Multivariate fault detection in WWTP

Kathryn B. Newhart

February 21, 2019

# Introduction

Effective municipal water and wastewater treatment is imperative to protecting the environment and the health of consumers. Most treatment facilities are monitored and controlled by rudimentary supervisory control and data acquisition systems (SCADA) which use operator-determined upper and lower limits for monitoring individual process variables. A process disturbance, characterized by a variable operating outside of these bounds, could be caused by something as simple as inefficient equipment performance or as severe as a process failure. While single-variable setpoints have a low false alarm rate, this approach is limited to substantial disturbances of a single, monitored variable. Single-variable setpoints can be very slow to detect faults because they do not account for the relationships between other variables. To consistently meet effluent quality standards at minimum cost, more advanced process control methods need to be integrated into water and wastewater treatment facilities to avoid costly process disturbances. Statistically-derived limits are one advanced process control approach that can be used for early fault detection in industrial applications (Kourti, Lee, and Macgregor 1996) (Kourti et al., 1996; MacGregor and Kourti, 1995). Using previously collected data, normal operating conditions or in-control (IC) conditions can be defined. By comparing current observations to the previously collected data, abnormal or out-of-control (OC) conditions can be identified. However, there are multiple methods of calculating these statistical thresholds and not all methods can be directly applied to data collected from water and wastewater treatment facilities. Data produced by water and wastewater treatment plants frequently have missing values, contain outliers, and exhibit interdependent, nonlinear, and nonstationary behavior (Banadda et al., 2011; Olsson et al., 2005; Rosen and Lennox, 2001). Hence, describing the treatment process using strictly mathematical models (e.g., activated sludge models) is often insufficient for early fault detection (Dürrenmatt and Gujer, 2012). It is also inaccurate to apply most standard statistical methods to water and wastewater treatment data because the data are not normally distributed. Changing influent quality and quantity, temperature, internal shifts in microbial ecology, and process control instability are a few causes of the observed non-normal behavior. Without knowledge of how the data is distributed, it is difficult to make inferences about IC or OC conditions in water and wastewater treatment. By subsetting data to only include a short window of time, data can be assumed to have a constant mean (i.e., stationary). Updating the subsetted data with only the most recent observations is called an adaptive modification. To account for autocorrelation (i.e., relationship of a variable to a previous timestep), a dynamic modification duplicates a variable in a dataset and lags it by a set number of timesteps. Using these modifications, statistical process control methods can be applied with more confidence. The method used in this work is principal component analysis (PCA). PCA has been used in process control for fault detection In this work, principal component analysis (PCA) is modified to monitor a decentralized wastewater treatment facility treating 7,000 gallons per day (GPD) of municipal wastewater.

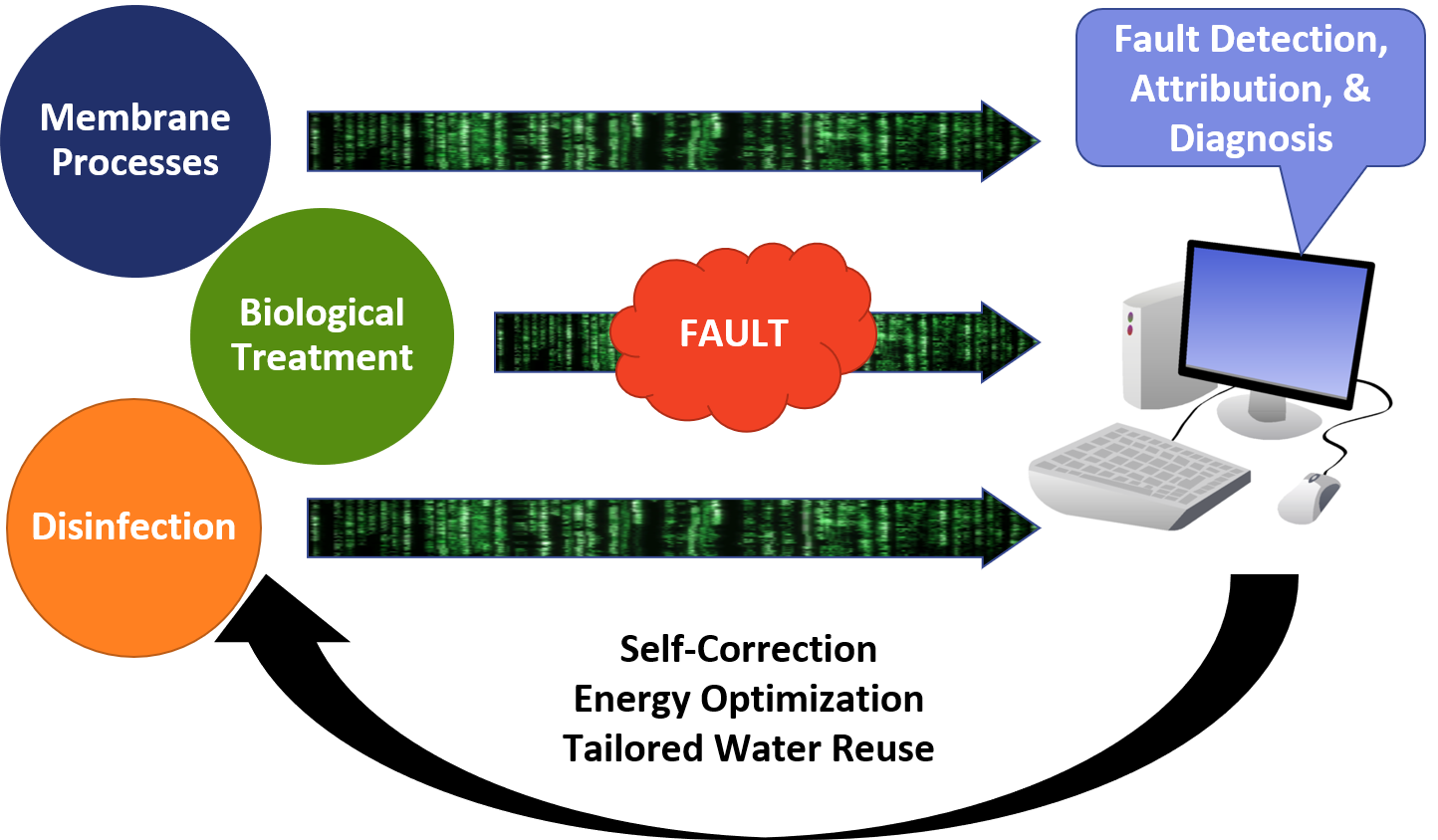
# Introduction to fault detection

* Single variable monitoring is **traditional approach** in WWTP
* Trend analysis is used, but methods are kept **private**
* Multivariate statistical process monitoring is **widespread in literature**, but *not in practice*

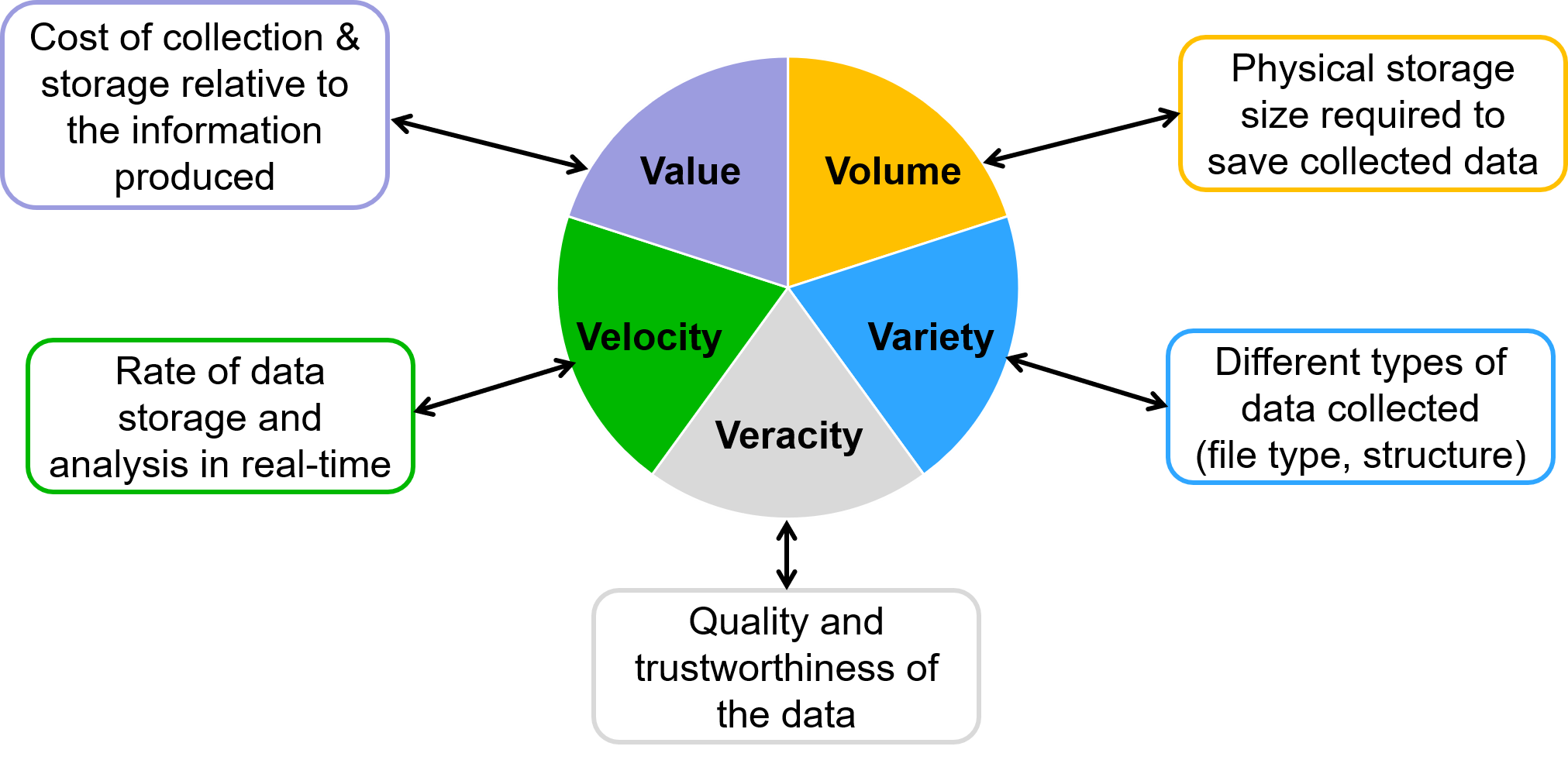
# Why use statistical process control?

* Statistical Process Control (SPE)
* There is **inherent variability** in WWTP that cannot be accounted for with an empirical model
* Operator-determined setpoints may not be valid for all operating conditions, making the system reliant on diligent supervison.
* **To improve efficiency, need to improve process control**

# Why use statistical process control?

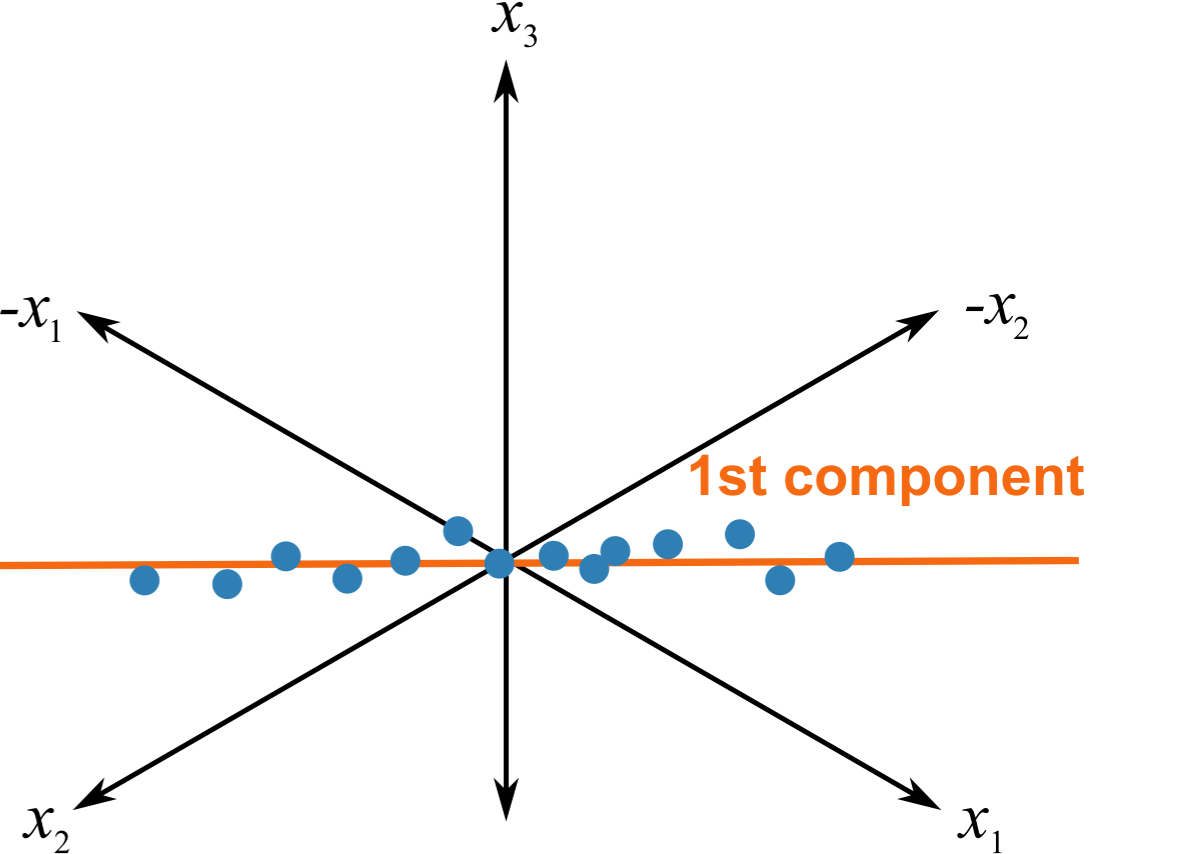


# Characteristics of big data



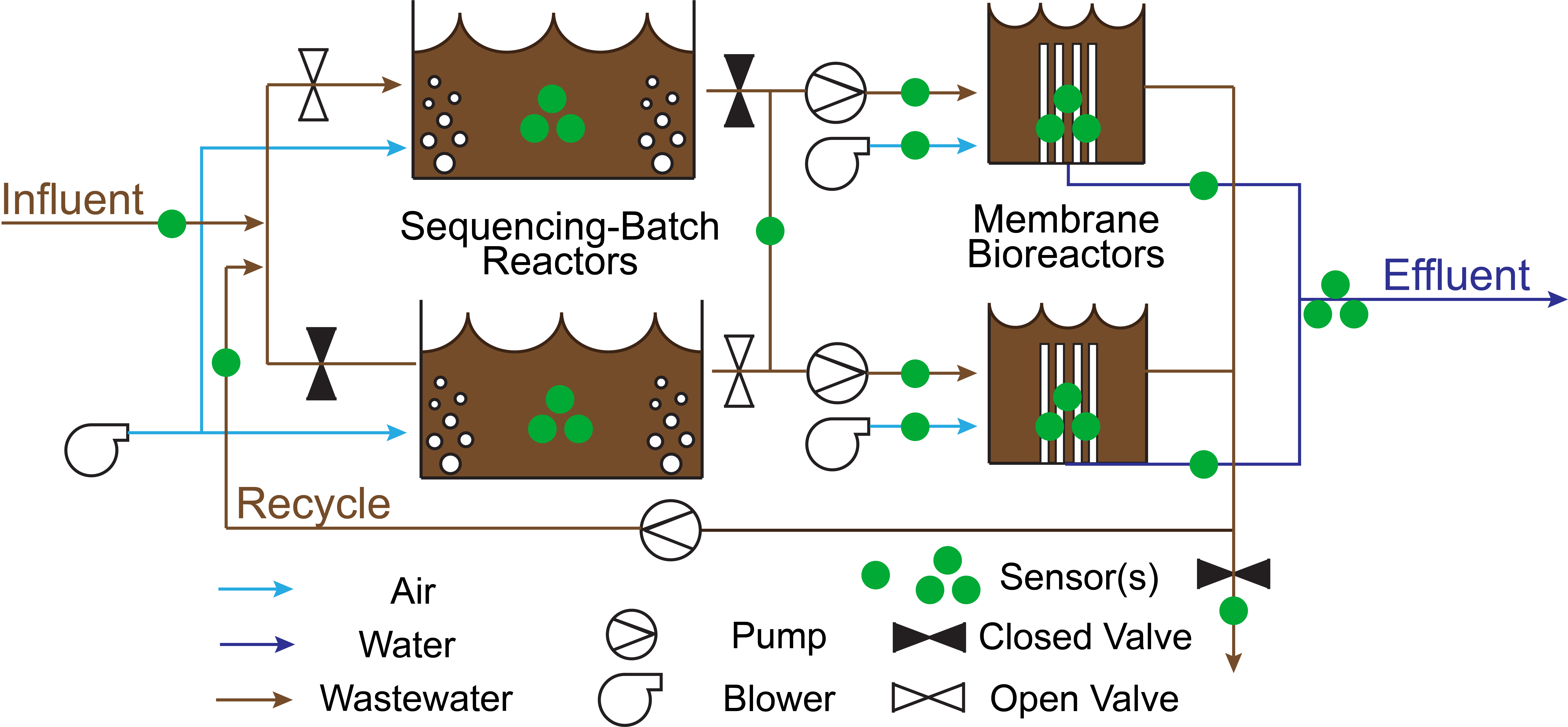
# Principal Component Analysis

* Data reduction technique
* Complex relationships between water quality variables
  + Interpretable linear combinations of data
* Which components represent the most variability?
  + 1st component = maximum variance



# Mines Park WWTP

# Mines Park WWTP



# Case studies

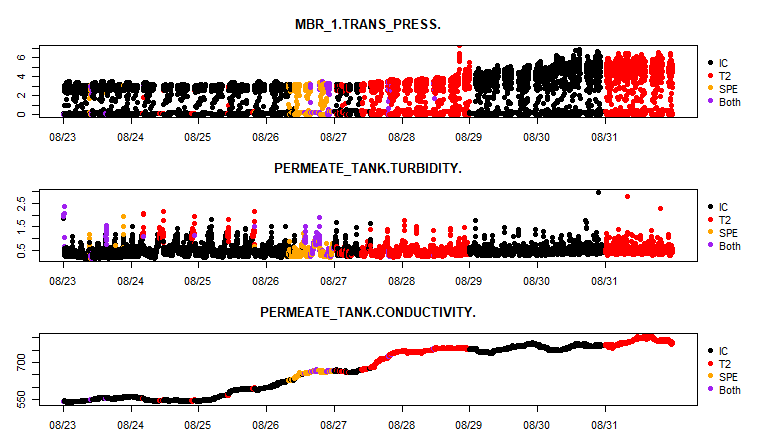
Steps for each case study:

1. Declare case study parameters
2. Compile & clean raw data
3. Train ADPCA model
4. Test ADPCA model
5. Calculate fault detection statistics
6. Summarize results

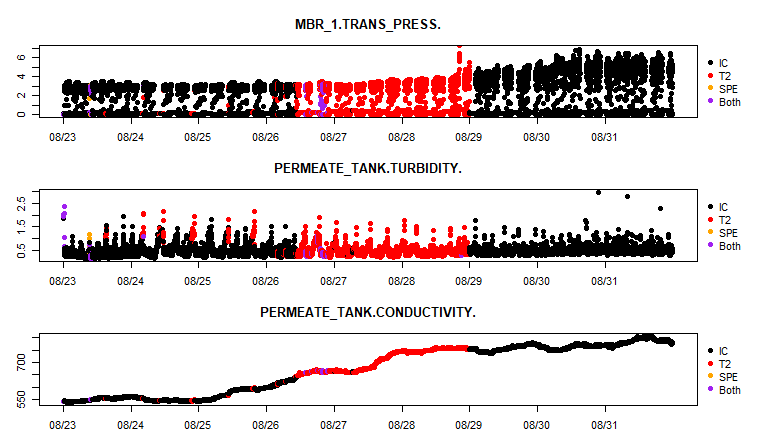
Required software:

* R (free)
* Student-developed packages (e.g., mvMonitoring, ADPCA)

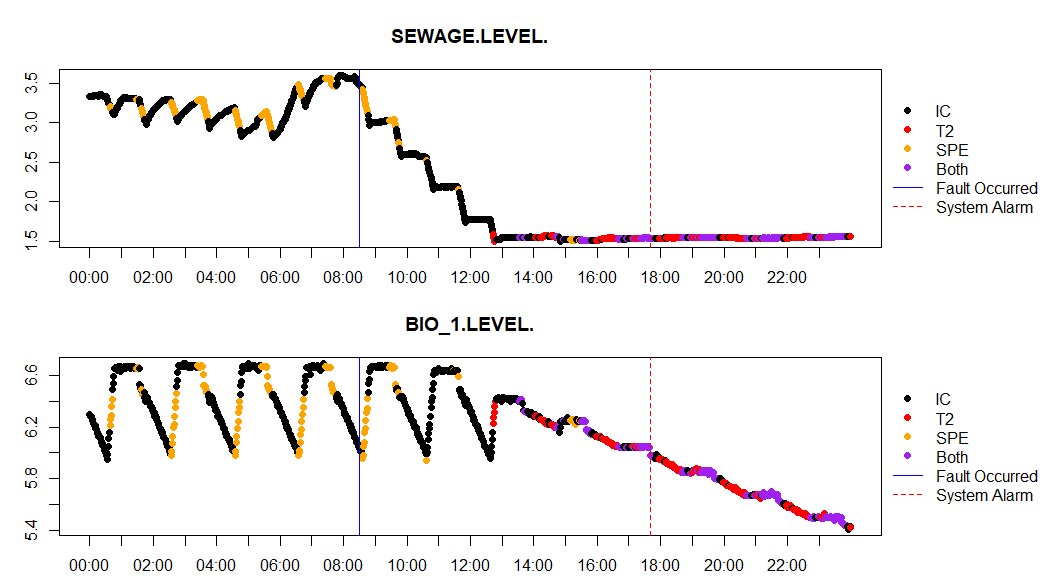
# Case study: membrane failure - 7 days



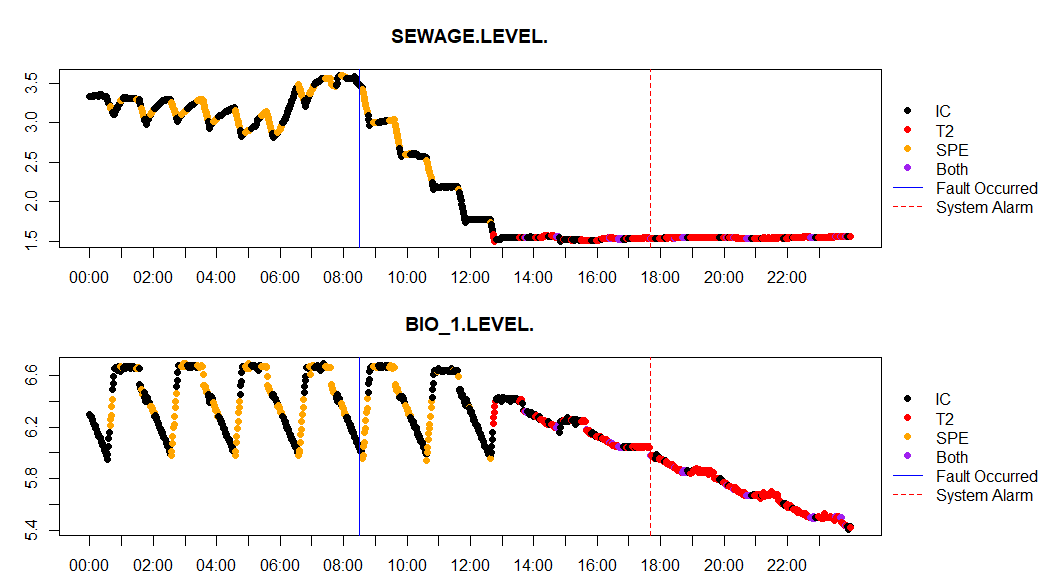
# Case study: membrane failure - 5 days



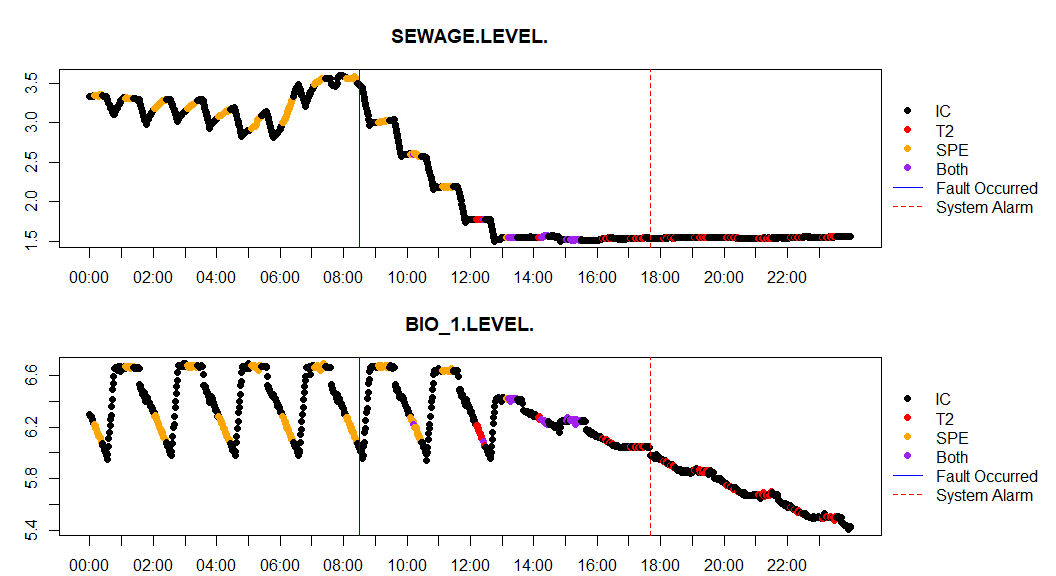
# Case study: pump shutdown - 10 days - ms



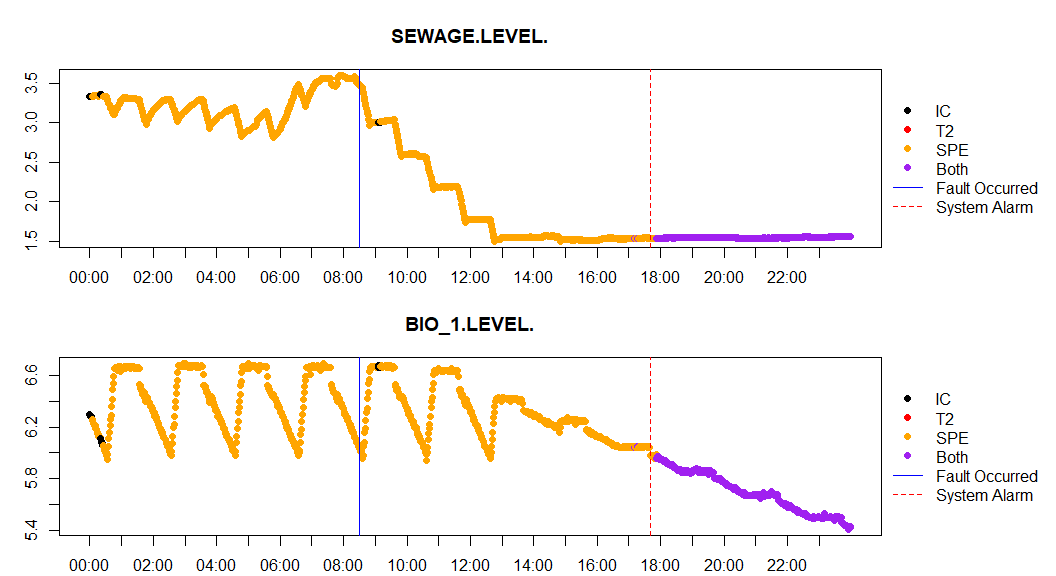
# Case study: pump shutdown - 7 days - ms



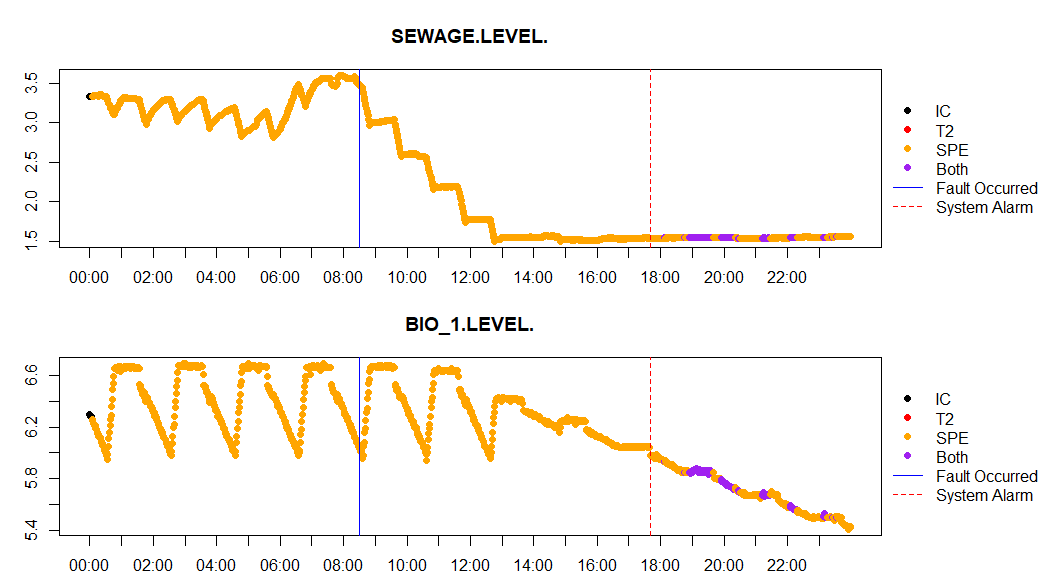
# Case study: pump shutdown - 5 days - ms



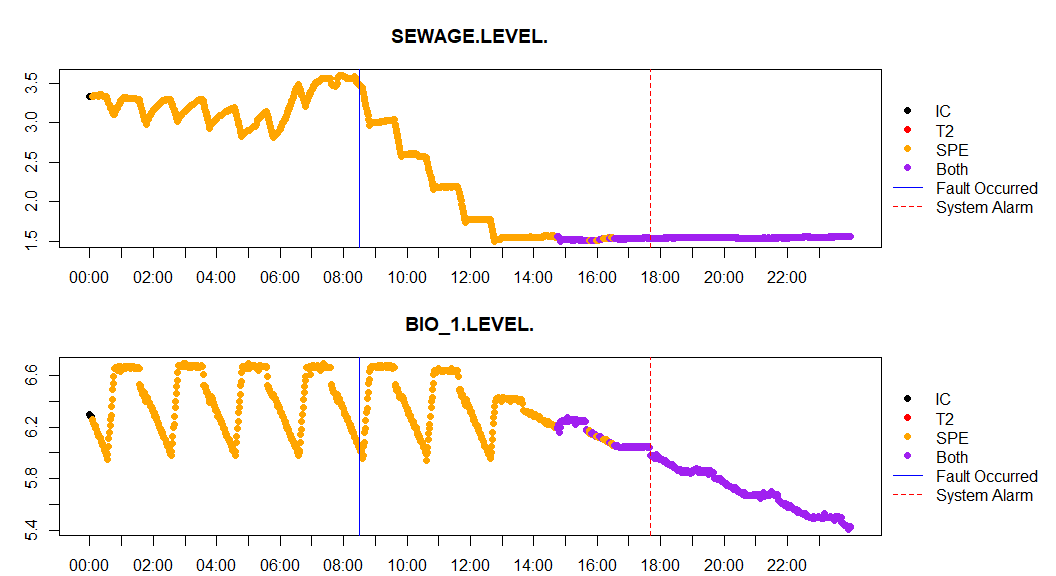
# Case study: pump shutdown - 10 days - ss



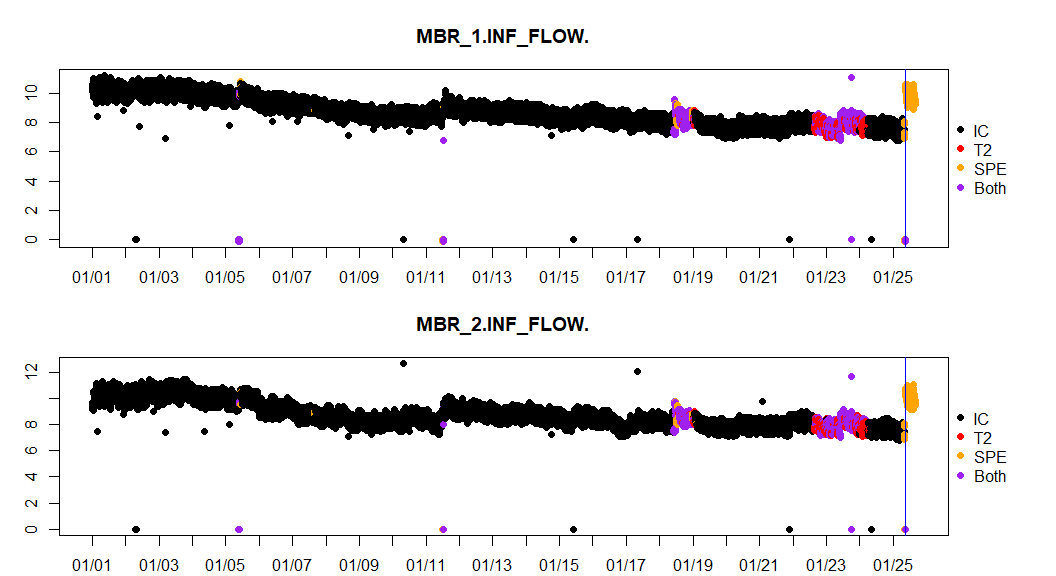
# Case study: pump shutdown - 7 days - ss



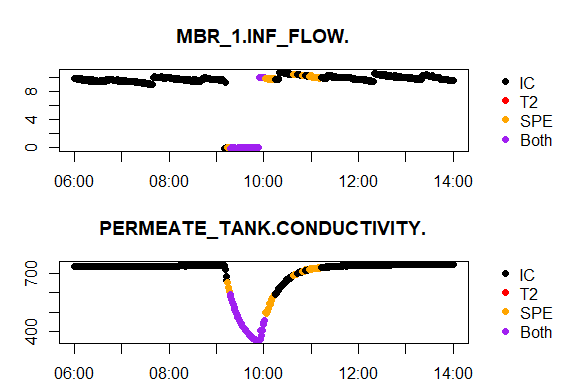
# Case study: pump shutdown - 5 days - ss



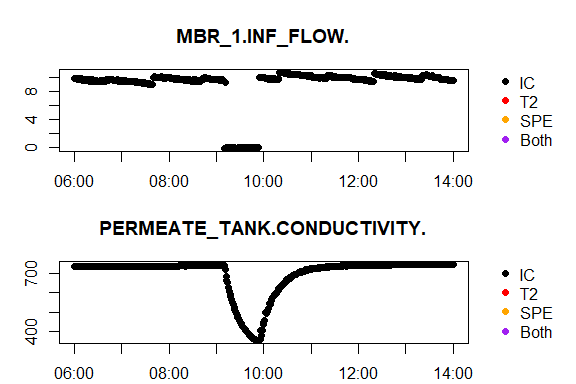
# Case study: pump clog - 5 days



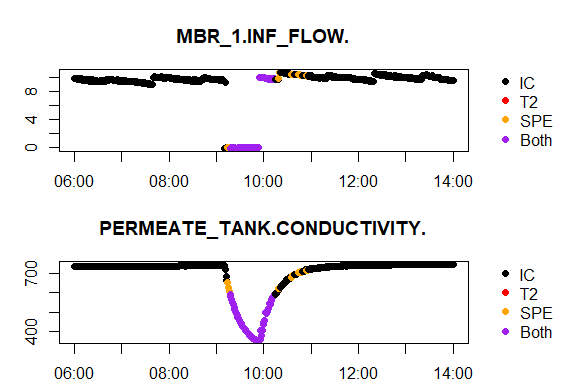
# Case study: membrane shutdown - 10 days - ss



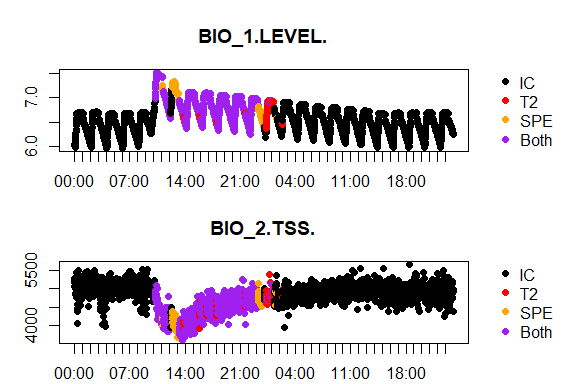
# Case study: membrane shutdown - 10 days - ms



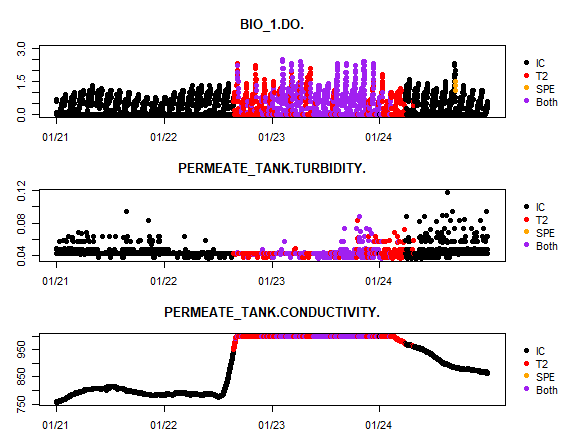
# Case study: membrane shutdown - 5 days - ss



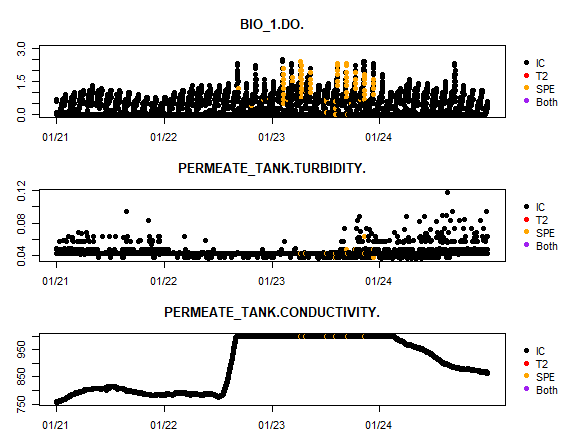
# Case study: influent overdose - 5 days



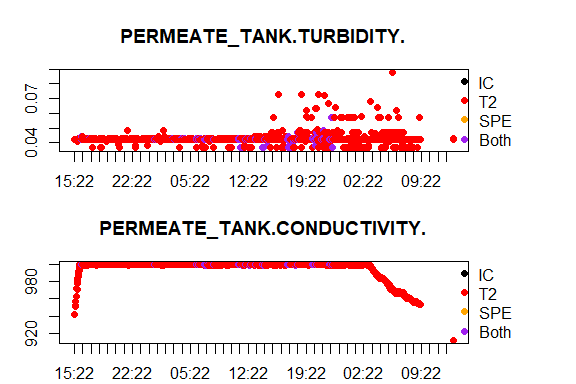
# Case study: inf/eff quality changes - 5 days - ss



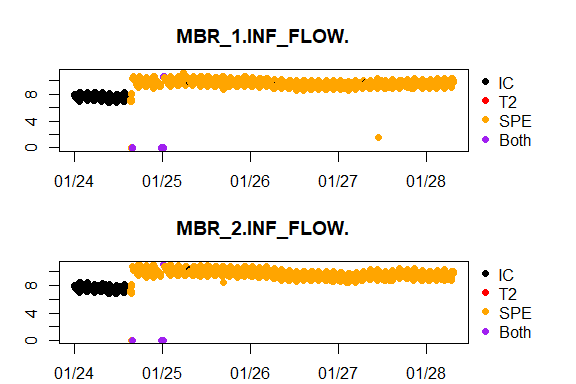
# Case study: inf/eff quality changes - 10 days - ss



# Case study: inf/eff quality changes - 10 days - ms



# Case study: Clear clog in pump - 5 days - ss



<img src="“images/5 day SS allGraphs 2018-01-24 2018-01-29.svg”>

# Alarm rates: January - 5 day

# Alarm rates: January - 10 day

Kourti, Theodora, Jennifer Lee, and John F. Macgregor. 1996. “Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control.” *Computers & Chemical Engineering* 20 (January): S745–S750. <https://doi.org/10.1016/0098-1354(96)00132-9>.