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 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, 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 1. Correlation between tryptophan fluorescence and HPCs in groundwater samples and. The right graph shows the entire data set (n=209) and the left one includes only the data with CFU<300. All regressions are statistically significant, with p-value < 0.0001 using a Student's t test.

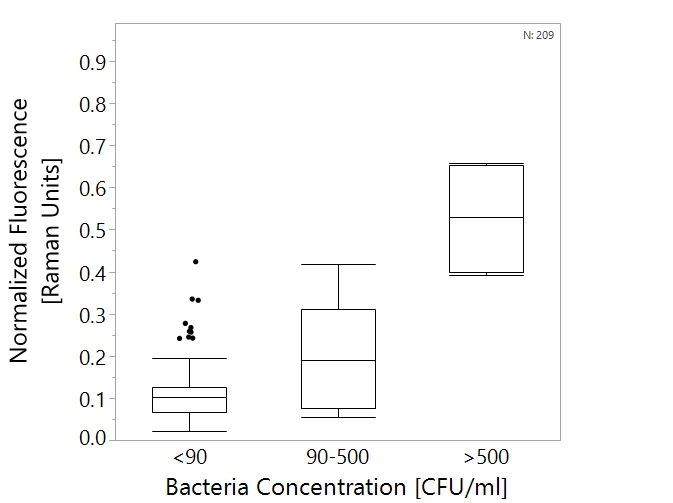
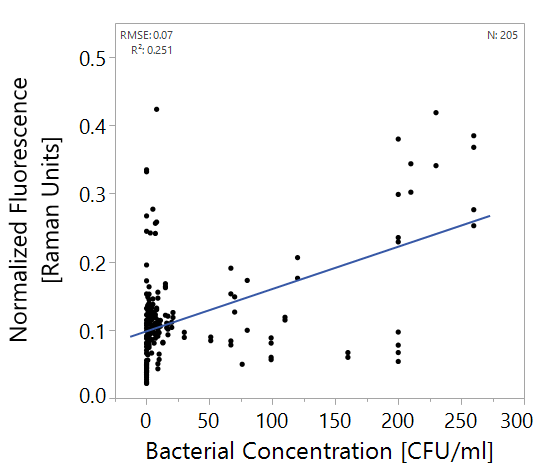
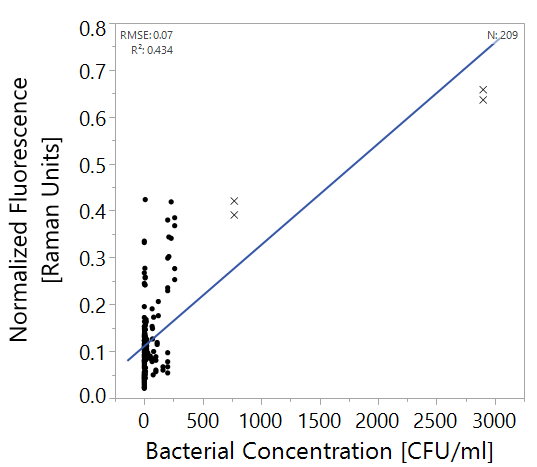


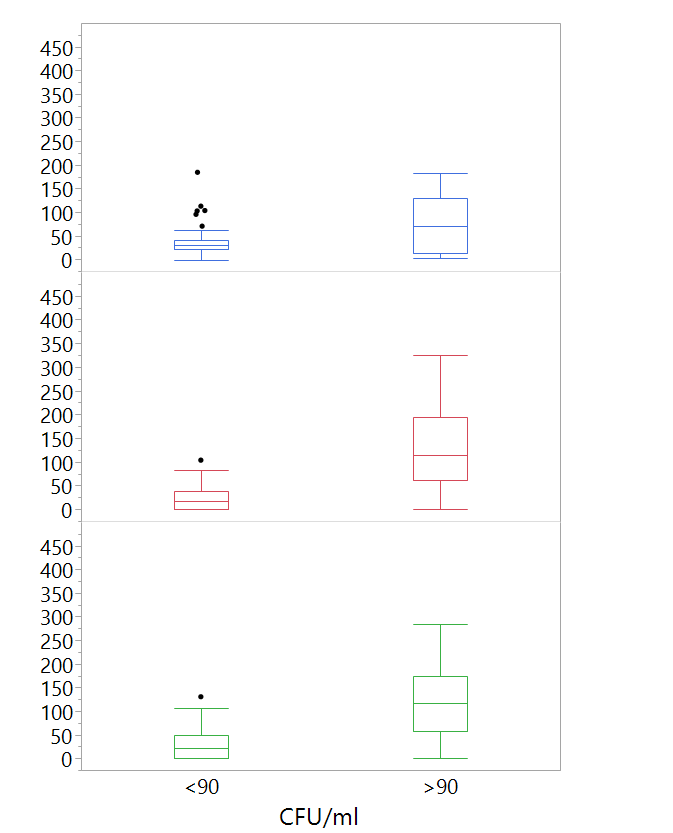
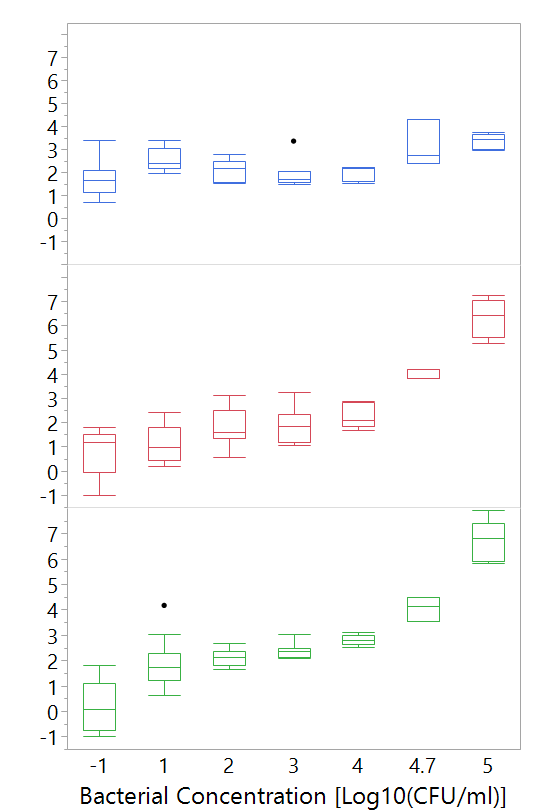
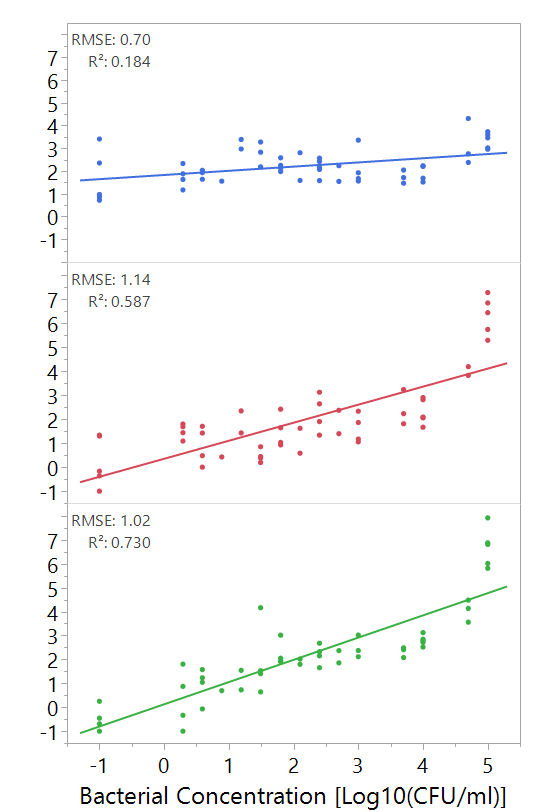
Figure 2. Box plots of tryptophan-like fluorescence measurements vs. heterotrophic bacteria counts in groundwater. All groups are significantly differenct from each other according to a Wilcoxon test, p<0.01. n=209. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown.

**\***

**\***

Figure 3. 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=53.

Predicted Bacterial Concentration [log10(CFU/ml)



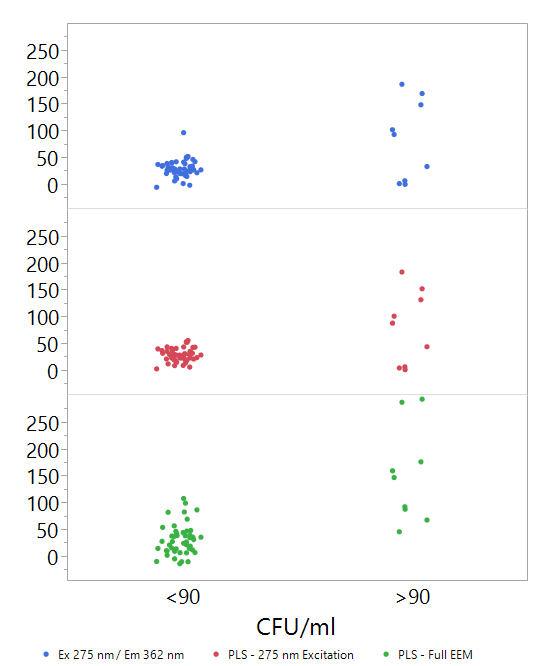
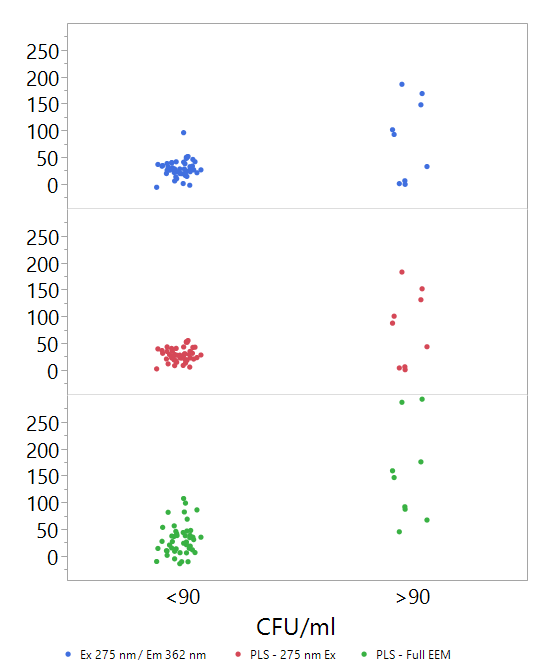
Predicted Bacterial Concentration [log10(CFU/ml)

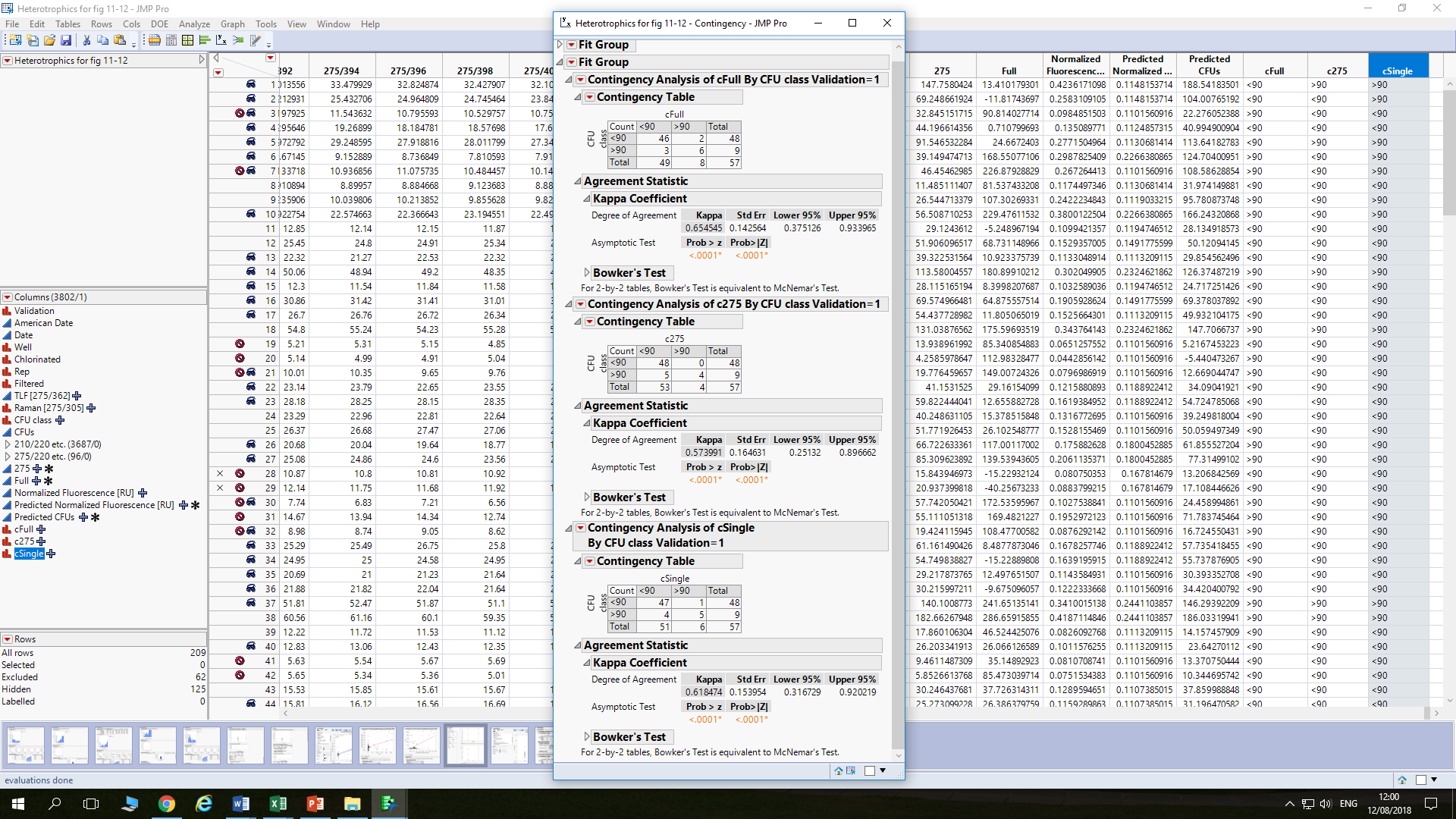
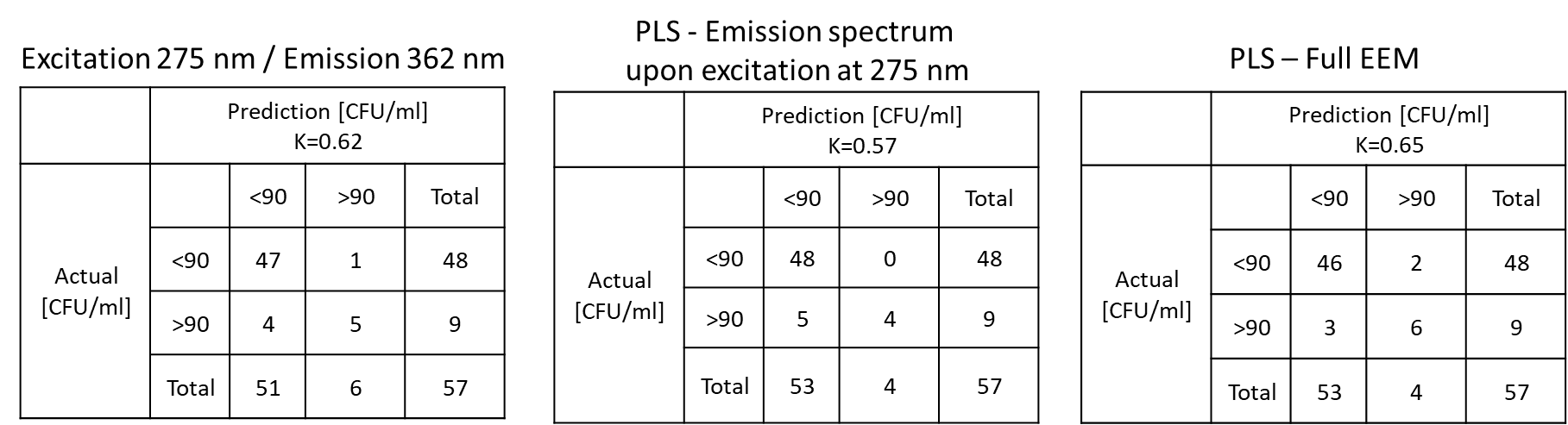


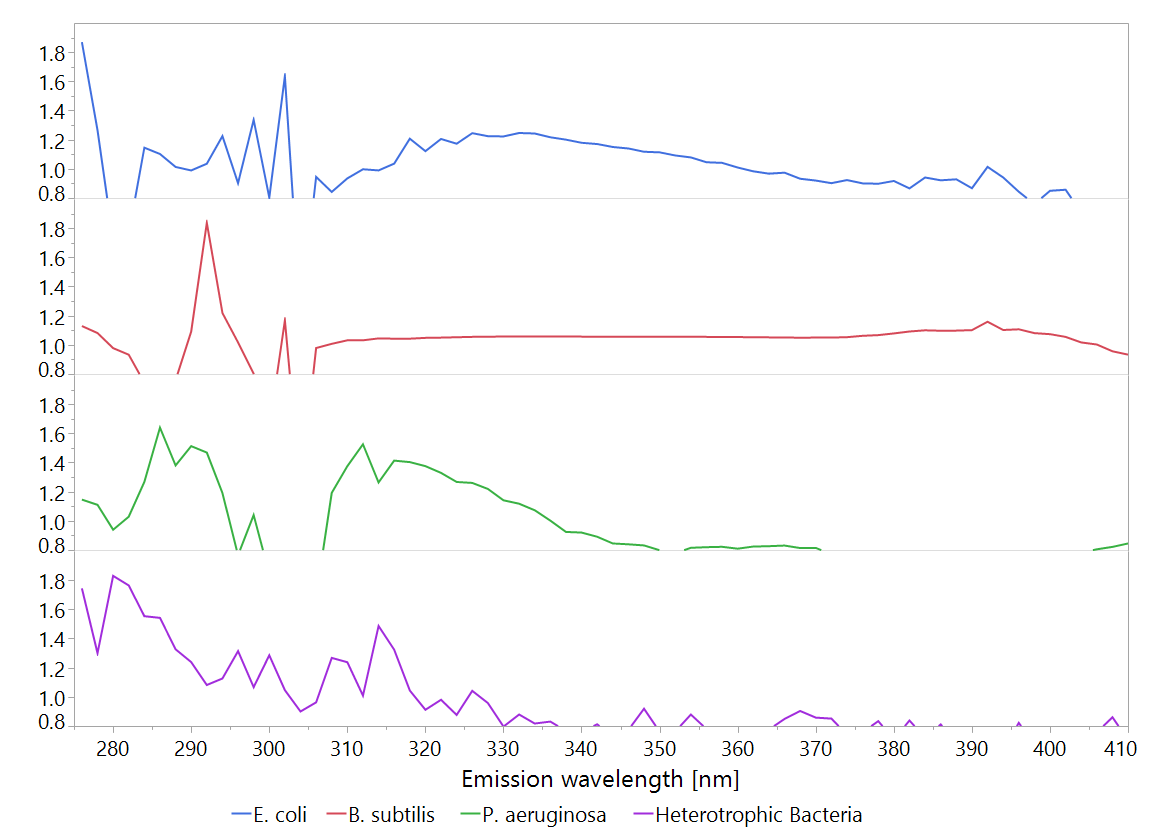
Figure 4. Box plots of predicted heterotrophic bacteria concentrations using different ??? versus real heterotrophic bacteria counts in groundwater. Astrixes mark significant difference between groups according to a Wilcoxon test, p<0.01. n=57. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile, outliers are shown.

**\***

**\***







A

Figure 20. Variable importance (VI) plots of PLS models made from the emission spectrum upon excitation at 275 nm spectra. The plots show only values >0.8. Shades A-F indicate areas of significan signals which may be indicative bacterial concentration.

VI

B

C

D

E

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

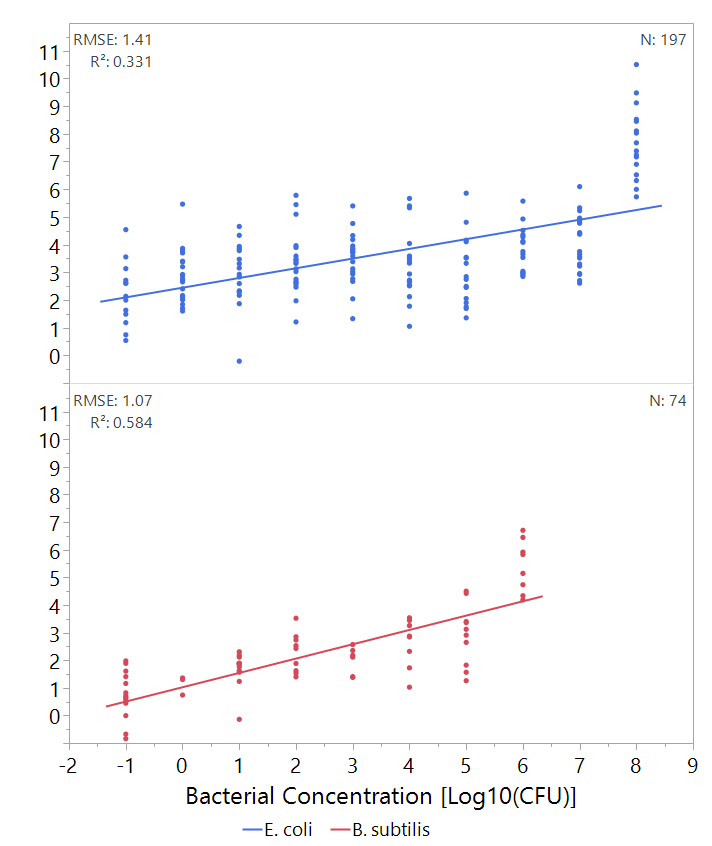
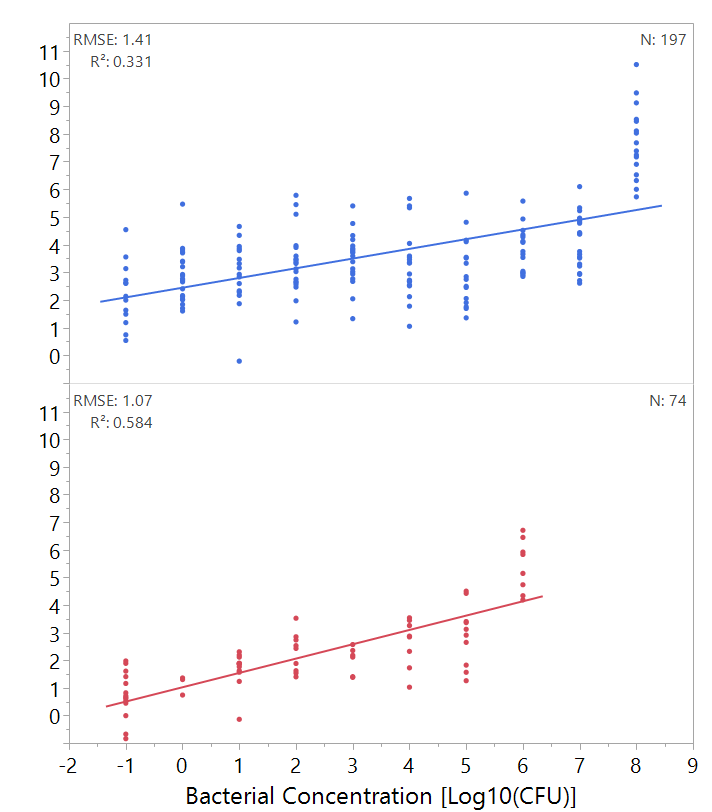
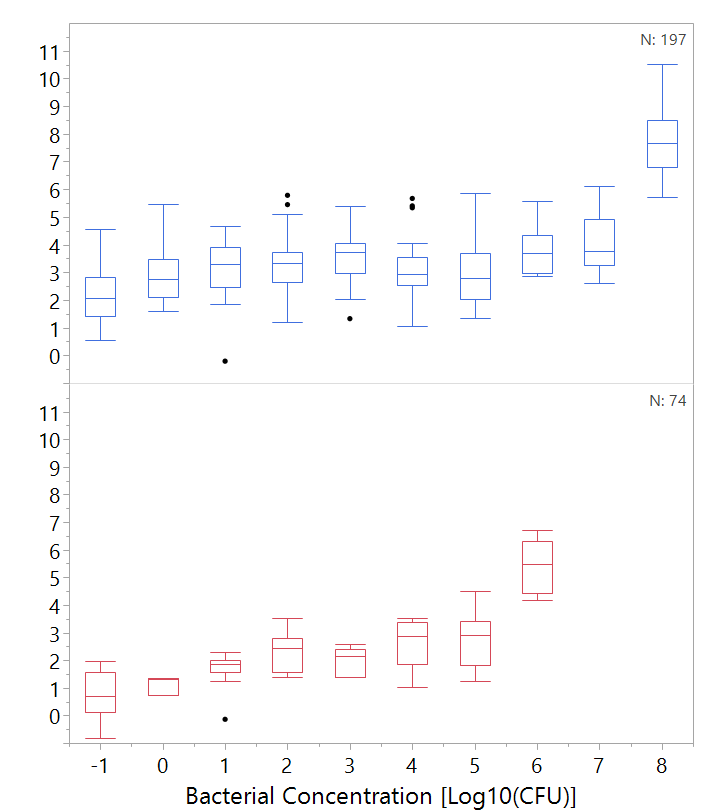
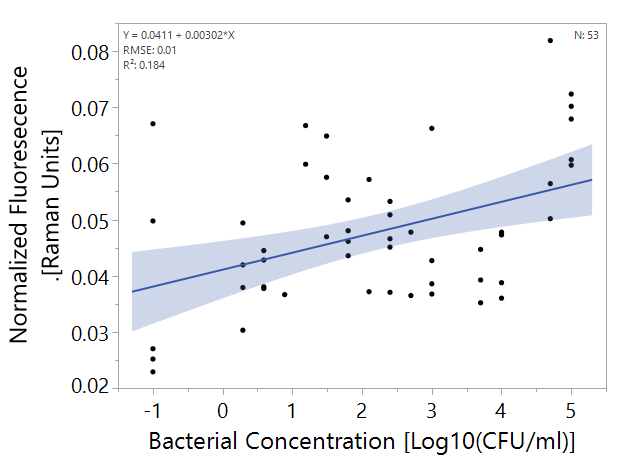


Figure 7. Top: regression of predicted bacterial concentration according to PLS model against real bacterial concentration of E. coli and B. subtilis. Bottom: box plot of the data as used for threshold analysis. Boxes illustrate median and interquartile range (IQR), whiskers indicate 25th and 75th percentile,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

Predicted Bacterial Concentration

Log10[log10(CFU/ml)

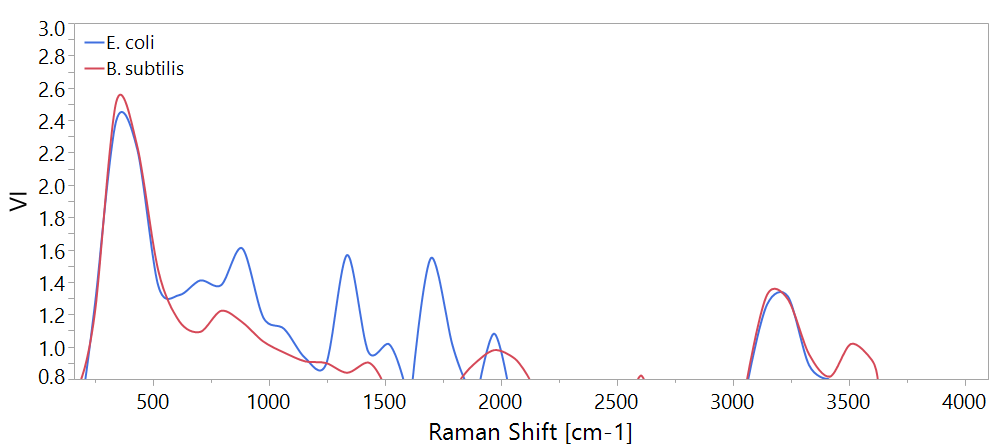


Figure 8. Variable importance (VI) plot of PLS models for quantification of bacteira according to Raman spectra. The plots show only values >1. A is the fingerprint region, B is the functional group region.

B

A