# Results

## 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. 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 108 and 107 CFU/ml (fig. 1).

Figure 1. An example of the Raman spectra of clean double-distilled water, E. coli at a concentration of 108 and B. subtilis at a concentration of 107 CFU/ml. Excitation by 785 nm laser, for 5 seconds, 3 scans were averaged.



Different attempts have been made to improve the data, by increasing time, cooling the samples to 4°C, 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 models was designed to try and expose underlying information.

### PLS model

Before starting to work on the PLS models, 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 imply on the complexity of the model, a general rule of thumb is that LVs of a good model should be between 3-10 (ref?), 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 R2; 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 Raman (280/360)
2. Centering
3. Scaling
4. 1/signal, log10(1/signal)
5. 1st derivative

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

Table 1. Comparison of different preprocessing approaches: RMSE and R2 values are calculated on validation set, n=74. LVs = Latent variables. The best method is the one that shows the highest R2 with the lowest possible RMSE, and is in **bold**.

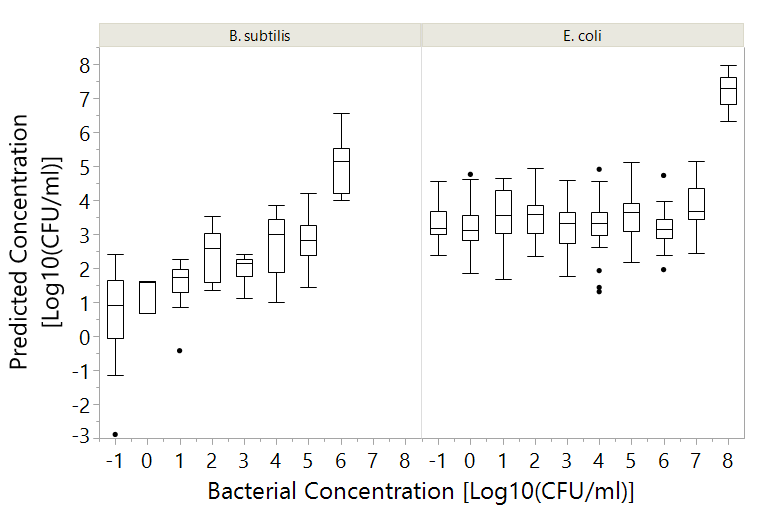
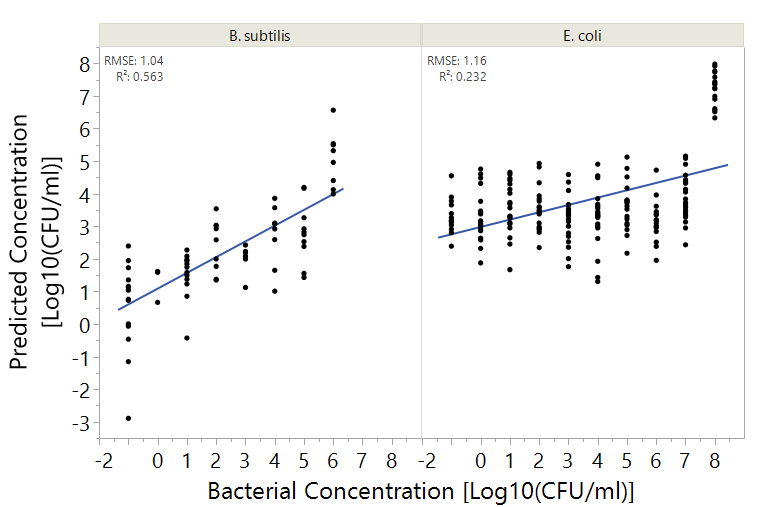
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Preprocessing 1 | Preprocessing 2 | LVs | RMSE | R2 |
| 1 | **None** | **None** | **7** | **1.04** | **0.56** |
| 2 | None | Centering | 7 | 1.14 | 0.58 |
| 3 | None | Scaling | 3 | 0.4 | 0.08 |
| 4 | None | Centering + Scaling | 5 | 1.35 | 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.

### 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 108 CFU/ml, *B. subtilis* prediction models enabled confident detection at 104 CFU/ml (fig. X).

Figure 2. 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 ussed 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.



**\***

**\***

**\***

**\***

### 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 (fig Y). 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 (C in fig Y) 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 (ref).

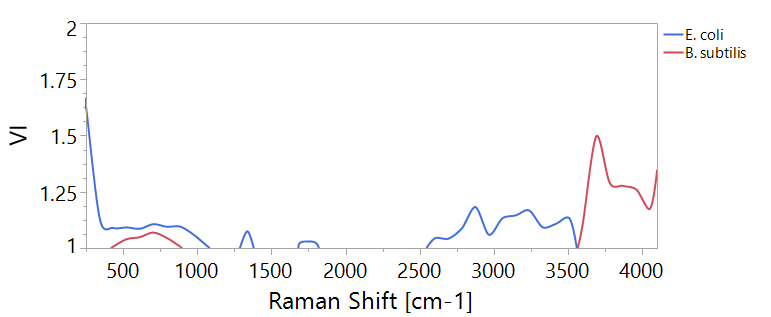


Figure 3. Variable importance (VI) plot of PLS models for quantification of bacteira according to Raman spectra. The plots show only values >1. Region (A) sinifies the

B

C

A

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 108 CFU/ml (n=42)
2. *B. subtilis* at 106 CFU/ml (n=51)
3. Clean water samples (n=67)

Table 2. Confusion matrix of classification of bacteria by PLS-DA, based on Raman spectra, at concentrations 108 and 106 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 |

A PLS-DA model was calculated, in an attempt to classify these samples correctly, with the same training-validation ratio of 6:4. Table X 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).

## Fluorescence

### 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, and in order to choose the best approach we compared the R2 values of different approaches with our approach based on the local maxima of the peak calculated in our experiment (Figure 1). The results are summarized in table 1, 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.

Table 3. Different approaches to describe the peak of tryptophan and their correlation with tryptophan concentration. All R2 values are significant p<0.01

|  |  |  |  |
| --- | --- | --- | --- |
|  | Tryptophan  Ex-Em | Raman  Ex-Em | R2 |
| 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, I normalized according to Bridgeman | 0.84 |
| Sorensen 2018 | Ex/Em peak at **280** ± 15 / **365** ± 27.5 nm. | Not reported, I normalized according to Bridgeman | 0.83 |
| Local Maxima | 275-362 | According to Simelane | 0.85 |

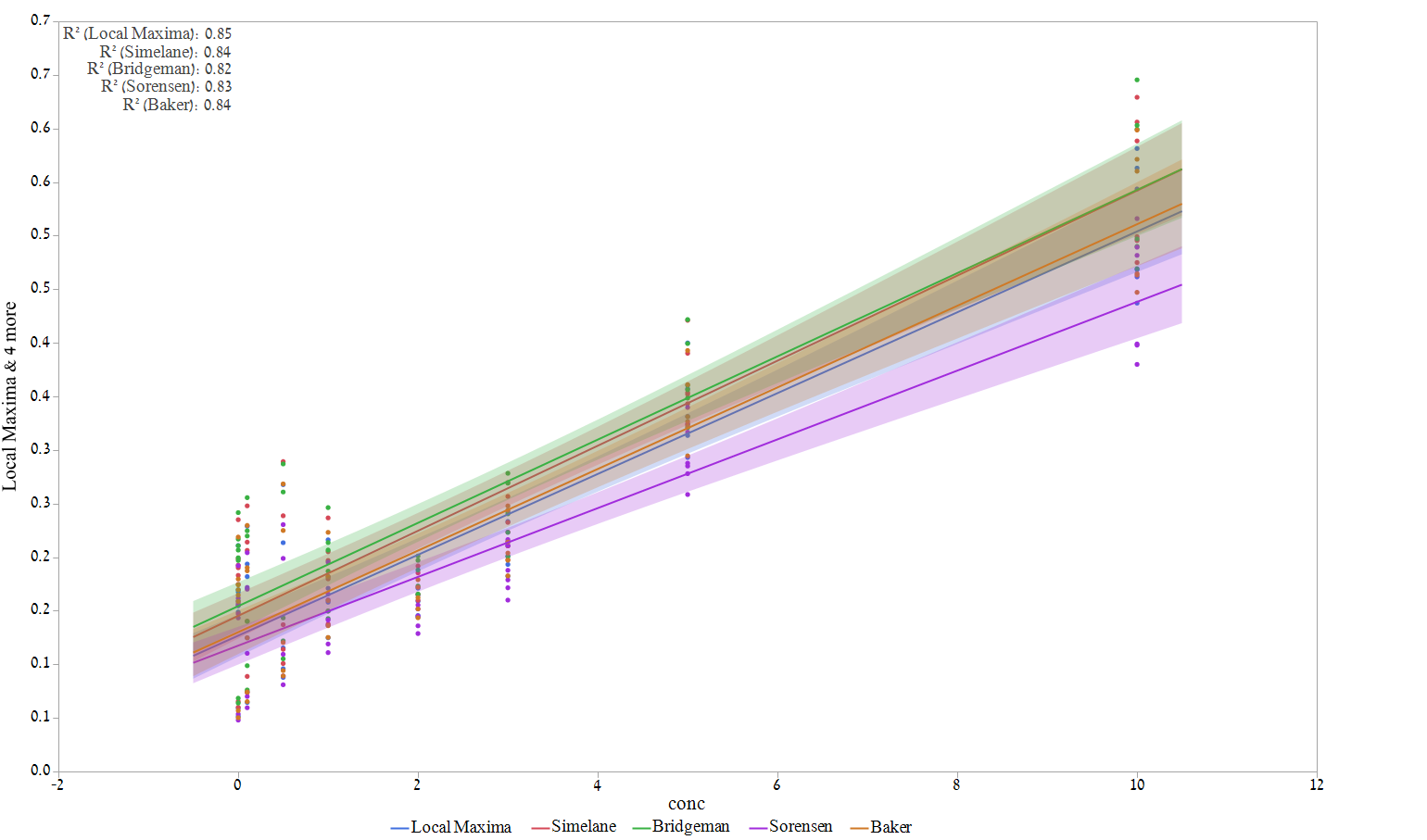


Figure 4 Correlation between different tryptophan fluorescence approaches. N=48 (51-3 excluded). Particular wavelengths of each approach elaborated in table. Halo is the confidence fit at α=0.05. [Script: "tryptophan calibration"]

Further, where applicable, tryptophan concentration will be gives as calculated from the regression line according to the following formula:

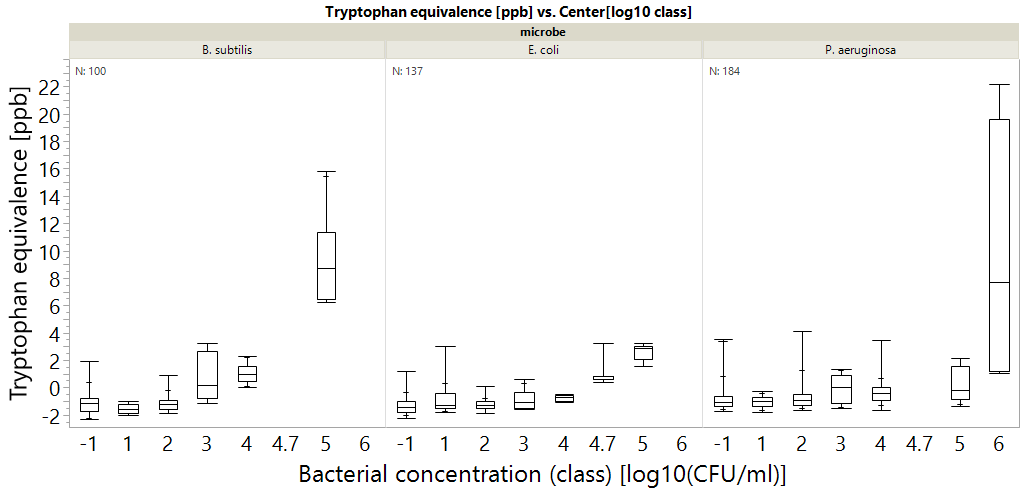
Normalized fluorescence = 0.1301063 + 0.0380656\*[Tryptophan ppb] 🡪

Tryptophan ppb = (Normalized fluorescence – 0.1301063)/0.0380656

Where the normalized fluorescence =

#### 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 measurements to detect bacteria in known concentration. Whenever the concentration is given as "-1" it is a concentration of 0 CFUs/ml. This is because you there is no log10(0). Figure 2 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 103 CFU/ml, however, the major difference in fluorescence can only be observed around 2 ppb of tryptophan.



\*

\*

\*

\*

\*

\*

\*

\*

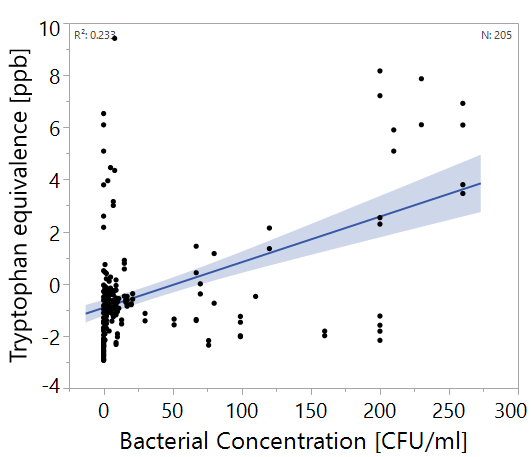
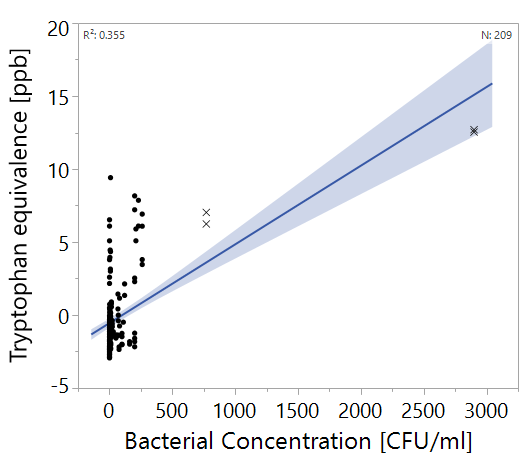
\*

Figure 5 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. Script saved as " Tryptophan classic of specific pathogens"

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

The correlation between fluorescence measurement in tryptophan equivalence and heterotrophic plate counts (HPCs) in drinking water is shown in figure 3. The right panel includes outliers 770, 2900 CFU/ml (marked as Xs) and the left does not. R2 are always significant according to a student's t-test (p-value<0.0001). Figure 4 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 6 - 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.



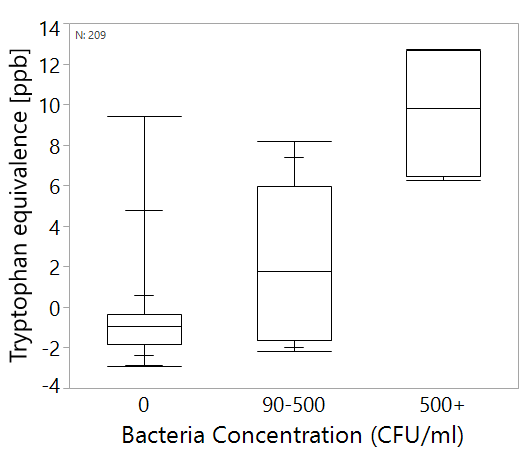


Figure 7. Tryptophan fluorescence measurements correlated heterotrophic bacteria in drinking water. All groups are significantly differenct according to a Wilcoxon test, p<0.01. n indicated at the top. Box plots shows the middle quantiles in box, with whiskers for bottom and top values. The crosshairs 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 has very strong fluorescence (fig. 5) before treatment, however, chlorination and lab filtration treatments weaken the signal almost completely.

Figure 8. EEM of shimron 7 water sampled on 22/5/2018. (A) raw water, (B) chlorinated water, (C) filtered raw water, (D) filtered chlorinated water.

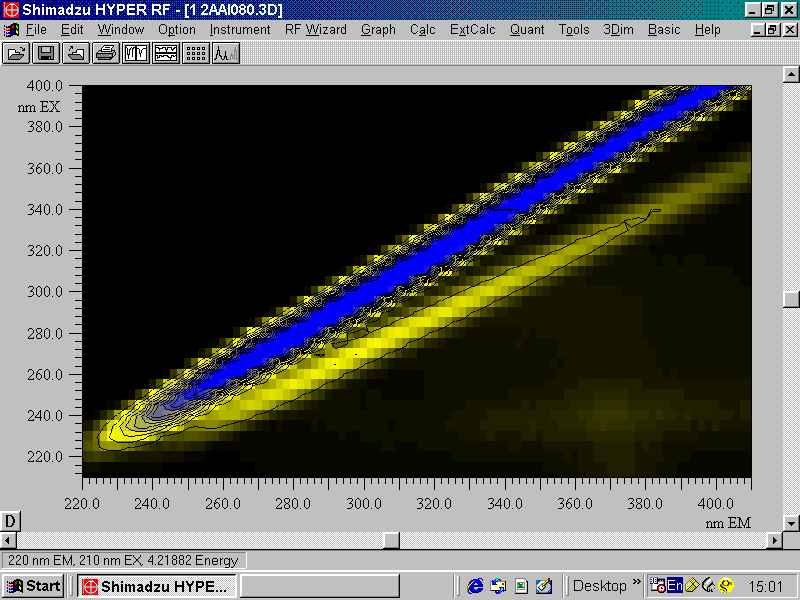
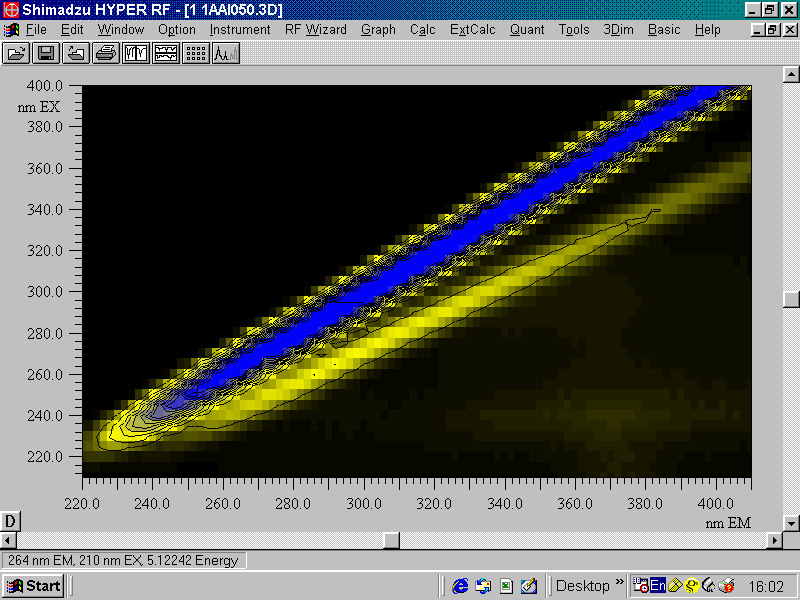
A

B

C

D

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 (fig. 6). 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 fig. 6) but no significant signals in the protein or any other regions.



CFUs/ml = 16

CFUs/ml = 140

Figure 9. EEMs of samples with both high and low 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.

### 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 a set protocol, different preprocessing approaches were tested on the *E. coli* data set (n= 137) (table 3). We tested 5 types of preprocessing elements:

1. Normalization to Raman (280/360)
2. Centering
3. Scaling
4. 1/signal, log10(1/signal)
5. 1st derivative

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

Table 4. Comparison of different preprocessing approaches: RMSE and R2 values are calculated on validation set, n=60. LVs = Latent variables. The best method is the one that shows the highest R2 with the lowest possible RMSE, and is in **bold**. Other format is in the "graphs edited out" section

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Preprocessing 1 | Preprocessing 2 | LVs | RMSE | R2 |
| 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. 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

This was tested on the *E. coli* sample. The idea is that if only a single excitation wavelength can be used, operational and computational costs can be saved. In this analysis we compared the detection threshold and R2 of regression of 4 possible detection methods:

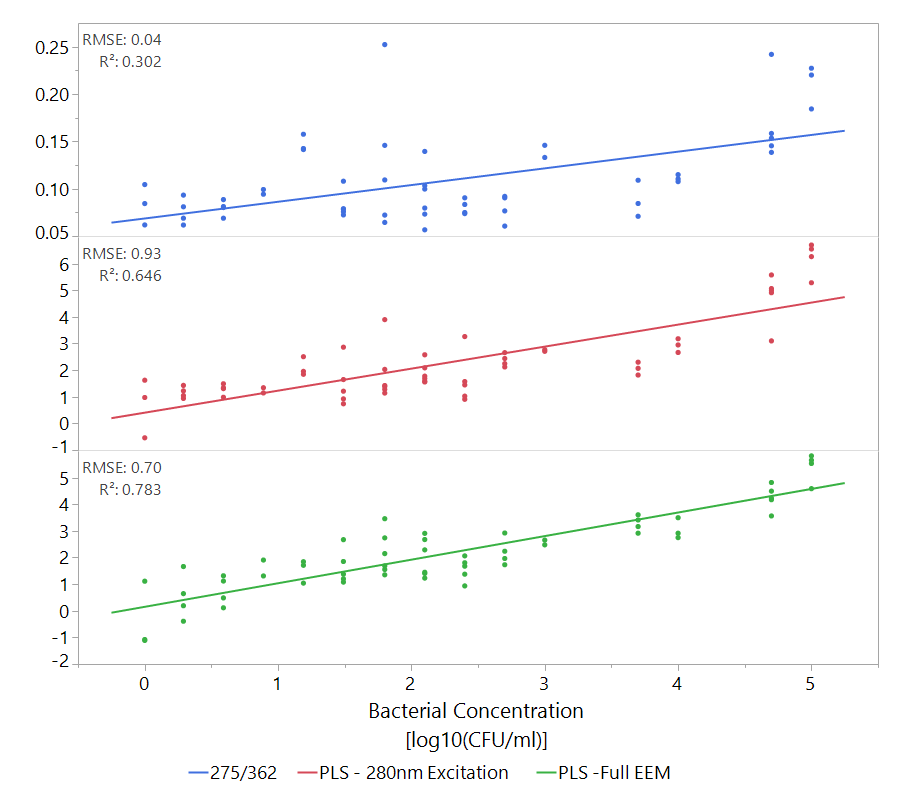
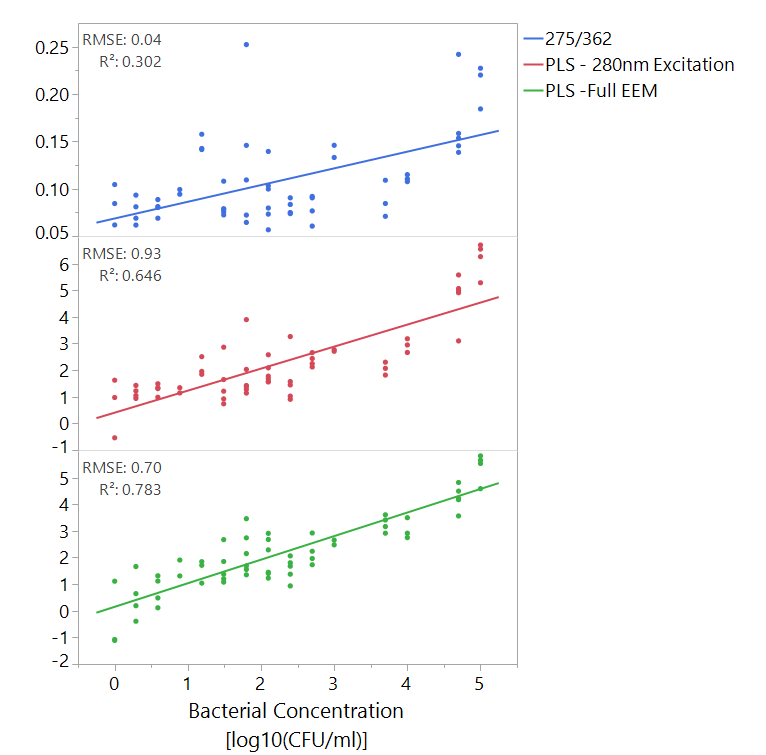
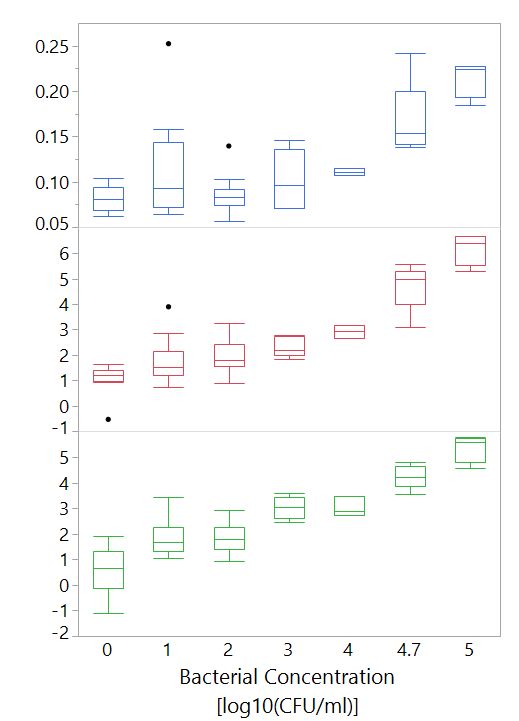
1. Single wavelength method – the intensity at 280/350
2. Single wavelength method – the intensity at 275/362
3. Single wavelength excitation whole spectra – the light emission at all wavelength (220-450 nm) excited by 280 nm light
4. The entire EEM with excitation between 210-400 and emission between 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 the first is only applicable to some of the methods, while the second is more robust (table 4, Figure 6).

Table 5 Comparison of methods for analysis. Analysis only includes the validation sets. Detection threshold is when p<0.01 between concentration and control. n=60

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **R2** | **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]**  **Log10(CFU/ml)** | **Detection Threshold**  **[Tukey Kramer] Log10(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 10. Left: regression of bacterial concentration with different detection methods, Right: box plot of the data as ussed 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*) groups 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 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 5.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Microbe** | **Single wavelength (280 nm)** | | | | **Full EEM** | | | |
|  | **RMSE**  **[Log10CFU/ml]** | **R2** | **Detection threshold (Wilcoxson)** | **Detection threshold (Tukey-Kramer)** | **RMSE**  **[Log10CFU/ml]** | **R2** | **Detection threshold (Wilcoxson)** | **Detection threshold (Tukey-Kramer)** |
| **E. coli**  **(N=58)** | 0.82 | 0.68 | 102 | 103 | 0.76 | 0.74 | 10 | 10 |
| **B. subtilis**  **(N=47)** | 0.84 | 0.58 | 105 | --- | 1.01 | 0.65 | 104 | 103 |
| **P. aeruginosa**  **(N=80)** | 0.86 | 0.23 | 106 | 106 | 1.04 | 0.55 | 10 | 10 |
| **All bacteria together**  **(N=185)** | 0.91 | 0.34 | 103 | --- | 1.06 | 0.48 | 100 | 100 |

Table 6. Comparison of PLS approaches ability to detect different bacteria. All R2 values are statistically significant. Wherever possible, both Wilcoxon and Tukey test were performed.

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

In this experiment, 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 figures 7 and 8.

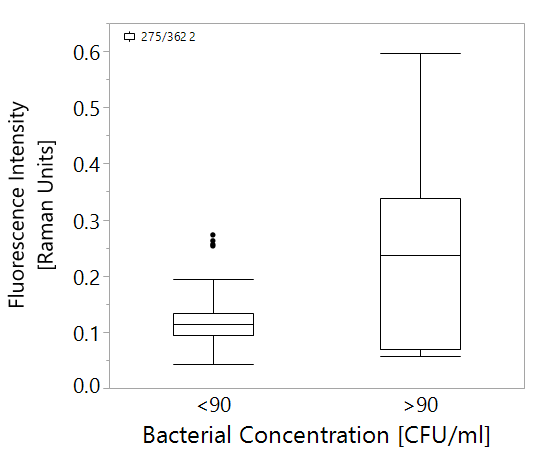
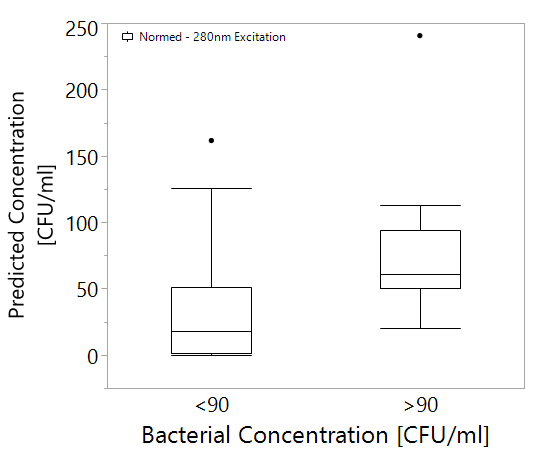
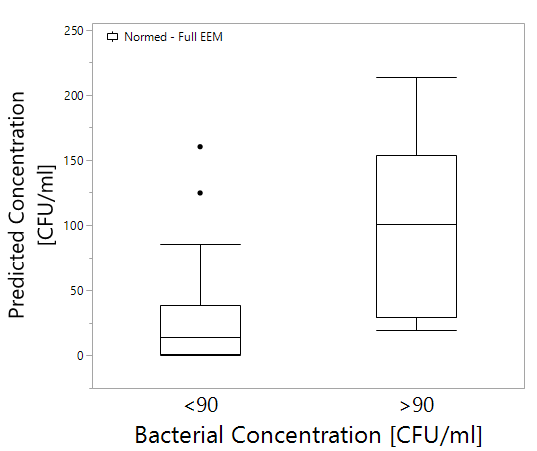


Figure 11. Boxplots of fluorescence intensity, prediction of bacterial concentration using either 280 nm excitation or full EEM against bacterial concentration categories. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown. Asterixes signify p<0.01 of Wilcoxon test. N = 70, "<90" category n=54, ">90" category n=16.

P of both is smaller than 0.0001 so I am not showing it.

\*

\*

275/362

Full EEM

PLS – 280 nm Excitation

\*

Figure 7 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 8), 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.

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

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

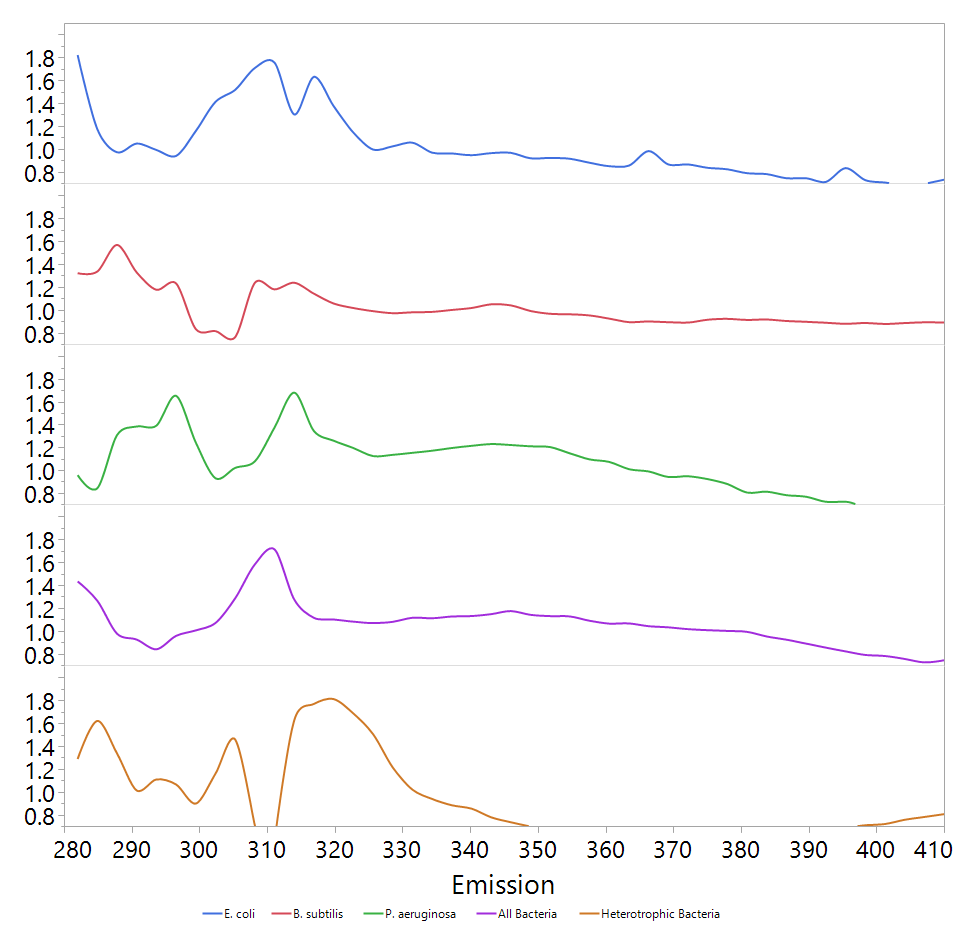
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 280 nm Excitation | Prediction [CFU/ml]  κ=0.213 | | | | |
| Actual  CFU/ml |  | <90 | >90 | Total |
| <90 | 50 | 4 | 54 |
| >90 | 12 | 4 | 16 |
| Total | 62 | 8 | 70 |

Figure 12. 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.

### 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 (JMP manual, Eriksson).

For this analysis, the VI values were plotted from the 280nm model (fig. 9). 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.



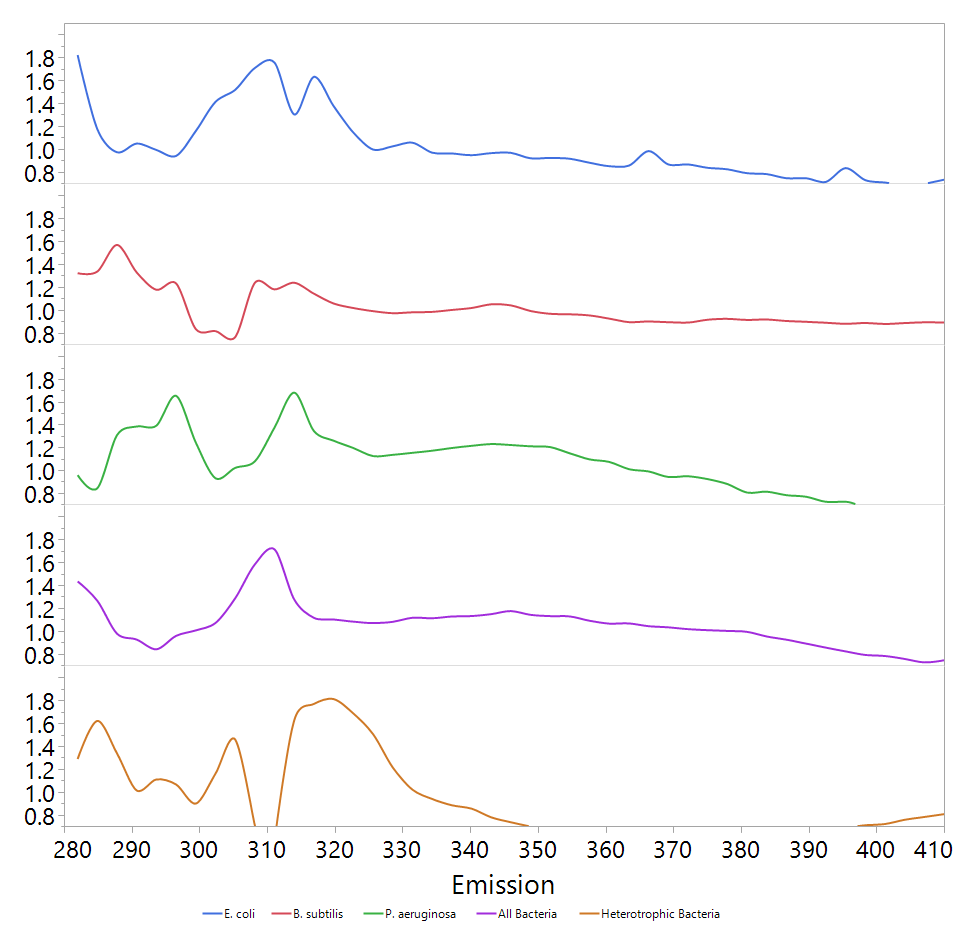
C

D

A

B

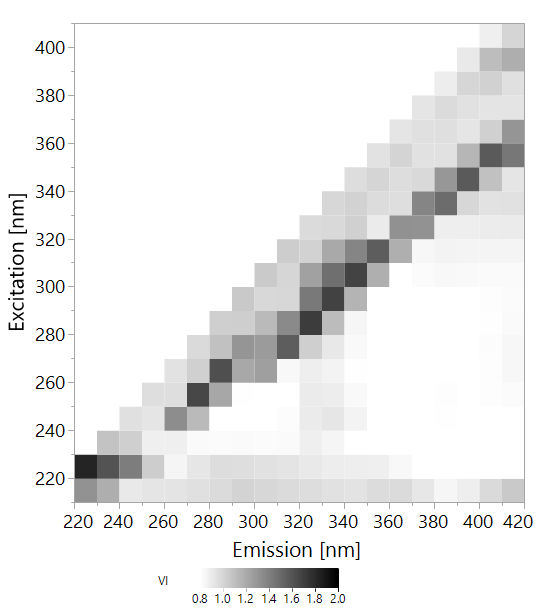
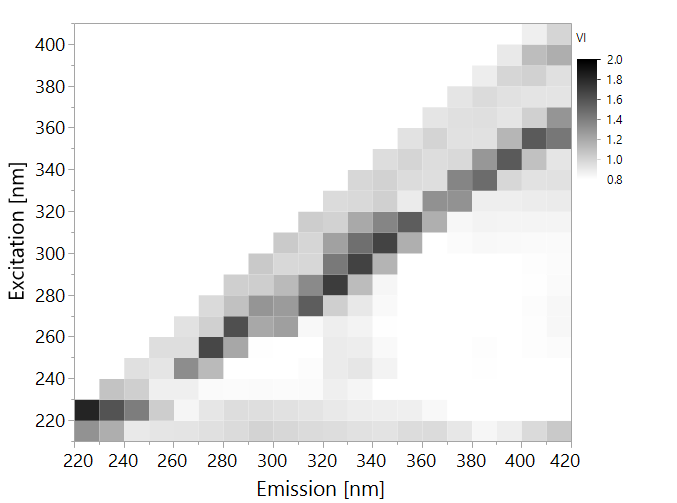
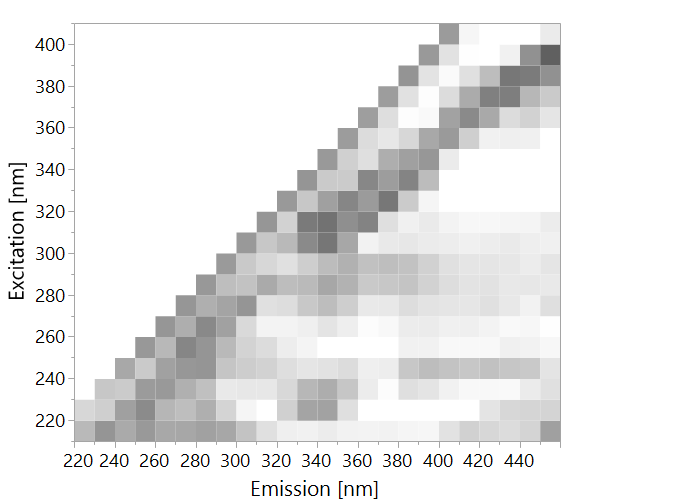
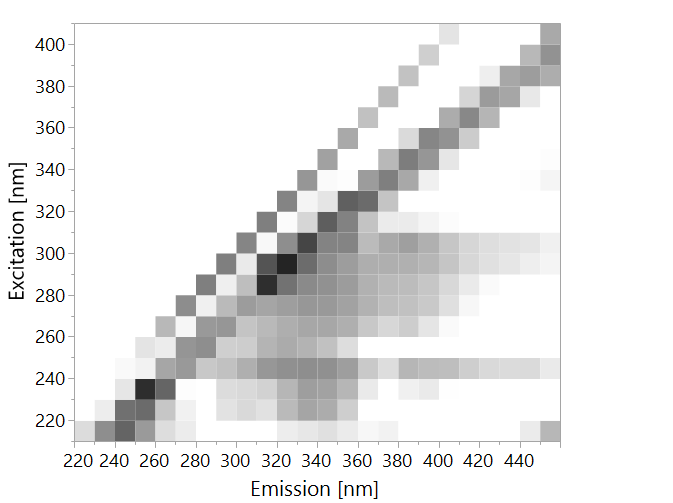
Figure 13. Variable importance (VI) plots of PLS models made from 280 nm excitation spectra. The plots show only values >0.8. Shades a,b,c and d indicate areas of peaks which may be indicative bacterial concentration.



VI

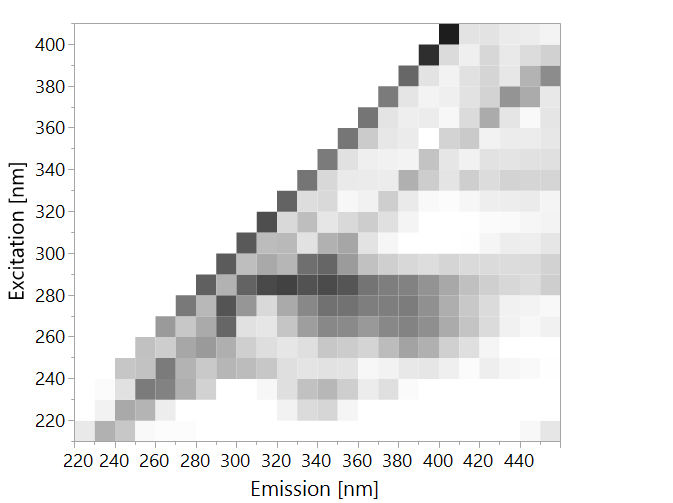
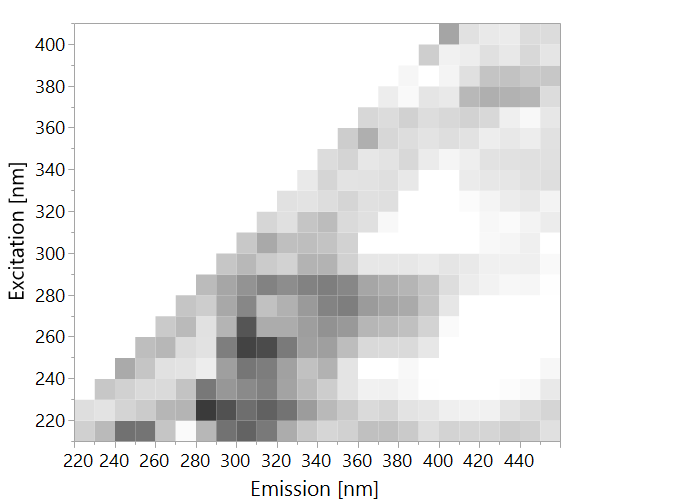
Emission [nm]

When the VI values of entire EEM, at excitation range of 210-400 nm and emission of 220-410 nm were plotted (fig. 10), 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.



E. coli

Figure 14. Variable importance (VI) plots of PLS models made from entire EEMs. The plots show only values >0.8.



B. subtilis

All Bacteria

P. aeruginosa

Heterotrophic Bacteria

### Differentiation between different species

In order to predict which bacteria is in the water, samples were divided into 4 categories:

1. *E. coli* >104 CFU/ml
2. B. subtilis > 104 CFU/ml
3. P. aeruginosa > 104 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 6 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 7. Confusion matrix of classification of microbes by PLS-DA, based on entire EEMs, at concentrations >104 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 |