MWRD PAA - Preliminary Data Analysis

Kate Newhart

3/28/2019

# Introduction

The Robert W. Hite Treatment Facility, operated by the Metro Wastewater Reclamation District (MWRD) of Denver, CO, treats ~130 million gallons per day (MGD) of wastewater produced by ~2 million people from the Denver-metro area and is the largest wastewater treatment facility in the Rocky Mountain west. In an effort to reduce the cost of disinfection, a peracetic acid (PAA) system was installed to replace the existing chloramine system. However, due to variable influent *E. coli* concentrations to the disinfection system, it has been difficult to optimize the dosing of PAA to keep below *E. coli* limits of 126 (most probable number [MPN])/100 mL based on a 30-day geometric mean and 252 MPN/100 mL based on a 7-day geometric mean. In practice, PAA is overdosed to ensure that MWRD is meeting it’s discharge limit. The goal of this work is to identify correlations between upstream operating conditions in the secondary activated sludge system, *E. coli* concentrations, and PAA dosing.

# Goals

Design a PAA disinfection dosing system that accounst for:

* Upstream secondary treatment performance
* Flowrate through the disinfection basin (i.e., hydraulic retention time or HRT)
* Flow conditions in receiving water body (e.g., low, mid-range, high, dry, moist)

# Questions

1. What effects pre-disinifection *E. coli*?
2. What effects PAA disinfection efficiency?

##### 

# Procedure

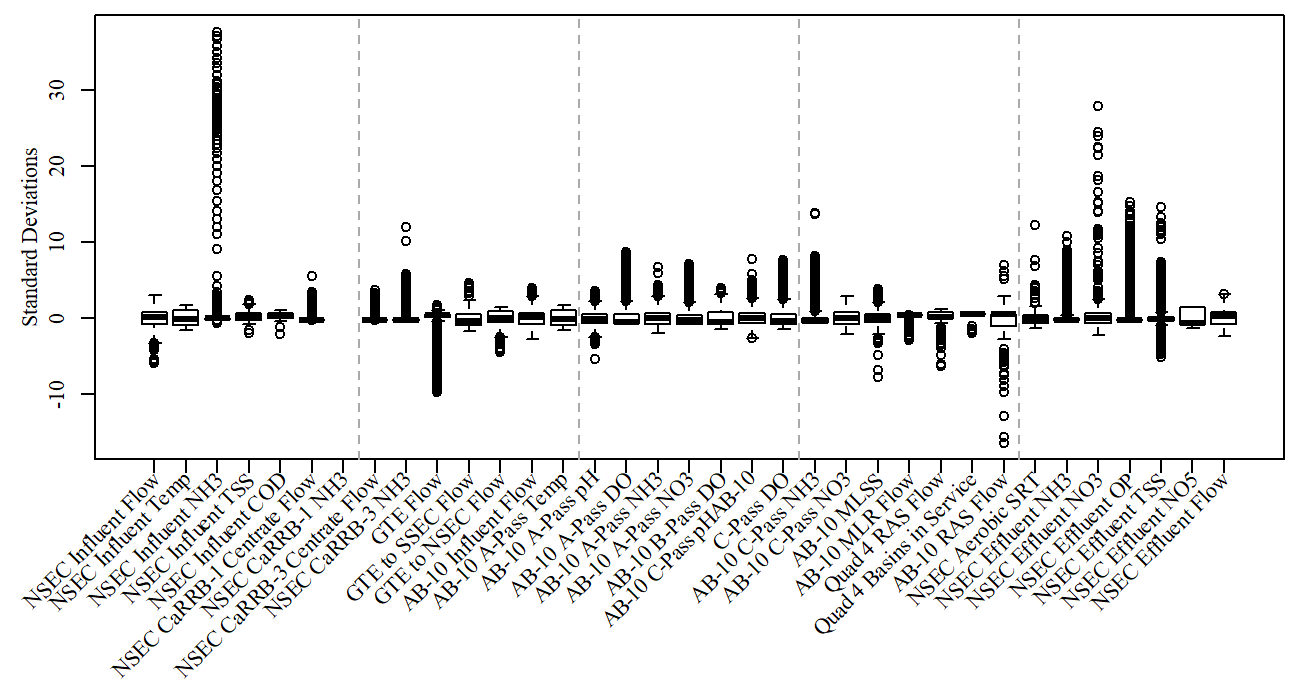
Data was provided by MWRD at a variety of intervals and locations in the treatment process:

|  |  |  |  |
| --- | --- | --- | --- |
| Location | Frequency | Source | Variables |
| North Secondary | 15 min | Sensors | Influent/recirculation/effluent flow, Temperature, Ammonia, TSS, COD, pH, DO, SRT, Nitrate, Ortho-P, Nitrite |
| North Disinfection | 15 min | Sensors | Influent flow, PAA residual, PAA pump flow, PAA setpoint, HRT |
| North Disinfection | Daily | Grab | PAA dose, Upstream residual, Pre-disinfection E. coli, Effluent flow, HRT, Effluent E. coli, CT |
| North Secondary Influent | Daily | 24 hr composite | BOD, Ammonia, TSS |
| North Secondary Effluent | Daily | 24 hr composite | cBOD, Ammonia, TSS |
| North Secondary Influent | 2-3 days | 24 hr composite | COD, Nitrate-nitrite, C:N, C:P, TP, TIN, TKN, TN |
| North Secondary Effluent | 2-3 days | 24 hr composite | Alkalinity, Nitrate-nitrite, C:N, C:P, TP, TIN, TKN, TN |
| North Secondary | 2-5 Days | Grab | SVI, TSS, VSS |
| North Secondary Influent | Weekly | 24 hr composite | Alkalinity, OP |
| North Secondary Influent | Weekly | 24 hr composite | COD, OP |

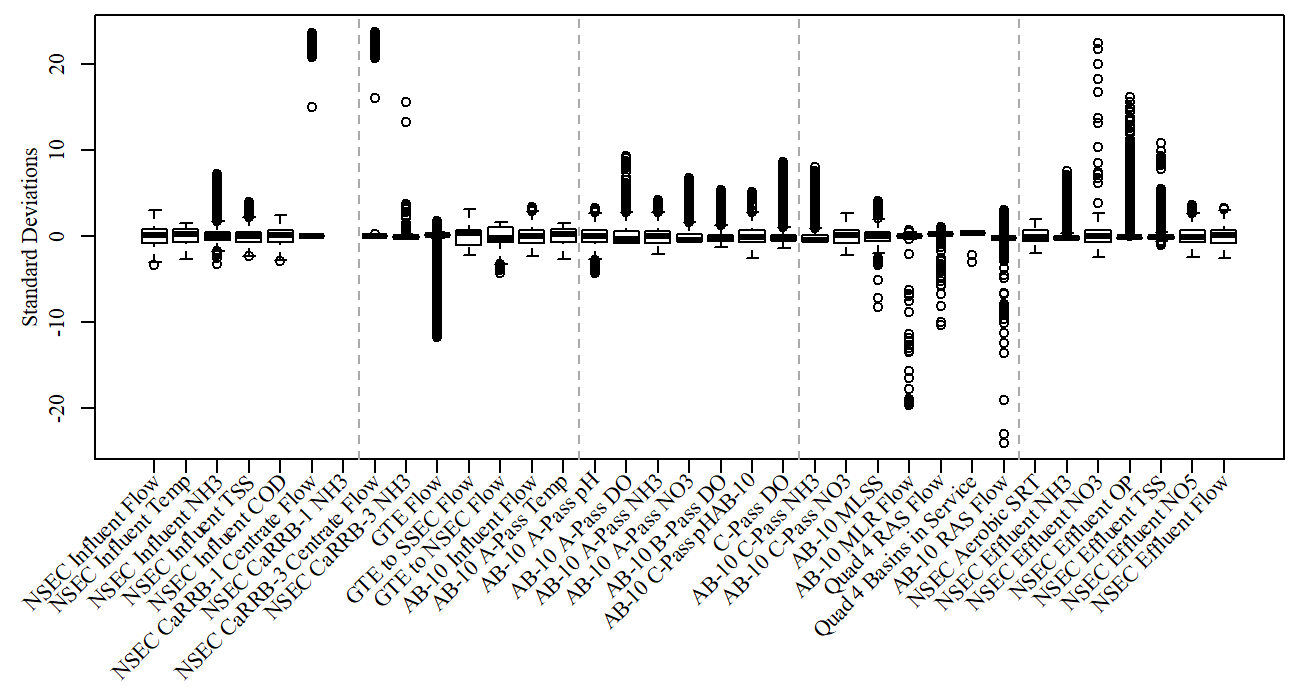
# Data cleaning

If the distribution of each variable is assumed to be univariate normal and scaled (i.e., zero mean, unit variance), boxplots can be constructed to visualize the range of observations in the dataset. The existance of numerous outliers, heavily shifted on either side of the variable’s median (Figures S1-S3) indicate that the majority of water quality variables are not normally distributed.

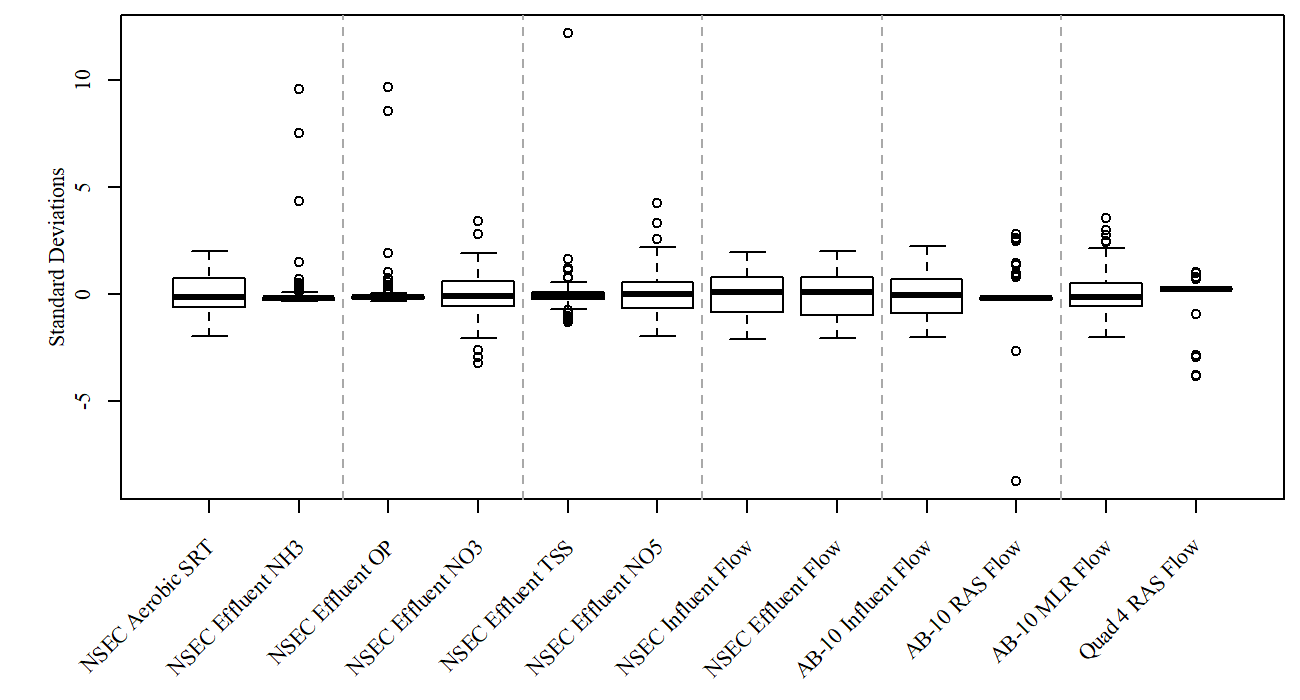
To achieve the goal of predicting *E. coli*, a log-transformed pre-disinfection *E. coli* grab data was merged with a reduced North secondary dataset (Figure 1) and data from a visual spectrum analyzer (Figure 2).



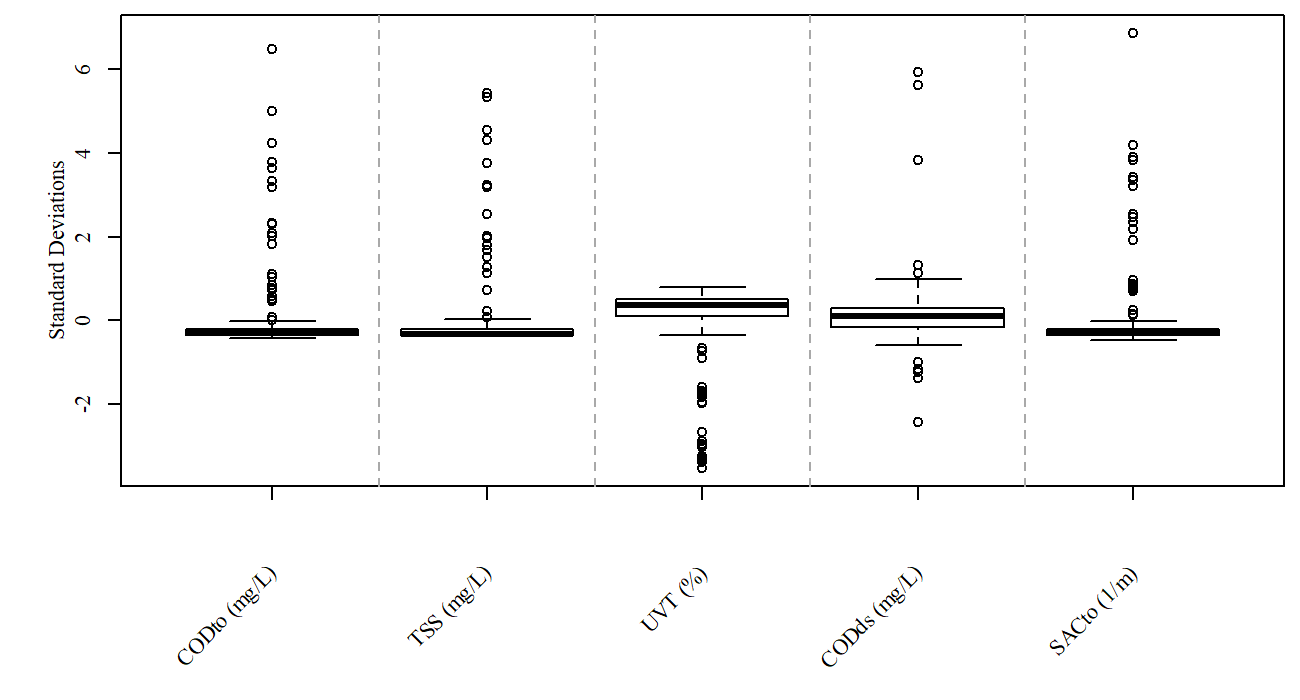
**Figure 1a.** Raw north secondary data.



**Figure 1b.** Cleaned north secondary dataset from 2018-06-01 / 2018-12-01.Removed outliers (such as negative flows and unrealistic sensor values) and missing values (such as influnet COD/TSS and effluent NO5).



**Figure 1c.** Final data after filtering for values <15 minutes from *E coli* sampling time.



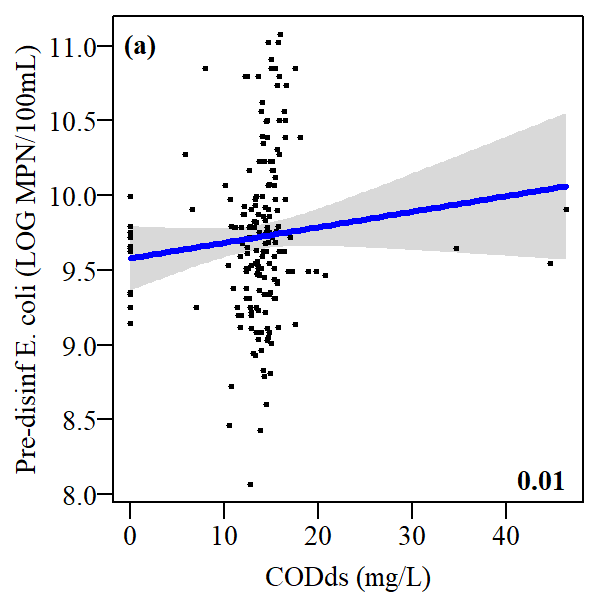
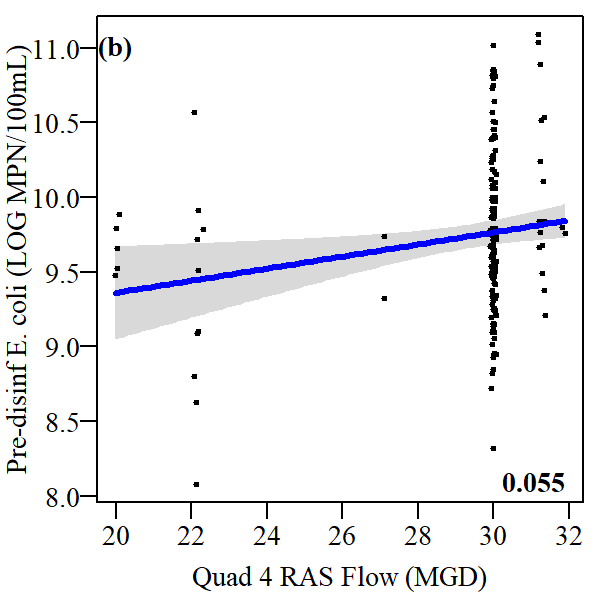
**Figure 2.** Boxplots of water quality variables immediately upstream of PAA dosing used to predict *E. coli* from 2018-06-01 / 2018-12-01.

##### 

# Methods

## Linear Regression

In simple linear regression, a model is constructed of a response variable (**Y**) that is a linear function of other variables (*xi*). The linear regression model assumes that **Y** is normally distributed, errors are normally distributed and independent, and **X** has constant variance.

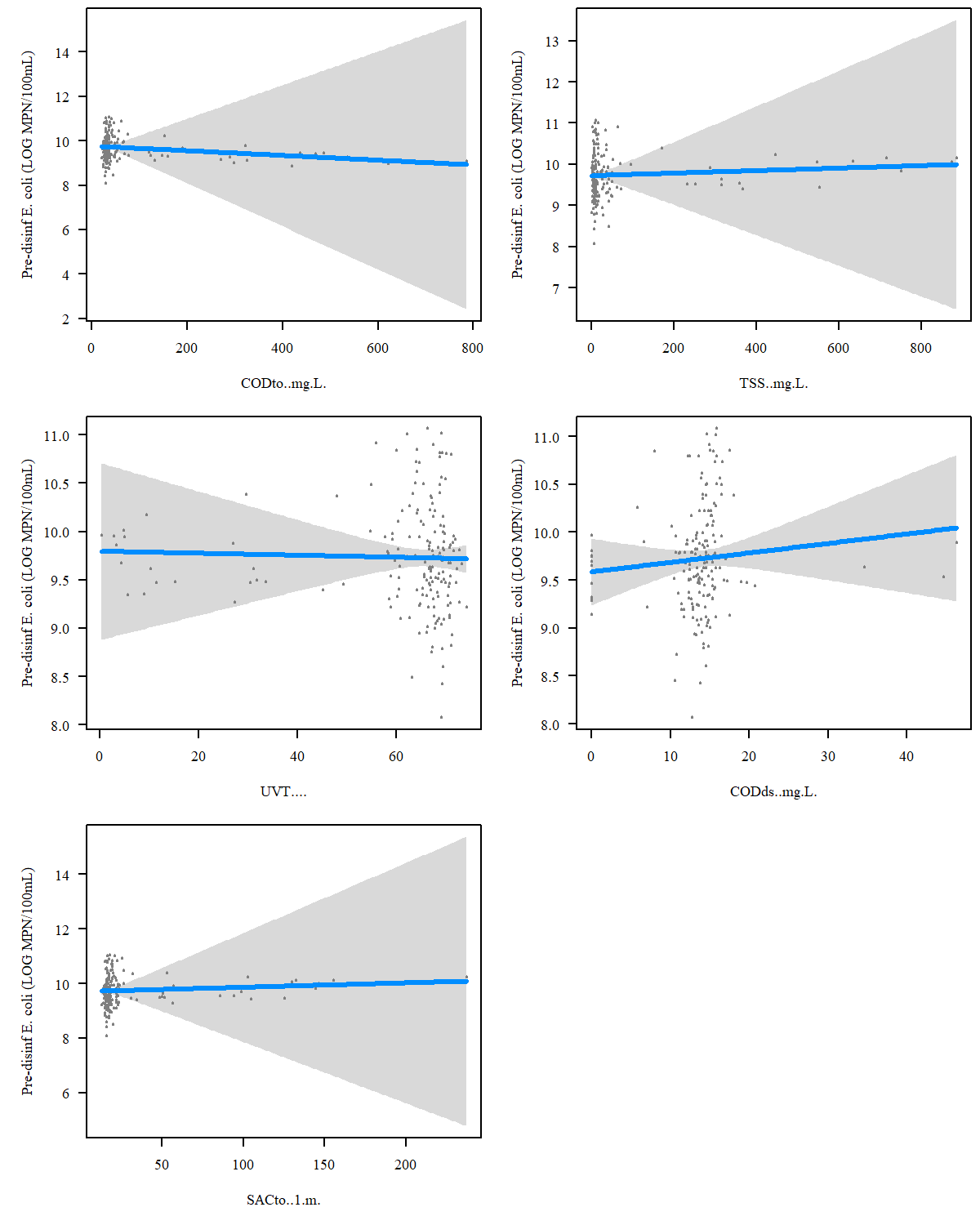
 

**Figure 2.** Multiple linear regression model fit for pre-disinfection *E. coli*. Black circles represent actual observations. R-squared value in lower right. (a) Variation of predicted pre-disinfection *E. coli* given a range of CODds measured by a sensor near the sampling point (b) Variation of predicted pre-disinfection *E. coli* given a range of RAS flows for Quad 4 which provides influent to the disinfection basin.

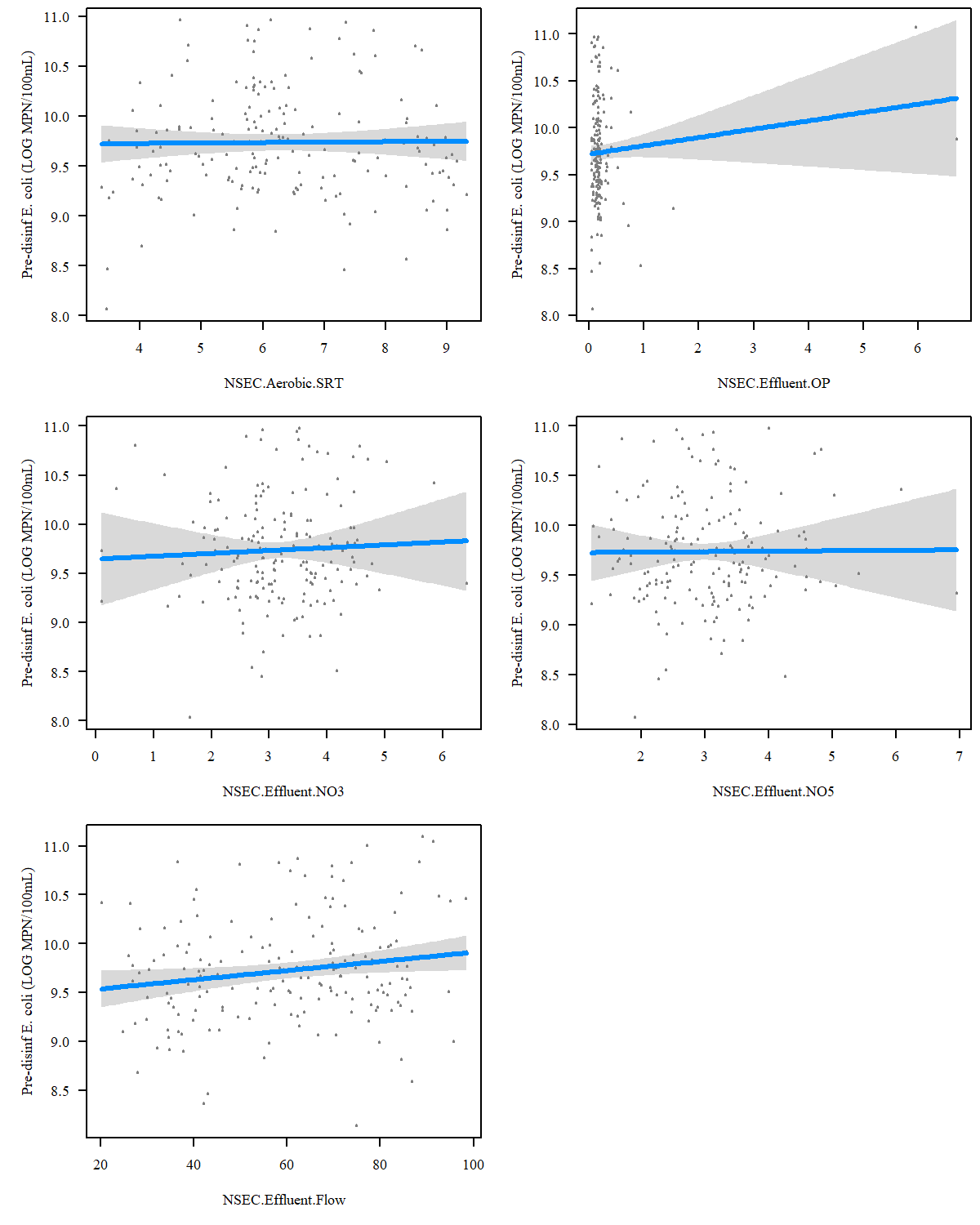
##### 

## Generalized Linear Model

The general linear model (LM) requires that the response variable follows the normal distribution whilst the generalized linear model (GLM) is an extension of the LM that allows the specification of models whose response variable follows different distributions.



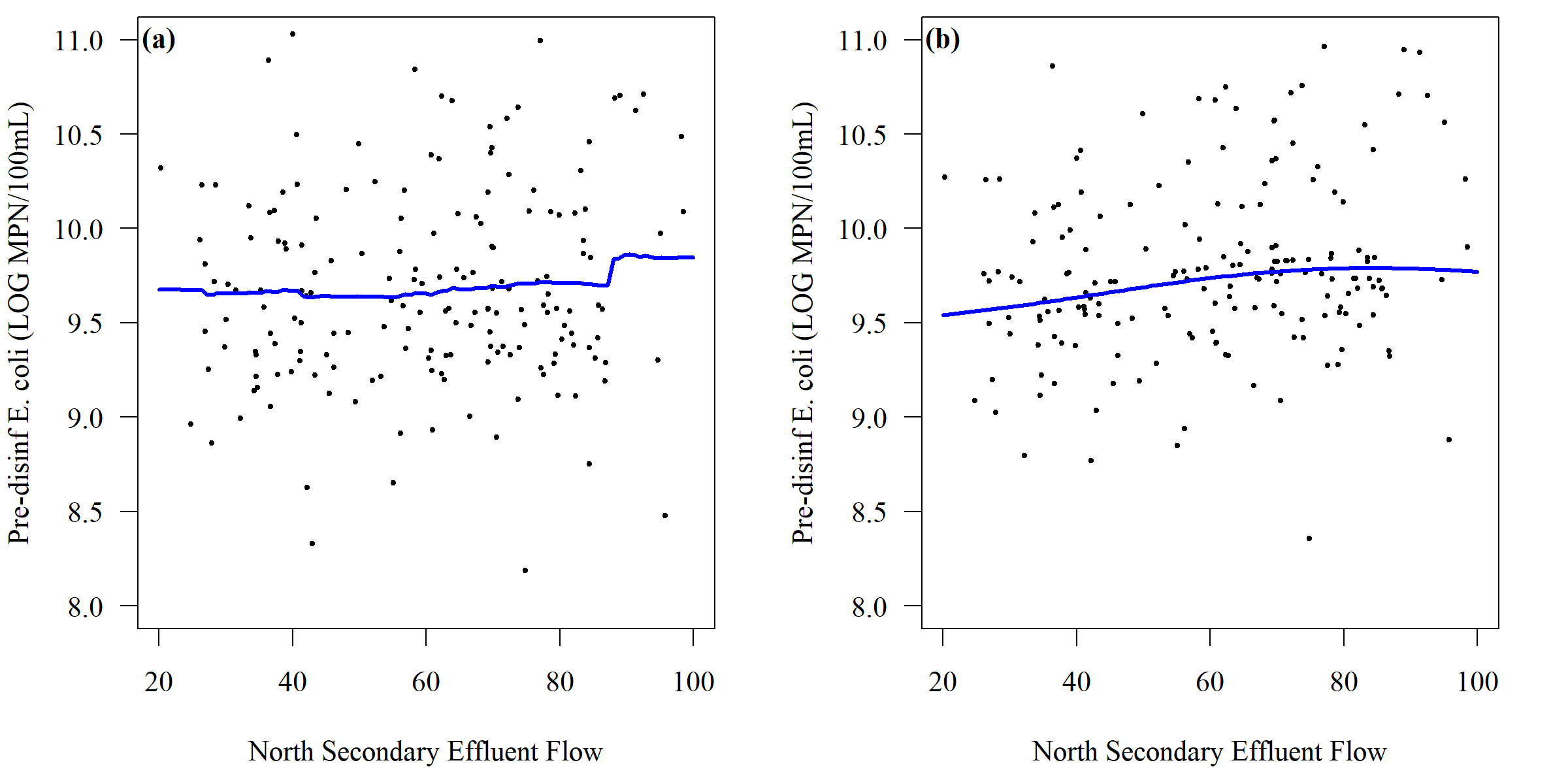
**Figure 3a.** GLM model constructed for visual spectrum data.

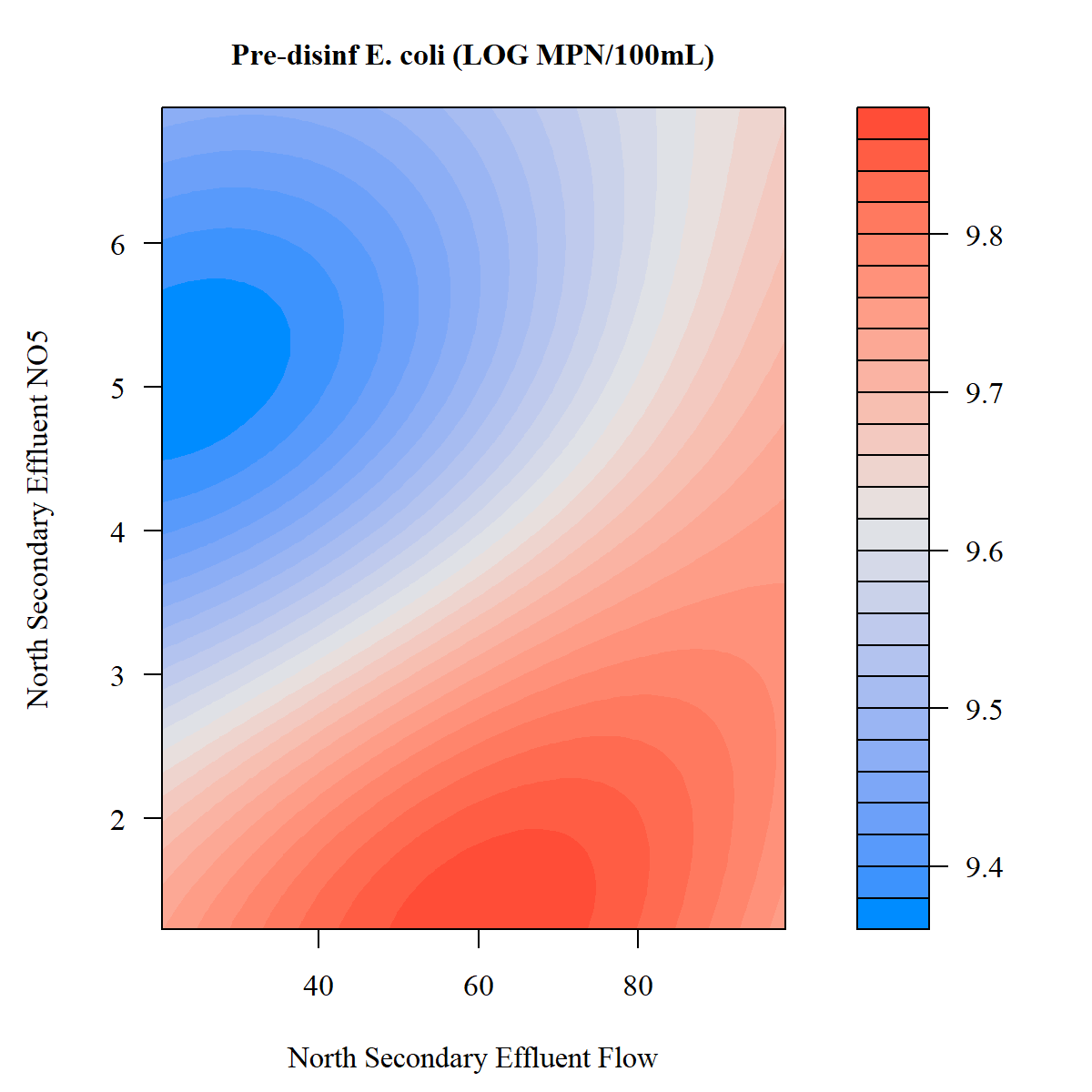


**Figure 3b.** GLM model constructed for north secondary process data.

##### 

## Non-Regression Model

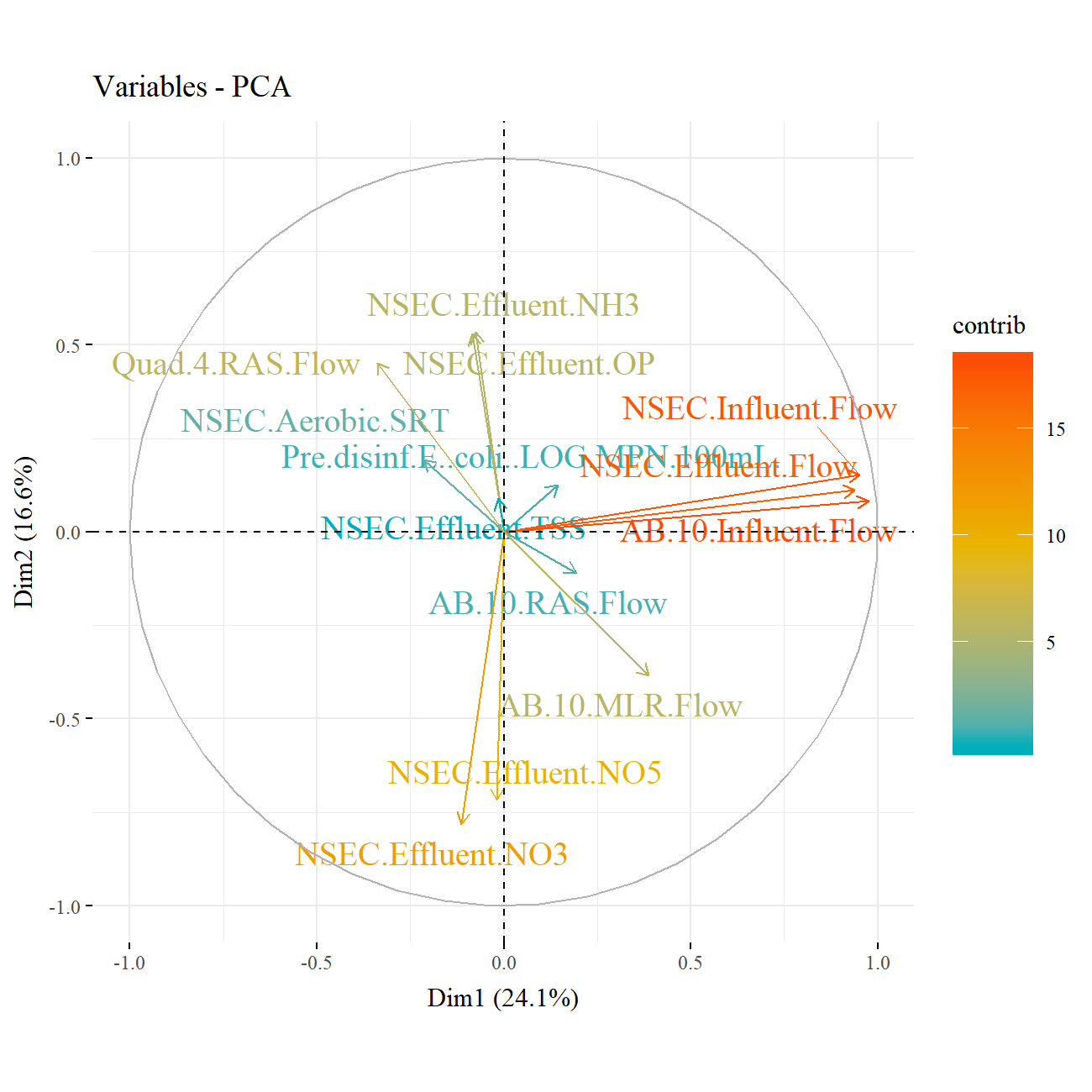
 **Figure 4.** Comparison of (a) random forest model and (b) support vector machine model for predicting *E. coli* into the PAA disinfection basin



**Figure 5.** Support vector machine model for predicting *E. coli* into the PAA disinfection basin

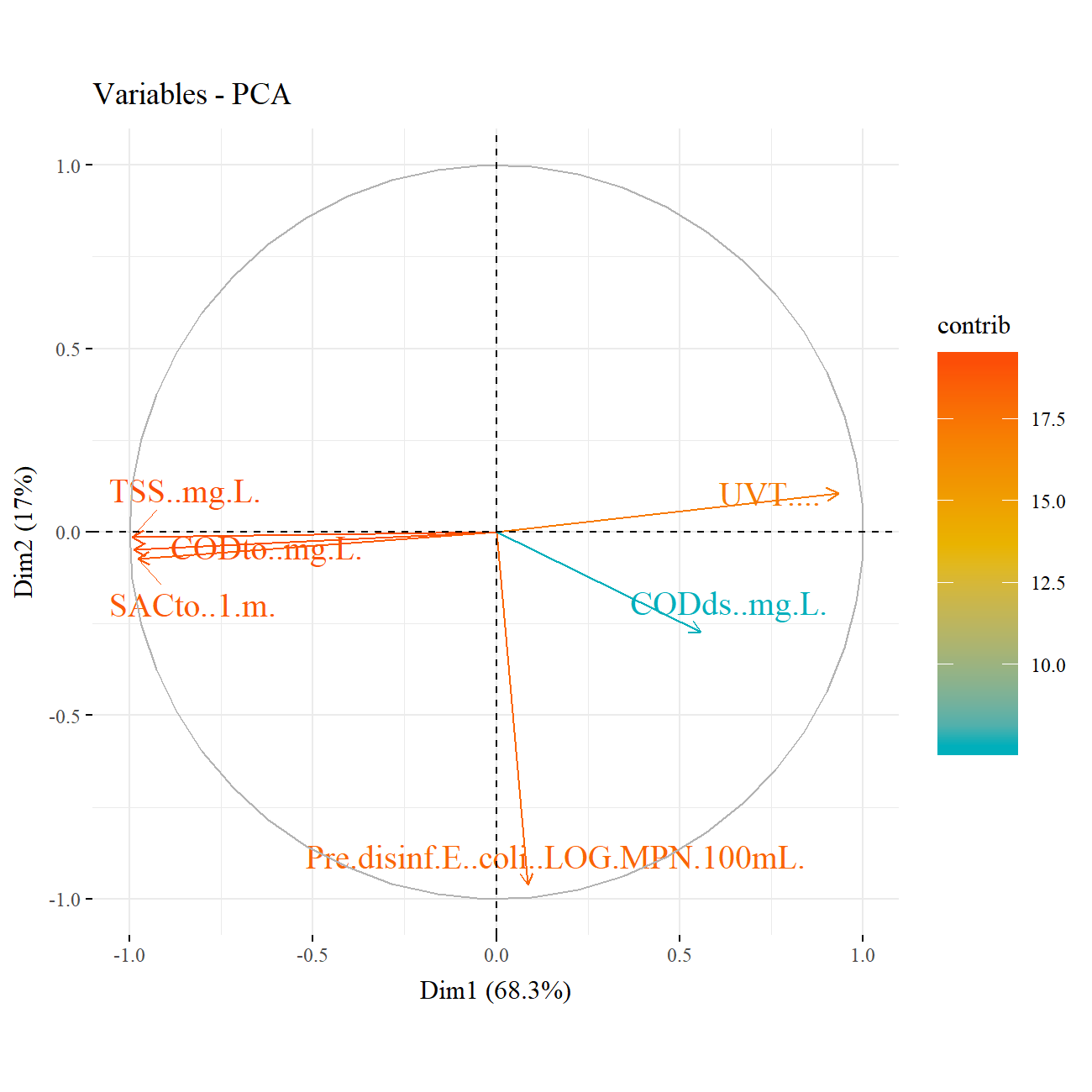
##### 

## Principal Component Analysis



**Figure 6.** PCA variable loading for effluent of north secondary

##### 

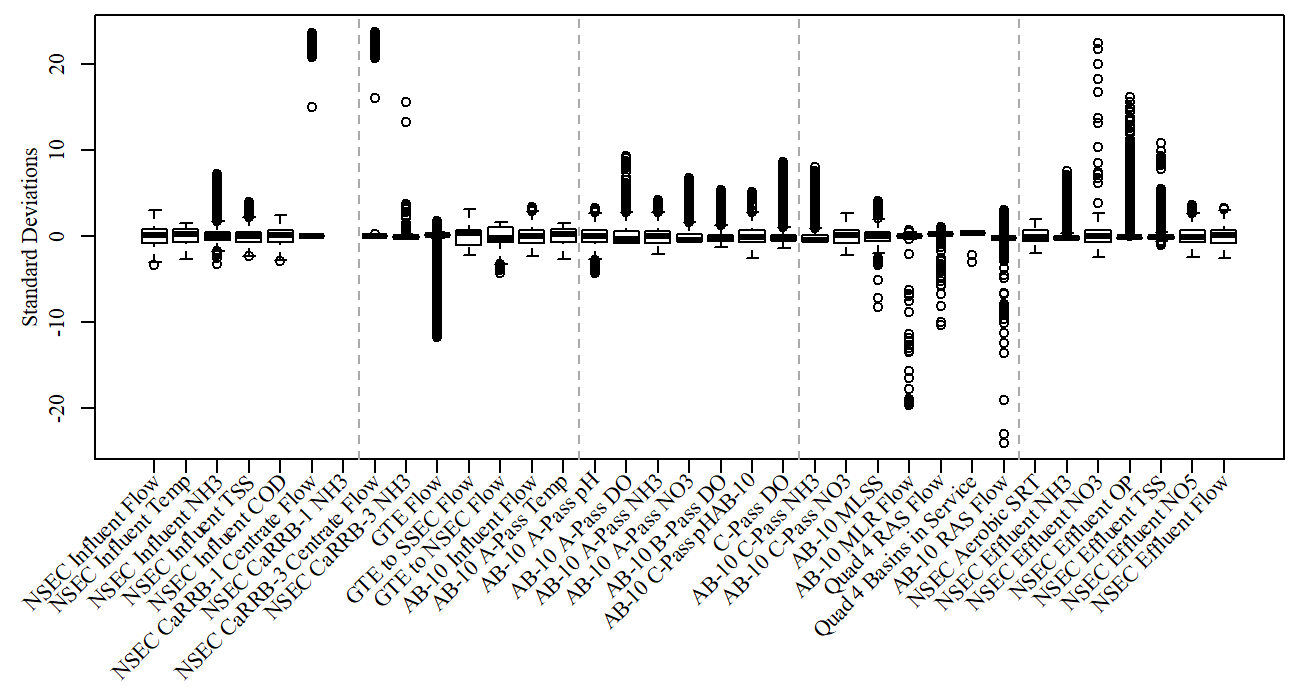


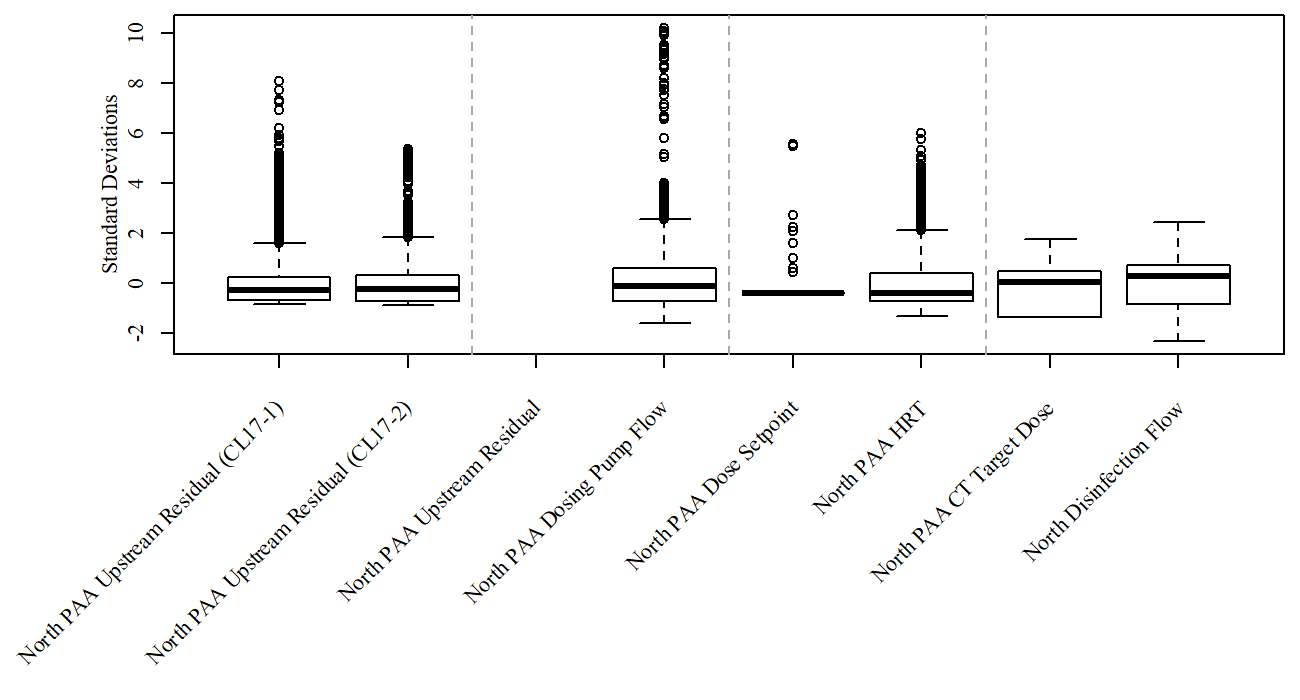
**Figure 7.** PCA variable loading for effluent of north secondary

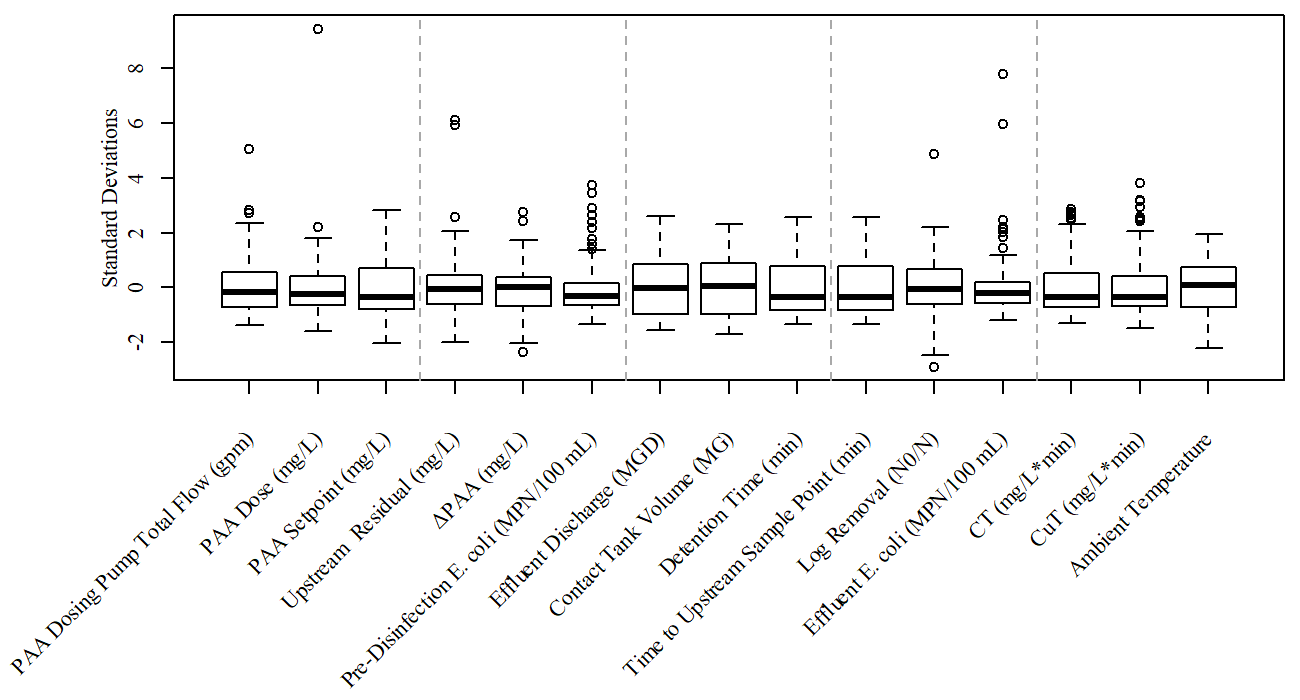
## Generalized Additive Models

# Supplementary Information

## Figures

 **Figure S1.** Centered and scaled boxplots of north secondary online data from MWRD

 **Figure S2.** Centered and scaled boxplots of north disinfection online data from MWRD

 **Figure S3.** Centered and scaled boxplots of north disinfection grab sample data from MWRD

##### 

## Tables

**Table S1.** PCA variable contributions to visual spectrum data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Dim.1 | Dim.2 | Dim.3 | Dim.4 | Dim.5 | Dim.6 |
| Pre.disinf.E..coli..LOG.MPN.100mL. | 0.18 | 90.88 | 8.95 | 0.00 | 0.00 | 0.00 |
| CODto..mg.L. | 23.76 | 0.24 | 0.92 | 13.70 | 0.51 | 60.87 |
| TSS..mg.L. | 23.98 | 0.02 | 0.05 | 4.14 | 60.95 | 10.85 |
| UVT…. | 21.18 | 1.10 | 6.49 | 70.63 | 0.12 | 0.49 |
| CODds..mg.L. | 7.59 | 7.24 | 80.72 | 2.19 | 2.25 | 0.01 |
| SACto..1.m. | 23.31 | 0.53 | 2.87 | 9.34 | 36.17 | 27.78 |

**Table S2.** PCA variable contributions to north secondary data

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Dim.1 | Dim.2 | Dim.3 | Dim.4 | Dim.5 | Dim.6 | Dim.7 | Dim.8 | Dim.9 | Dim.10 | Dim.11 | Dim.12 | Dim.13 |
| Pre.disinf.E..coli..LOG.MPN.100mL. | 0.7 | 0.7 | 3.8 | 10.4 | 3.4 | 54.2 | 20.2 | 3.4 | 1.0 | 2.0 | 0.2 | 0.0 | 0.0 |
| NSEC.Aerobic.SRT | 1.5 | 1.7 | 14.0 | 0.1 | 28.4 | 0.0 | 4.3 | 35.8 | 2.7 | 2.7 | 8.3 | 0.4 | 0.1 |
| NSEC.Effluent.NH3 | 0.2 | 13.2 | 10.6 | 18.4 | 2.0 | 6.0 | 0.1 | 0.9 | 41.9 | 6.5 | 0.1 | 0.1 | 0.0 |
| NSEC.Effluent.OP | 0.2 | 13.0 | 10.0 | 24.3 | 0.8 | 0.5 | 0.5 | 0.1 | 48.0 | 0.6 | 2.2 | 0.0 | 0.0 |
| NSEC.Effluent.NO3 | 0.4 | 28.3 | 4.1 | 13.2 | 0.7 | 0.6 | 2.2 | 1.0 | 0.0 | 8.2 | 41.4 | 0.0 | 0.1 |
| NSEC.Effluent.TSS | 0.0 | 0.4 | 0.0 | 0.9 | 56.3 | 1.9 | 11.8 | 27.4 | 0.1 | 0.9 | 0.2 | 0.0 | 0.0 |
| NSEC.Effluent.NO5 | 0.0 | 23.8 | 0.3 | 27.5 | 0.0 | 0.9 | 0.4 | 3.5 | 0.0 | 3.5 | 39.5 | 0.2 | 0.4 |
| NSEC.Influent.Flow | 28.9 | 1.1 | 2.3 | 0.2 | 0.0 | 1.0 | 0.3 | 0.0 | 0.0 | 0.3 | 0.4 | 11.2 | 54.3 |
| NSEC.Effluent.Flow | 28.1 | 0.6 | 3.6 | 0.5 | 0.1 | 0.6 | 0.5 | 0.3 | 0.4 | 0.3 | 0.2 | 64.0 | 0.9 |
| AB.10.Influent.Flow | 30.3 | 0.3 | 0.7 | 0.4 | 0.1 | 0.3 | 0.4 | 0.2 | 0.1 | 0.5 | 0.1 | 23.1 | 43.5 |
| AB.10.RAS.Flow | 1.2 | 0.6 | 16.8 | 0.4 | 0.4 | 27.9 | 42.6 | 1.4 | 0.3 | 7.6 | 0.7 | 0.0 | 0.1 |
| AB.10.MLR.Flow | 4.8 | 6.8 | 17.8 | 0.1 | 7.2 | 1.6 | 1.0 | 20.2 | 1.7 | 35.0 | 2.4 | 0.8 | 0.5 |
| Quad.4.RAS.Flow | 3.7 | 9.5 | 16.0 | 3.9 | 0.7 | 4.5 | 15.7 | 5.9 | 3.7 | 31.9 | 4.5 | 0.1 | 0.1 |