 220-450 nm when excited by 280 nm light, or the entire EEM measured in the excitation range of 210-400 nm and emission of 220-450 nm. The reason for this comparison is because we believe that perhaps the emission spectrum of a single wavelength could be enough for the detection of bacteria in water, which will make future instrument design simpler since only a single light source will be required. We chose the 280 nm wavelength since it falls within the proteinaceous fluorescence zone and has shown good results. The difference in correlation and detection threshold between using a single wavelength, a PLS model based on single excitation of 280 nm and a PLS model based on the entire EEM was significant (Table 7). We found that by applying the PLS model on either the single excitation or the entire EEM we could lower the detection threshold by 2.5-3.5 orders of magnitude. Furthermore, this method had less variation between groups which meant the Tukey-Kramer test can be applied, signifying a more robust result.

We followed up this finding by testing out the different detection thresholds of *B. subtilis* and *P. aeruginosa*. These bacterial species were chosen because a) they represent the gram positive and gram negative groups, which we expect to have a different chemical composition (Jean and Simorre, 2018) and b) because *P. aeruginosa* is a very common environmental pathogen which may be acquired from drinking water (Costa et al., 2015). The detection thresholds of the different species were surprisingly different when using the entire EEM or only the 280 nm excitation spectrum; with differences of between 3-6 orders of magnitude (Table 8). This indicates that the 280 nm excitation spectrum does not capture the entire spectral fingerprint of the bacteria. This is likely because the proteinaceous region is a large 3-dimentional space and a large proportion of it is lost when a single 2-dimentional slice of it is analysed. The variation in protein structures, the different positions of aromatic amino acids within proteins and the different aromatic moieties in bacterial cells are far more diverse than what can be captured by looking at the spectrum of emission from excitation at 280 nm, or likely any other single wavelength. It is interesting to note that the increase in sensitivity from 1-dimentional single wavelength (280 nm excitation, 360 nm emission) to a 2-dimetional spectrum at a single excitation wavelength (280

nm excitation) is ~3 orders of magnitude, and the increase from single excitation spectra to full 3-dimensional EEM is again several orders of magnitude (depending on the bacterial species).

Additionally, different species were found to have different detection thresholds, varying between 10-1,000 CFU/ml (Table 8). The difference, surprisingly, corresponds to the bacterial gram type; gram negative bacteria (*E. coli, P. aeruginosa*) were detected at a concentration of 10 CFU/ml, while gram positive bacteria (*B. subtilis*) were detected only at 10^3 CFU/ml. 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 which does not contain aromatic amino acids (Jean and Simorre, 2018) and may block the light from reaching the inside of the cells. In addition, gram negative bacteria are known to have an outer layer which is abundant with proteins, some of which might contain aromatic side-chains and structures (Madigan et al., 2006). This difference in detection threshold hints that each bacterial species is likely to have its own spectral fingerprint, which may be used to differentiate different bacteria in water samples.

5.4. Detection of heterotrophic bacteria in drinking water using fluorescence spectroscopy

Since the detection threshold for the detection of bacteria in water was equal to or lower than 1,000 CFU/ml (Table 8), this method may be applied for the detection of heterotrophic bacteria in drinking water. The detection limit required by the ministry of health and other regulatory bodies is 1,000 CFU/ml for heterotrophic plate counts (2013; Allen et al., 2004).

In order to test this hypothesis, we scanned entire EEMs of raw drinking water collected in wells in the north of Israel, and tried to train a PLS model to predict the concentration of heterotrophic bacteria in them. Although we collected water over a yearly period, in drinking water wells which were expected to have occasional microbial contaminations, only one sample over the entire sampling period had more than 1,000 CFU/ml and another over 500 CFU/ml. The rest of the samples (n=97) contained between 0-300 CFU/ml of heterotrophic bacteria as measured by standard methods. A model was then designed to differentiate between samples with either more or less than 90 CFU/ml. This threshold was chosen as it approximated 100 CFU/ml, but still contains several samples which had 99 or 98 CFU/ml. While the model did manage to significantly differentiate samples with either over or under 90 CFU/ml (**Error! Reference source not found.**), the method did not show very good classification ability (Figure 17). This means that while using the entire EEM for the detection of bacteria in real drinking water improves the sensitivity of fluorescence spectroscopy for the task, it is still not accurate enough for reliable differentiation ($\kappa<0.6$). We considered training a PLS model for the detection of higher concentration, but since only 2 samples had more than 500 CFU/ml over the entire measuring period, no prediction model could be calculated. Collecting a larger, more diverse data set, which involves more contamination events is likely to yield significantly better results.

It is difficult to tell which approach, the use of EEMs, single excitation wavelength spectrum or single wavelength fluorescence measurements, will be best applied for the detection of low concentrations of heterotrophic bacteria in drinking water based on this study. It is however,

highly probable that fluorescence spectroscopy can be used for the detection of microbial contamination of water.

5.5. Differentiation of different species of bacteria

Since different detection thresholds were calculated for different bacterial species, it is expected that each species has a different spectral fingerprint. In order to understand the spectral fingerprint of different bacteria, the VI of the PLS models used to quantify bacteria were plotted and analysed (Figure 18 and Figure 19). It is interesting to note that all bacterial strains, as well as heterotrophic bacteria have a spectral fingerprint in the TLF region (at 225–237 nm excitation with 340–381 nm emission and 270–280 nm excitation and 330–368 nm emission) however, this fingerprint is different across species. Furthermore, it is surprising to see the importance of the Rayleigh scatter region and the Raman scatter region for the quantification of bacteria. It appears that all strains have some effect on light scattering, which was surprising since Raman spectrometry did not yield good results. One explanation for this is that while the Raman spectrum measured in earlier experiments was based on excitation by near-infrared light (785 nm), the spectrofluorometer excited the samples using UV light (200-400 nm) which has much higher energy levels. It is difficult to explain what molecular moieties are responsible for the different fluorescence spectra of different species. It is clear that proteinaceous substances form a large part of the spectral fingerprint, but the source of these could be entire cells, cell wall components or dissolved proteins and peptides which were excreted by the bacteria and it is not clear from the spectra which of these actually contributes to the spectral fingerprint. Further studies are needed to elucidate the precise mechanism of the fluorescence reaction. It is clear however that these fluorescence fingerprints are correlated to colony-forming-units, and may thus be further employed in industrial environments.

Considering the different spectral fingerprints of different species, we hypothesised that a PLS discriminant model can be designed to classify different bacterial species. The EEMs of *E. coli, B. subtilis* and *P. aeruginosa* pure cultures in distilled water were used alongside the EEMs of clean water to try and classify the samples according to their spectra. Only samples at a high concentration ($>10^4$ CFU/ml) were used, and a good classification model was created (Table 9). This model has substantial agreement with real classifications (K=0.78), and shows near perfect classification of both *E. coli* and *B. subtilis* (16/17 correctly classified samples in the validation set). The model poorly distinguishes *P. aeruginosa* samples from water, although it does not confuse these with other species. This strongly suggests that fluorescence spectroscopy full EMM analysis can be used to differentiate between species of bacteria in pure, dense cultures. It is not clear exactly why *P. aeruginosa* is poorly distinguished. A bias is not expected since water samples were taken evenly from all sampling days where bacteria were tested equally.

5.6. The importance of multivariate statistics

This study demonstrates the breakthroughs that can be achieved when using advanced multivariate statistical models like PLS, instead of "peak picking" or visual assessment, on spectral data. In different experiments, data which seemed to have no significance to the naked eye was calibrated and modelled to enable higher sensitivities by several orders of magnitude

(Figure 15, Table 7). It is obvious that with the advancement in computer power, the speed of calculation and the accessibility of smart algorithms, spectral data can now be much better understood. Better models can probably be made with more advanced algorithms than PLS, some studies employ support vector machine (SVM) algorithms to better describe non-linear phenomena (Rösch et al., 2005), others use Parallel Factor Analysis (PARAFAC) to resolve overlapping peaks in fluorescence spectra (Baghoth et al., 2011; Borisover et al., 2009; Nebbioso and Piccolo) and few even use complex machine learning algorithms which include sparse coding and neural networks to understand the obscure interplay between variables (Frolich et al., 2017; Stöckel et al., 2010).

6. Conclusions

In conclusion, this study demonstrates a high potential for the application of fluorescence spectroscopy in water. This modern method can quickly and accurately detect bacterial presence in water, without the need for complex sample preparation nor expensive reagents. It can also be applied for fast discrimination between different bacterial species in dense cultures, which has industrial and medical significance, such as quickly identifying a specific pathogen in a patient's urine and prescribing the suitable antibiotic (Pazos-Perez et al., 2016), or being able to tell if the wine-fermenting bacteria in a barrel are actually spoilers rather than the fermenters of aroma (Rodriguez et al., 2013).

That being said, further studies are needed to optimize the system for real drinking water generally, and in Israel particularly. Collection of a larger set of data, that will be more diverse and include more contamination events, will probably enable accurate detection of bacteria in drinking water, at industrial scales, quickly and efficiently.

7. References

תקציר

זיהום בקטריאלי של מי שתייה מהווה סכנה רבה לבריאות הציבור. מדי שנה, אלפי אנשים חולים במחלות הנישאות במים במדינות מפותחות, ומיליונים במדינות מתפתחות. בעוד רשויות המים מבצעות פעולות רבות כדי להבטיח את בטיחות מי השתייה, השיטות לבדיקות סניטציה של מים הן איטיות מדי, ולוקחות בין 12- 72 שעות כדי לייצר תוצאות. שיטות מודרניות מולקולריות הוצעו כתחליפים לשיטות הסטנדרטיות, כולל הגברת רצפי דנ"א, Enzyme-Linked Immunosorbent Assay (ELISA) וספקטרומטריית מסות. על אף זאת, שיטות אלה אינן מתאימות לבדיקות בקנה מידה תעשייתי היות והן יקרות מדי או אינן מדויקות מספיק. גישות מבוססות ספקטרוסקופיית אור, כגון ראמאן וספקטרוסקופיית פלואורסנציה הן שיטות אנליטיות מודרניות ומהירות אשר ניתן להחיל על גילוי של חיידקים במי שתייה. שיטות אלה מבוססות על הארת דוגמה באמצעות מקור אור ומדידות ההשפעה של הדוגמה על האור הנפלט.

במחקר זה, ספקטרוסקופיית ראמאן ברזולוציה נמוכה וספקטרוסקופיית פלואורסנציה שימשו כדי לנסות לגלות חיידקים, תחילה במערכת מודל ולאחר מכן במי שתייה אמיתיים. חיידקי *E. coli*, B. subtilis ו- *P. aeruginosa* שימשו להערכת ספי הגילוי של השיטות השונות, על ידי מדידת ספקטרום האור באמצעות ספקטרומטר ראמאן או בעזרת ספקטרופלואורומטר. אנליזת Partial Least Squares (PLS) שימשה לניתוח הספקטרה המורכבים, כיוון שלא נמצאו שיאים (Peaks) מובהקים. לאחר מכן, נמדדו מי שתייה אמיתיים מבארות בצפון ישראל באמצעות ספקטרומטריית פלואורסנציה, על מנת לקבוע את היכולת של השיטה לזהות זיהום של מי שתייה אמיתיים.

נמצא כי ספקטרוסקופיית ראמאן ברזולוציה נמוכה אינה כלי בעל ערך רב עבור זיהוי של חיידקים במים, עם ספי גילוי של 10^8 ו-10^4 CFU/ml עבור *E. coli* ו-*B. subtilis* בהתאמה. ספקטרוסקופיית פלואורסנציה לעומת זאת, הראתה פוטנציאל גדול לזיהוי וכימות של חיידקים במים, גילוי חיידקים בסף גילוי נמוך עד כדי 10 CFU/ml עבור מינים מסוימים. יתר על כן, מודל סיווג היה מסוגל להבדיל במדויק בין *E. coli*, *B. subtilis* ו-*P. aeruginosa* בריכוזים גבוהים, ממצא המצביע על טביעות אצבע ספקטרליות שונות בין המינים. במי השתייה האמיתיים נמצא כי במערך הנתונים יש מעט מאוד מקרים של זיהום, כאשר רק מדגם אחד מזוהם באופן רשמי לאורך כל תקופת הדגימה. במדגם זה ניתן, בכל זאת, להבחין בזיהום בקלות באמצעות ספקטרוסקופיית פלואורסנציה, וטיפול במים על ידי כלור או סינון הפחית את טביעת אצבע הספקטרלית כמעט לחלוטין, ממצא המציין את הקשר בין זיהום מיקרוביאלי ופלואורסנציה.

הן בניסויים מבוקרים במעבדה והן במי שתייה אמיתיים, ספקטרוסקופיית פלואורסנציה הראתה יכולת לגלות חיידקים בריכוז של 1,000 CFU/ml המהווה את הסטנדרט התעשייתי הנדרש. יתר על כן, המחקר ממחיש את בשיפור ברגישות אשר ניתן להשיג באמצעות ניתוח הספקטרום התלת-ממדי המורכב של ספקטרוסקופיית פלואורסנציה לעומת שימוש באורך גל יחיד "Peak Picking". בניסוי מבוקר, אלגוריתם PLS שחושב על מפת עירור-פליטה מלאה (EEM ,Excitation Emission Map) בטווח של 400-210 ננומטר עירור ו-450-220 ננומטר פליטה הגדיל את הרגישות מ-10^5 CFU/ml עד 10 CFU/ml. שיטה זו גם מאפשרת הבחנה בין מינים שונים על פי טביעת האצבע הספקטרלית שלהם. מחקר זה ממחיש את הפוטנציאל הגבוה של ספקטרוסקופיה פלואורסנציה לזיהוי מהיר ומדויק של חיידקים במי השתייה.

. Kaiser Optical Systems - Raman Spectroscopy - A Tutorial.

(2013). תקנות בריאות העם - איכותם התברואית של מי־שתיה ומיתקני מי שתיה.

(2016). הנחיות לדיגום מים.

(2017). הנחיות המנהל להגשת תכנית, לתפעול וניטור מתקן טיפול במי שתיה.

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. Curr Environ Health Rep *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. (1995). Special techniques and applications. In Infrared and Raman Spectroscopy - Methods and Application, S. Bernhard, ed. (Weinheim, Federal Republic of Germany: VCH Verlagsgesellschaft).

Bernhard, S. (1995). General survey of vibrational spectroscopy. In Infrared and Raman Spectroscopy - Methods and Application, S. Bernhard, ed. (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. Int J Environ Res 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. Epidemiol Infect *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. J Hosp Infect *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.

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.

Duanis-Assaf, D., Duanis-Assaf, T., Zeng, G., Meyer, R.L., Reches, M., Steinberg, D., and Shemesh, M. (2018). Cell wall associated protein TasA provides an initial binding component to extracellular polysaccharides in dual-species biofilm. Sci Rep *8*, 9350.

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 Fau - Matthaus, C., Matthaus C Fau - Schmitt, M., Schmitt M Fau - Popp, J., and Popp, J. Advantages and limitations of Raman spectroscopy for molecular diagnostics: an update.

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. Symp Ser Soc Appl Microbiol, 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. J AOAC Int *78*, 1537-1542.

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.

George, S. (2001). Infrared and Raman Characteristic Group Frequencies (Chichester, England: John Wiley & sons Ltd.).

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.

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 Microbiol Rev *36*, 815-836.

Information, N.C.f.B. (2018). Tyrosine CID=6057, Phenylalanine CID=6305, Tryptophan CID=6140 (PubChem Compound Database).

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.

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., C, and Simorre, J.P. (2018). The Structure of Bacterial Cell Wall.

Krafft, C., and Popp, J. The many facets of Raman spectroscopy for biomedical analysis.

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.

L, Z., and S, E. (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.

Lakowicz, J.R. (2006). **Introduction to Fluorescence**. In *Principles of Fluorescence Spectroscopy*, J.R. Lakowicz, ed. (Maryland, USA: Springer).

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.

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. Anal Chem *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. Appl Environ Microbiol *78*, 5575-5583.

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

Mello, C., Ribeiro, D., Novaes, F., and Poppi, R.J. (2005). Rapid differentiation among bacteria that cause gastroenteritis by use of low-resolution Raman spectroscopy and PLS discriminant analysis. Anal Bioanal Chem *383*, 701-706.

Mizrach, A., Schmilovitch, Z.e., Korotic, R., Irudayaraj, J., and Shapira, R. (2007). Yeast Detection in Apple Juice Using Raman Spectroscopy and Chemometric Methods.

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. Molecular characterization of dissolved organic matter (DOM): a critical review.

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. Anal Chem *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. Adv Drug Deliv Rev *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. Sci Rep *6*, 29014.

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. Anal Bioanal Chem *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.

R., L.J. (2006). Protein Fluorescnce. In Principles of Fluorescence Spectroscopy (Singapore: Springer), pp. 529-570

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., *et al.* (2005). Chemotaxonomic identification of single bacteria by micro-Raman spectroscopy: application to clean-room-relevant biological contaminations. Appl Environ Microbiol *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. In <div aria-hidden="false" class="word-break layout-column" layout="column" ng-repeat="curValue in $ctrlgetValuesOfValueItems(value) track by $index" ng-show="$first || !detailisCollapsible || !detailshowMore" style="box-sizing: border-box; display: flex; flex-direction: column; word-wrap: break-word;"> <prm-highlight ng-if="!$ctrlisSuprima()" style="box-sizing: border-box;" terms="$ctrlhl(valuekeyt)" text="curValue">Faculty of Agriculture, Food and Environment</prm-highlight> (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. Appl Environ Microbiol *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. Int J Food Microbiol *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.

Walsh, J.D., Hyman, J.M., Borzhemskaya, L., Bowen, A., McKellar, C., Ullery, M., Mathias, E., Ronsick, C., Link, J., Wilson, M.*, et al.* (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. Int J Food Microbiol *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.*, et al.* (2009). Optical fingerprinting in bacterial epidemiology: Raman spectroscopy as a real-time typing method. J Clin Microbiol *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 / ed. by Phil Williams, Karl Norris.

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. Appl Spectrosc *58*, 33-40.

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. Anal Chem *86*, 1525-1533.