An Evaluation of Multivariate Statistical Process Monitoring in a Biological and Membrane Treatment System

Kathryn B. Newhart1, Molly Klanderman2, Tzahi Y. Cath1, Amanda S. Hering2

1Civil and Environmental Engineering, Colorado School of Mines, Golden, Colorado, USA | 2Baylor University, Waco, Texas, USA

Table of Contents

Abstract

Fault detection in wastewater treatment plants (WWTP) is currently reliant on operator-determined thresholds, which limits detection to faults that have already impacted the treatment process. The ability of statistical process control methods, like principal component analysis (PCA), to monitor the relationships between variables provides an additional layer of monitoring not achieved by the current fault detection paradigm. However, many traditional statistical analysis methods cannot be directly applied due to many unique features of WWTP data including non-stationarity, autocorrelation, and co-correlation. To account for these features, a modified version of PCA that includes adaptive, dynamic, and multistate adaptations is proposed to detect faults in a decentralized municipal WWTP with biological and membrane treatment units. False alarm rate and time-to-detection are used to compare different PCA configurations in order to identify best practices for implementation at WWTP.

Highlights

Keywords

* wastewater treatment
* statistical process control
* principal component analysis
* membrane treatment
* activated sludge

Abbreviations

AD-PCA. Adaptive-dynamic principal component analysis  
ARL. Average run length  
DO. Dissolved oxygen  
FDR. False-detection rate  
FICR. False-in-control rate  
IC. In-control  
MBR. Membrane bioreactor  
MSAD-PCA. Multistate adaptive dynamic principal component analysis  
OC. Out-of-control  
PC. Principal component  
PCA. Principal component analysis  
RAS. Return activated sludge  
SB-MBR. Sequencing-batch membrane bioreactor  
SBR. Sequencing-batch reactor  
SCADA. Supervisory control and data acquisition  
*SPE*. Squared prediction error  
SSAD-PCA. Single-state adaptive dynamic principal component analysis  
*T2*. Hotelling’s *T2*  
TMP. Transmembrane pressure  
TSS. Total suspended solids  
UF. Ultrafiltration  
WTP. Water treatment plants  
WWTP. Wastewater treatment plants

# Introduction

Effective municipal water and wastewater treatment is imperative to protecting the environment and the health of consumers. While treatment technologies themselves have improved over the past few decades, control is still archaic; heavily reliant on human operators to detect and respond to faults. Precision process control can help water treatment plants (WTP) and wastewater treatment plants (WWTP) achieve their effluent quality goals (e.g., regulated nutrient limits) in a cost-effective and reliable manner. A fault could be caused by something as simple as inefficient equipment performance or as severe as a process failure. However, most treatment facilities are monitored and controlled by standard supervisory control and data acquisition (SCADA) systems that use operator-determined upper and lower limits for monitoring individual process variables. Single-variable limits are established based on operator experience and are inherently limited in their capacity to detect process abnormalities or faults prior to system failure. This is due to the variety of operating and environmental conditions experienced at WTP and WWTP throughout the year. Additionally, the multivariate nature of treatment processes weaken the single-variable monitoring paradigm because of the correlation between certain process variables. In these cases, a wider range of static process limits are required to accommodate all “normal” operating conditions. While single-variable setpoints have a low false alarm rate (i.e., if a process variable is measured below an operator setpoint, a failure involving that particular variable has most likely occurred), this approach limits process monitoring and fault detection to substantial disturbances of a single, monitored variable. To consistently meet effluent quality standards at minimum cost, more advanced process control and monitoring methods need to be integrated into WTP and WWTP to avoid costly process disturbances.

Statistically-derived process limits are one advanced monitoring approach that can be used for early fault detection in industrial applications (Kourti, Lee, & Macgregor, [1996](#ref-kourti1996); MacGregor & Kourti, [1995](#ref-macgregor1995)). Using previously collected data, normal operating conditions (i.e., in-control (IC) conditions) can be defined. By comparing current observations to the previously collected data, abnormal (i.e., 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 WTP and WWTP. Data produced in WTP and WWTP frequently have missing values, contain outliers, and exhibit interdependent, nonlinear, and nonstationary behavior (Banadda, Nhapi, & Kimwaga, [2011](#ref-banadda2011); Olsson, Nielsen, Yuan, Lynggaard-Jensen, & Steyer, [2005](#ref-olsson2005); Rosen & Lennox, [2001](#ref-rosen2001)). Hence, describing the treatment process using strictly mathematical models (e.g., activated sludge models, first order disinfection kinetics) is often insufficient for early fault detection (Dürrenmatt & Gujer, [2012](#ref-durrenmatt2012)). It is also inappropriate to apply most standard statistical methods to WTP and WWTP 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 are distributed, it is difficult to make inferences about IC or OC conditions.

Many statistical process control methods have been applied to WTP and WWTP including control charts (Alves, Henning, Samohyl, Cristina Konrath, & Formigoni Carvalho Walter, [2013](#ref-alves2013); Corominas et al., [2011](#ref-corominas2011); Rosen, Röttorp, & Jeppsson, [2003](#ref-rosen2003); Wold, [1994](#ref-wold1994)) and partial least squares (Choi & Lee, [n.d.](#ref-choi2005); Qin, Gao, & Chen, [n.d.](#ref-qin2012)). The primary tool used in this work is principal component analysis (PCA); a widely used statistical method to pre-process data in order to monitor multiple variables simultaneously (Jackson, [1991](#ref-jackson1991)). PCA monitors the relationships between variables rather than the variables themselves; identifying independent, linear combinations of principal components (PCs) to capture as much variation in the data as possible while eliminating noise and redundancy (Wise & Gallagher, [1996](#ref-wise1996)). Some applications of PCA in WTP and WWTP include exploratory data analysis, fault detection (Patent No. 3723429343, [2006](#ref-king2006)), data reconstruction (C. Lee et al., [2006](#ref-lee2006a); Schraa, Tole, & Copp, [2006](#ref-schraa2006)), variable reduction for multiple regression models (Wallace, Champagne, & Hall, [2016](#ref-wallace2016); Wang, Ratnaweera, Holm, & Olsbu, [2017](#ref-wang2017)), and microbiological sample analysis (Jałowiecki et al., [2016](#ref-jalowiecki2016)).

Within a short period of time, data can be assumed to have a constant mean (i.e., stationary). Updating only the most recent data (as opposed to all historical observations) is called an *adaptive* modification. Adaptive PCA is the most popular extension of PCA for WTP and WWTP monitoring (Baggiani & Marsili-Libelli, [2009](#ref-baggiani2009); Kazor, Holloway, Cath, & Hering, [2016](#ref-kazor2016); Lee & Vanrolleghem, [2004](#ref-lee2004a); Rosen & Lennox, [2001](#ref-rosen2001)). Adaptive PCA updates the PCA model based on a “rolling window” of training observations. However, the effectiveness of adaptive PCA is sensitive to the size of the training window. If the training window is too large, faults could be ignored because there is too much variation or error in the training dataset (Baggiani & Marsili-Libelli, [2009](#ref-baggiani2009)). If the training window is too small, normal observations could be flagged as faults because an insufficient amount of variation is included in the training dataset (Rosen & Lennox, [2001](#ref-rosen2001)).

To account for autocorrelation (i.e., relationship of a variable to its previous value), a *dynamic* modification duplicates a variable in a dataset and lags it by a set number of timesteps (Ku, Storer, & Georgakis, [1995](#ref-ku1995)). Autocorrelation plots are valuable tools in this instance for identifying the strongest correlation between present and previous timesteps of an individual variable. Dynamic PCA is another common extension of PCA for monitoring WTP and WWTP (Lee et al., [2006](#ref-lee2006a); Dae Sung Lee et al., [2006](#ref-lee2006b); Mina & Verde, [2007](#ref-mina2007)). By including lagged observations (i.e., duplicating a variable and shifting observations back a certain number of timesteps that best represents the time-dependent nature of the process), the dynamic PCA extension accounts for autocorrelation among variables (Kruger, Zhou, & Irwin, [2004](#ref-kruger2004); Ku et al., [1995](#ref-ku1995)). For most WTP and WWTP applications, a lag of a single-timestep is sufficient (Kazor et al., [2016](#ref-kazor2016)).

Previous work combined PCA extensions to monitor a decentralized municipal WWTP using adaptive-dynamic PCA (AD-PCA). Kazor et al. ([2016](#ref-kazor2016)) evaluated a pH fault in the biological WWTP used in this work, caused by a seasonal change of influent water quality, and found that AD-PCA performed much better than conventional PCA at detecting the pH change. Additionally, AD-PCA was able to capture nonlinear process behavior as well as nonlinear dimension reduction approaches (e.g., kernel PCA, locally linear embedding), which are more computationally intense, and that the use of non-parametric thresholds greatly reduced false alarm rate. Odom et al. ([2018](#ref-odom2018)) further improved the AD-PCA paradigm by incorporating “state” information. A “state” is defined as a set of operating parameters that produce unique conditions. At a WTP or WWTP, this is representative of how a process is being operated (i.e., recycle streams, process setpoints). Statistically, this manifests as different covariance matrices. In the case of Odom et al. ([2018](#ref-odom2018)) PCA models were built for different states of blower operation. In an activated sludge WWTP, many monitored process variables (e.g., dissolved oxygen or DO) change depending on whether the conditions are aerobic or anaerobic (e.g., whether the air blower is on or off). By applying AD-PCA to different operating states (i.e., multistate AD-PCA or MSAD-PCA), as opposed to all data, the same pH fault evaluated by Kazor et al. ([2016](#ref-kazor2016)) was detected more quickly and consistently.

In this work, Kazor et al. ([2016](#ref-kazor2016))’s single-state AD-PCA (SSAD-PCA) and Odom et al ([2018](#ref-odom2018))’s MSAD-PCA is used to continuously monitor the decentralized municipal WWTP for a variety of fault scenarios, beyond the pH fault. Section 2 describes the specific WWTP, AD-PCA, and a generalized analysis approach for each case study. Section 3 compares the traditional, single-variable fault detection paradigm to SSAD-PCA and MSAD-PCA at the WWTP. Section 4 details how the results from each case study are interpreted, and Section 5 provides lessons learned from this work and suggestions for how to best implement AD-PCA at municipal WTP and WWTP.

# Materials and methods

## Wastewater treatment facility

The sequencing-batch membrane bioreactor (SB-MBR) at the Mines Park student housing complex (Colorado School of Mines, Golden, CO) is a coupled activated sludge and membrane system treating municipal wastewater produced by the residents of Mines Park (Figure 1). The Mines Park research site intercepts the municipal sewer line and diverts raw sewage to a 2,500-gal (9.5 m3) underground holding tank where the contents are withdrawn hourly in batch sizes of 324 gal by a submerged grinder pump. Influent is screened (2 mm) before being fed into one of two 4,500-gal (17 m3) sequencing-batch reactors (SBRs). The SBRs operate in alternating 2-hour cycles where screened sewage is added, the mixture of sewage and activated sludge is exposed to a combination of aeration conditions in order to remove chemical and biological constituents of interest, and the treated water is separated from the biologically-active solids. In traditional SBRs, the solid-liquid separation phase is achieved using gravity separation. The settled solids that are returned to the SBR are referred to as return activated sludge (RAS) and is measured as a concentration of total suspended solids (TSS). At the Mines Park facility, solid-liquid separation is achieved using ultrafiltration (UF) in the membrane bioreactors (MBRs). The PURON® hollow-fiber UF membranes (Koch Membrane Systems, Inc., Wilmington, Massachusetts) used have a nominal pore size of (i.e., remove contaminants greater than) 0.03 µm. The UF permeate is used for a variety of research purposes including potable and non-potable water reuse (Prieto et al., [2013](#ref-prieto2013)).

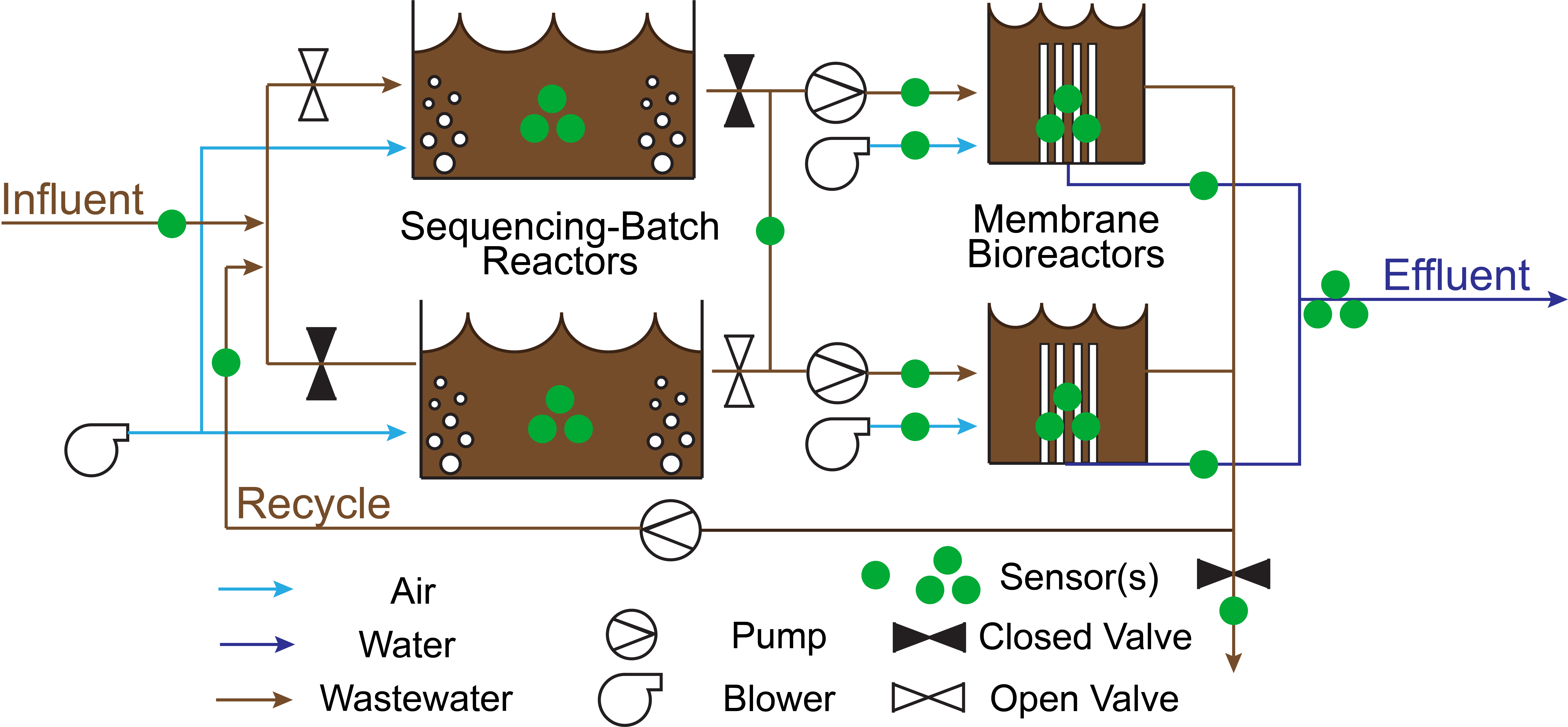


Figure 1. Process flow diagram of the SB-MBR integrated system, visualizing gas, liquid, and solid streams as well as the locations of where real-time process measurements are taking place. A single dot indicates one variable where multiple dots indicate multiple variables being measured.

The SB-MBR is controlled and monitored using a SCADA system that collects data from a variety of sensors and controls at high frequency. The data is then saved at a user-defined 1-minute interval (Table 1). Water quality measurements are made once a week using 24-hour time composite samples of the influent and effluent. Due to the nature of decentralized systems serving seasonal communities, the influent water quality can be highly variable and affect removal of certain contaminants. For example, at Mines Park, most tenants are undergraduate students whom vacate during holidays and summer months. Hence, decentralized facilities need to closely monitor and respond to process changes caused by a changing influent or other process modifications to prevent failure (i.e., no or impaired contaminant removal). In this work, faults that occurred in the SB-MBR are used as case studies for a fault detection system designed to provide WWTP operators with advance notice of a process change or failure.

Table 1. Monitored SB-MBR process variables. Variables are indicated as either state variables (indicate how a process is being operated, e.g., blower is on or off) or process variables (sensor measurements used to monitor or control a process). Variables are only included in the AD-PCA model if they directly measure a unit process, indicated here by the SBR-only, MBR-only, or both SBR and MBR AD-PCA models.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Monitored Variable | State (S) or Process (P) | SBR or MBR Subsystem |
| 1 | Bioreactor 1 Phase | S | SBR |
| 2 | Bioreactor 2 Phase | S | SBR |
| 3 | Bioreactor Blower 1 Running | S | SBR |
| 4 | Bioreactor Blower 2 Running | S | SBR |
| 5 | MBR 1 Mode | S | MBR |
| 6 | MBR 2 Mode | S | MBR |
| 7 | MBR 1 & 2 Flux Mode | S | MBR |
| 8 | MBR1 State | S | MBR |
| 9 | MBR 2 State | S | MBR |
| 10 | MBR 1 Air Scour Valve Position | S | MBR |
| 11 | MBR 2 Air Scour Valve Position | S | MBR |
| 12 | Bioreactor 1 Dissolved Oxygen | P | SBR |
| 13 | Bioreactor 2 Dissolved Oxygen | P | SBR |
| 14 | Bioreactor Blower 1 Flow | P | SBR |
| 15 | Bioreactor Blower 2 Flow | P | SBR |
| 16 | Sewage Flow | P | SBR |
| 17 | Sewage Level | P | SBR |
| 18 | Ambient Temperature | P | SBR/MBR |
| 19 | Bioreactor 1 Level | P | SBR |
| 20 | Bioreactor 1 Temperature | P | SBR |
| 21 | Bioreactor 2 Level | P | SBR |
| 22 | Bioreactor 2 Temperature | P | SBR |
| 23 | RAS pH | P | SBR/MBR |
| 24 | RAS Temperature | P | SBR/MBR |
| 25 | RAS Total Suspended Solids | P | SBR/MBR |
| 26 | MBR 1 Permeate Flow | P | MBR |
| 27 | MBR 1 Permeate Pressure | P | MBR |
| 28 | MBR 2 Permeate Flow | P | MBR |
| 29 | MBR 2 Permeate Pressure | P | MBR |
| 30 | MBR 1 Transmembrane Pressure | P | MBR |
| 31 | MBR 2 Transmembrane Pressure | P | MBR |
| 32 | MBR 1 Air Scour Flow | P | MBR |
| 33 | MBR 2 Air Scour Flow | P | MBR |
| 34 | MBR 1 Air Scour Pressure | P | MBR |
| 35 | MBR 2 Air Scour Pressure | P | MBR |
| 36 | MBR 1 Level | P | MBR |
| 37 | MBR 2 Level | P | MBR |
| 38 | MBR 1 Influent Flow | P | MBR |
| 39 | MBR 2 Influent Flow | P | MBR |
| 40 | MBR 1 Total Suspended Solids | P | MBR |
| 41 | MBR 2 Total Suspended Solids | P | MBR |
| 42 | Permeate Tank Conductivity | P | MBR |
| 43 | Permeate Tank Level | P | MBR |
| 44 | Permeate Tank Turbidity | P | MBR |

## Fault events and testing

Two categories of testing were performed, differentiated by the absence or presence of OC observations in the training dataset:

* *Ideal*: The initial widow of time used for training includes only IC observations, as initially identified by WWTP operators. The PCA model is tested for various durations (depending on the continuous length of time until the number of sequential OC observations was equal to 24 hours training or 1440 observations). The goal of the continuous testing is to calculate bulk fault detection metrics to compare overall performance of each method.
* *Non-ideal*: A major fault is identified by operators and PCA models are compared by their ability to adequately alert operators of OC conditions. Due to a lack of totally IC conditions prior to the events, the training data includes some OC observations. The model is allowed to completely update (e.g., if initially trained on 3 days of observations, model is tested and re-trained on a total of 3 more days of observations) prior to the fault date. The goal of the non-ideal testing is to determine which method is able to overcome inherent inaccuracies and OC conditions at WWTP.

The faults tested in both the ideal and non-ideal approaches are outlined below.

### Pump failure

Raw wastewater enters the Mines Park WWTP via an interceptor to the municipal sewer. Sewage flows by gravity to a holding tank where a submerged pump provides hourly batches of wastewater to the SBRs. In this event, the interceptor pipe was clogged and flow to the holding tank stopped. The level in the holding tank declined each time a new batch was pumped to the SB-MBR until the previously-submerged pump was exposed, causing the pump to overheat and shutdown each time it attempted to pump a new batch of wastewater to the SB-MBR. The SB-MBR continued to treat water present in the system despite not receiving any influent, albeit at a slower rate due to the lower batch size (as the level in the SBRs is used to calculate the MBR production rate). The first system alarm was triggered when the levels in the SBRs dropped below an operator threshold, which is the lowest level that the SBR should be operated. Due to a recent change in operating conditions, critical alarm levels were not adjusted to the new 6 ft SBR limit, and operators were not notified by the SCADA system until the SBR levels dropped below the previously used alarm level of 5 ft.

### Change to influent quality

The SB-MBR is designed to remove grit, suspended solids, organic compounds, and nutrients through a combination of physical, chemical, and biological reactions. However, the SB-MBR is not designed to removed dissolved solids (i.e., salts) and an indication of a changing influent water quality is a temporary increase in dissolved solids concentration of the permeate which is measured using electrical conductivity. In most cases, this change affects additional process variables that are otherwise unexplained such as DO in the SBRs and permeate turbidity in the MBR.

### Influent overdose

During routine maintenance, the influent pump to the SB-MBR was left in the manual position by an inexperienced operator. The addition of excess raw wastewater caused a shift fault in SBR level and total suspended solids (TSS) concentration, but was not detected by operators until a high-level threshold was exceeded in the SBRs. In conventional WWTP that rely on gravity for clarification, a substantial shift in TSS could impact settling and the quality of the product water. Since the Mines Park WWTP uses membranes for solid-liquid separation, solids are retained within the system and can quickly recover from the TSS shift. However, transmembrane pressure (TMP) in the MBR is dependent on TSS and flowrate. TMP may temporarily decrease when TSS declines, but TMP will increase with the need to increase flowrate to treat the additional volume in the bioreactor.

### Sensor faults

In the event of a power surge, re-calibration, cleaning, or position adjustment, online sensors can exhibit a brief spike in measurements. During preliminary data analysis for engineering design, it is critical to detect and remove these abnormal values before using the data to estimate average and peak flows and mass loading (e.g., nutrients, TSS). Manually removing these values is time consuming and often skews the data by removing all values above and below visually-determined (from time-series plots) thresholds.

## Statistical process control

### Principal component analysis

PCA is a linear transformation process that identifies a data subspace representing the optimal correlation between process variables. Every process variable is first centered to zero mean and scaled to unit variance in order to standardize the comparison among variables of different units and operating ranges. The covariance matrix is calculated from the scaled data and singular value decomposition is performed to calculate the eigenvalues and eigenvectors. The eigenvectors are the PCs of the model and the eigenvalues represent the amount of variation captured by the associated PC.

The subspace defined by the PC loadings (i.e., PCA model) indicates the direction in which the data varies the most. The first PC captures the most variability, the second PC captures the second most variability independent of the first, etc. By capturing the linear relationships among variables in each component, the original dataset can be reduced to contain fewer dimensions. The number of PCs is selected based on the desired percent of variability to be captured by the PCA model (90% here). PCA retains the information provided by each observation in a lower-dimensional space, which can reduce computational time for data analysis and reduce noise for fault detection. A more detailed description of how to apply and interpret PCA can be found in Jackson ([1991](#ref-jackson1991)) and Pankavich and Swanson ([2015](#ref-pankavich2015)).

### Monitoring statistics and thresholds

When defining a PCA model, a dataset that is known to be fault-free is usually used as the model subspace should represent IC conditions. By plotting new observations in the PCA subspace, a process is considered IC if the observations are within a certain distance from the model subspace. Two measures of distance from the PCA model are the squared prediction error (*SPE*) and Hotelling’s *T2*. *T2* is a measure of where an observation is within the PCA model; whereas *SPE* is a measure of variation that is not captured by the model (Figure 2). If an observation’s *SPE* or *T2* is above the calculated threshold, it is flagged as abnormal. Due to noise and missing data, even in an IC process, multiple sequential observations must be flagged to consider that the process is truly OC. In this work, 5 sequential observations must be above the *T2* or *SPE* threshold to trigger an alarm.

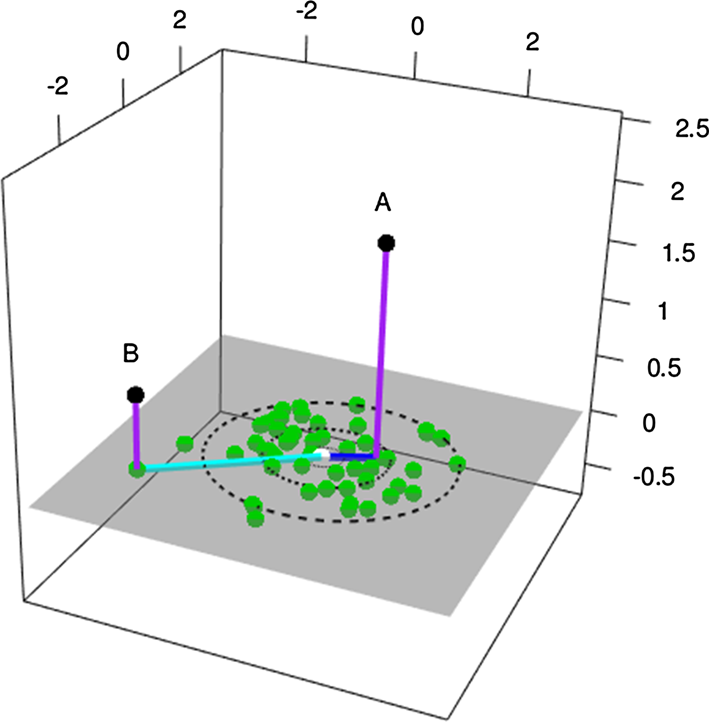


Figure 2. Three-dimensional observations (black) are plotted in their two-dimensional PCA model (green). Observation **A** plotted in the model subspace has a large *SPE* and a small *T2* compared to observation **B**, which has a small *SPE* but large *T2*. Adopted from Kazor et al. ([2016](#ref-kazor2016))

The conventional method of calculating thresholds for *SPE* and *T2* makes assumptions about the distribution of the IC data. However, data produced in WTP and WWTP rarely fit any one distribution (Berthouex, [1989](#ref-berthouex1989)). Following Odom et al. ([2018](#ref-odom2018)), we calculate nonparametric (i.e., distribution-free) thresholds (Chen et al., 2016; Qiu and Li, 2011) via kernel density estimation with a Sheather-Jones bandwidth and a Gaussian kernel (Sheather, 2009). A threshold is made more or less strict by decreasing or increasing the α value, respectively. α is defined as the probability that an IC observation will be flagged as an OC observation. If α is too high, then faults could be missed. If α is too low, then the false alarm rate could be high. An α = 0.01 is used for this work.

### Adaptive-dynamic extensions

Kazor et al. ([2016](#ref-kazor2016)) found that AD-PCA is capable of overcoming some of the limitations found in WWTP (e.g., nonlinearity, nonstationary, autocorrelation). The adaptive modification uses a “rolling window” of time over which process variables can be assumed to be linearly-related and stationary. Only IC observations are used to update the model as the “window” of time moves forward, and the oldest observations are forgotten (i.e., removed from the training dataset). The dynamic modification accounts for autocorrelation by including a lagged copy of each variable (i.e., 1-lag includes the previous observation and the current observation), doubling the size of the dataset but not the number of PCs (i.e., dimension of the PCA model).

### Multistate extension

In Kazor et al. ([2016](#ref-kazor2016)), all past observations were included in the PCA model in a SSAD-PCA configuration. MSAD-PCA repeats the AD-PCA process for each set of operating conditions that may have different relationships among variables. For example, the relationship between TMP and flow through the membranes is entirely dependent on whether the membrane is permeating, backwashing, or undergoing a chemical clean. In MSAD-PCA, an AD-PCA model is constructed for each set to achieve greater precision in the covariance matrix than if all of the sets were aggregated.

Odom et al. ([2018](#ref-odom2018)) investigated MSAD-PCA for a pH fault in the SB-MBR including information about whether a blower was on or off. Odom et al. concluded that given the large number of potential states, states should only be included that have at least observations (where is the number of process variables). Additionally, Odom et al. ([2018](#ref-odom2018)) recommended only including states that are representative of normal function. For example, a state occurring 1% of the time would be considered too infrequent to be included. In this work, each unique combination of 11 state variables (Table 1) is considered a “state.” If the number of observations for a state is less than , then it is not used in the MSAD-PCA monitoring scheme.

While Odom et al. ([2018](#ref-odom2018)) only examined states relative to the function of two air blowers, this work considers all possible equipment and process configurations to define states. Some state variables are binary (e.g., air blower is off, represented by 0, or on, represented by 1) but some can take on more than two values (e.g., bioreactor phase is either 1 for fill, 2 for react, 3 for recirculation with MBR, or 0 for offline). To further improve results, the SBR and MBR are treated separately. Combinations of 4 state variables for the SBR and 7 state variables for the MBR are used (Table 1).

## Fault detection metrics

Klanderman et al. (2019) describe multiple measures of evaluating fault detection methods. For each study, we use the true and false detection rates which quantify a method’s ability to currently identify IC and OC conditions. Here, an “alarm” is defined as three consecutive OC observations for a particular method. The true detection rate (TDR) is the total number of true alarms (i.e., alarmed OC observations) divided by the total number of OC observations. The false-detection rate (FDR) is calculated by dividing the number of false alarms (i.e., alarmed IC observations) by the total number of IC observations. The OC period is assessed by a method’s time-to-detection (TTD): the time between when a fault occurs and when an alarm is signaled. Each method’s TTD is compared to the TTD for the Mines Park facility’s build-in alarm system that uses single-variable monitoring with operator determined setpoints (i.e., SCADA system). If TTD is positive, it indicates that the *T2* or *SPE* alarm predicted the fault prior when operator’s identified it. If the TTD is negative, it indicates that the *T2* or *SPE* alarm detected the OC conditions after the fault.

## Case study procedure

All data analysis is performed in the open-source statistical software R, and AD-PCA is executed using a modified version the mvMonitoring package developed by Odom et al. ([2018](#ref-odom2018)). To apply SSAD-PCA and MSAD-PCA to the SB-MBR data, the following procedure is followed for each case study: 1. Establish initial case study parameters (e.g., α, faults-to-trigger-alarm, state and process variables, lag-size) 2. Compile and clean raw data (e.g., remove process variables with insufficient unique observations in each state to calculate the covariance matrix and eigenvectors appropriately) 3. Train the PCA model and calculate *SPE* or *T2* thresholds 4. Apply the PCA model by calculating *SPE* or *T2* for each new observation in 1 day 5. For the adaptive PCA modification: remove observations that exceed the *SPE* or *T2* thresholds and remove observations from the earliest training day. Re-train the PCA model and calculate new *SPE* or *T2* thresholds 6. Compare the actual IC and OC observations (operator-determined) to the observations alarmed by *SPE* or *T2* by calculating detection rate metrics 7. Iterate through Steps 2-6 by varying case study parameters until the desired detection rate is achieved for a given dataset

AD-PCA was performed on the earliest observations in a dataset for the given number of training days (ntraining) (Step 3). Observations from the following day (day [ntraining + 1]) are classified as IC or OC (Step 4). The training dataset is updated by removing the oldest observations (day 1) and add IC observations from day [ntraining + 1] (Step 5). Data was compiled such that each case study dataset included [2·ntraining + ntesting] days, where ntesting is the number of testing days to compare (Step 2). This ensured that the final AD-PCA model used for testing would not include any of the original/first training dataset, in the event that there were OC conditions not detected by operators.

All case studies described here used the following AD-PCA parameters: PCs retained = 0.9; α = 0.01; faults-to-trigger-alarm = 3; lag-size = 1; re-trained every 1440 observations (e.g., every day at midnight for SSAD-PCA). MSAD-PCA included either SBR or MBR process variables. States were identified by concatenating state variables (Table 1) and only retaining those with a sufficient number of observations (). Additionally, variables were only included that had sufficient number of unique observations to conduct singular value decomposition for PCA (> 7, found by experimental iteration). Autocorrelation function (ACF) plots were used to identify the necessary lag for the dynamic PCA extension. For most variables, a lag of a one timestep is found to account for the most autocorrelation (see [Appendix](#appendix)).

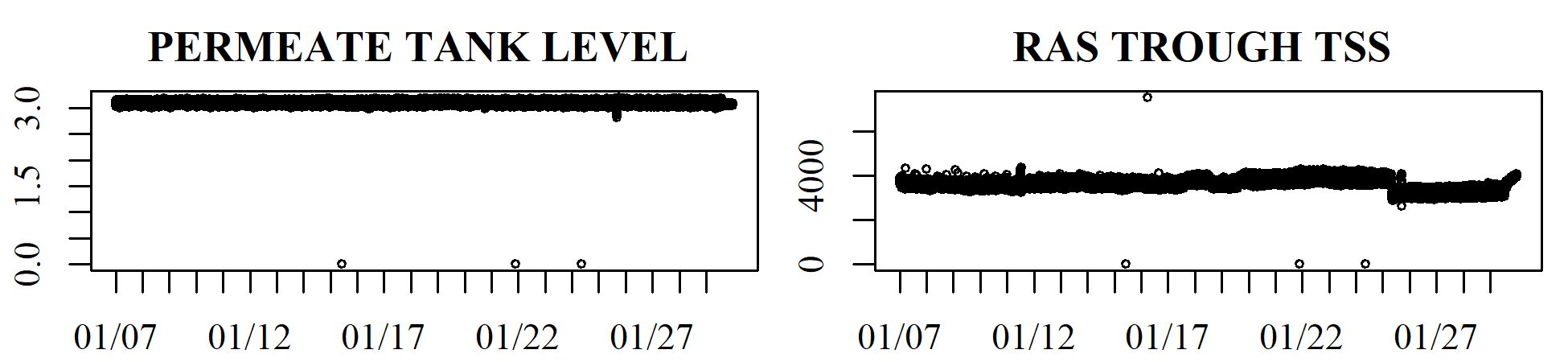
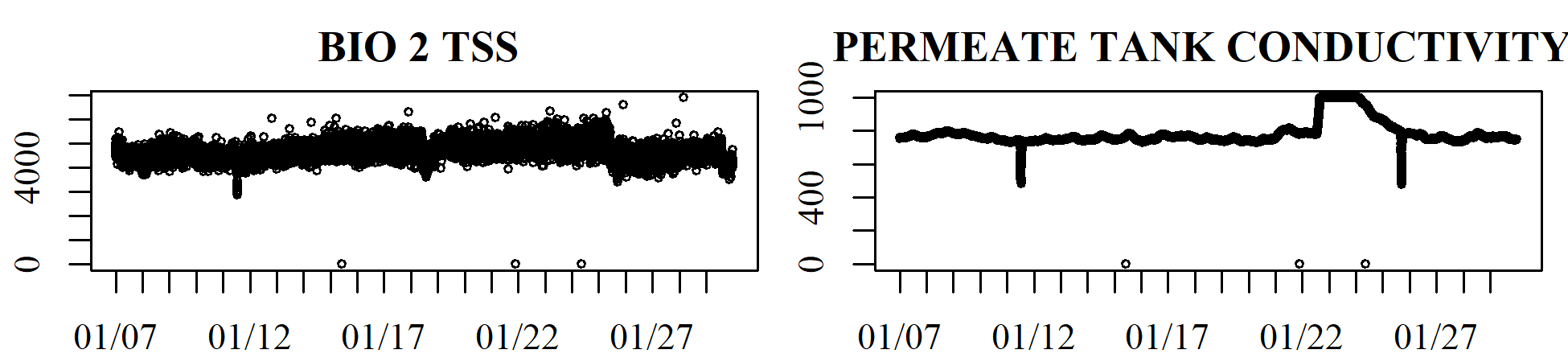
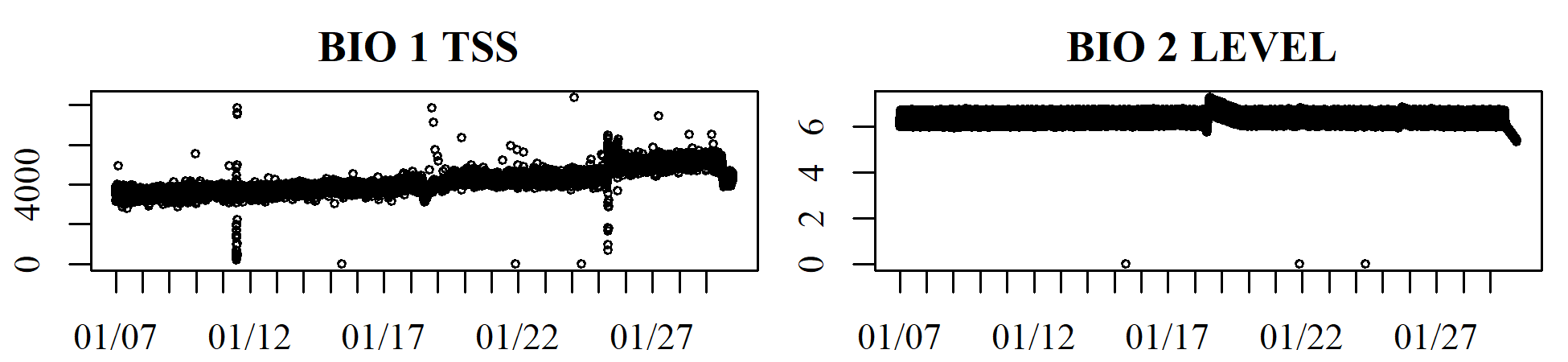
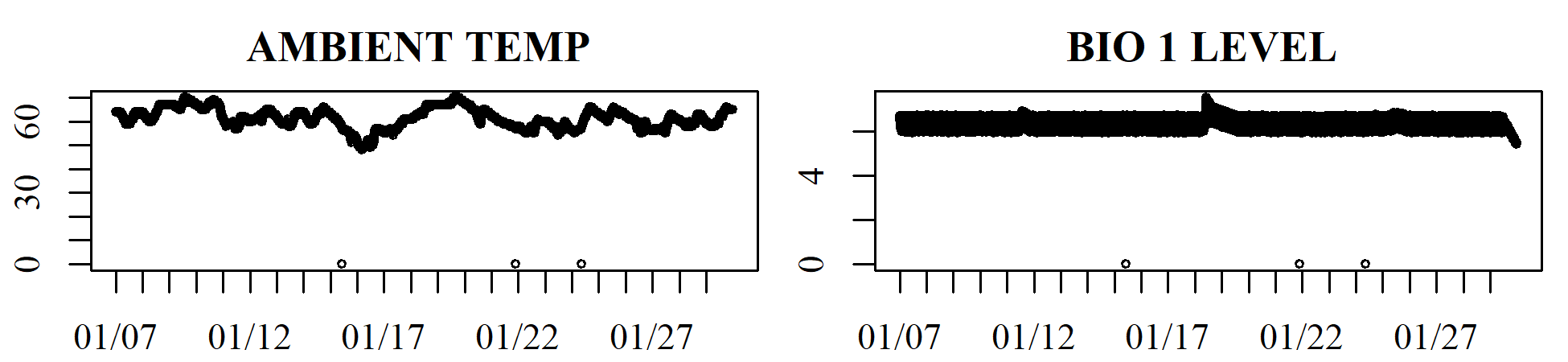
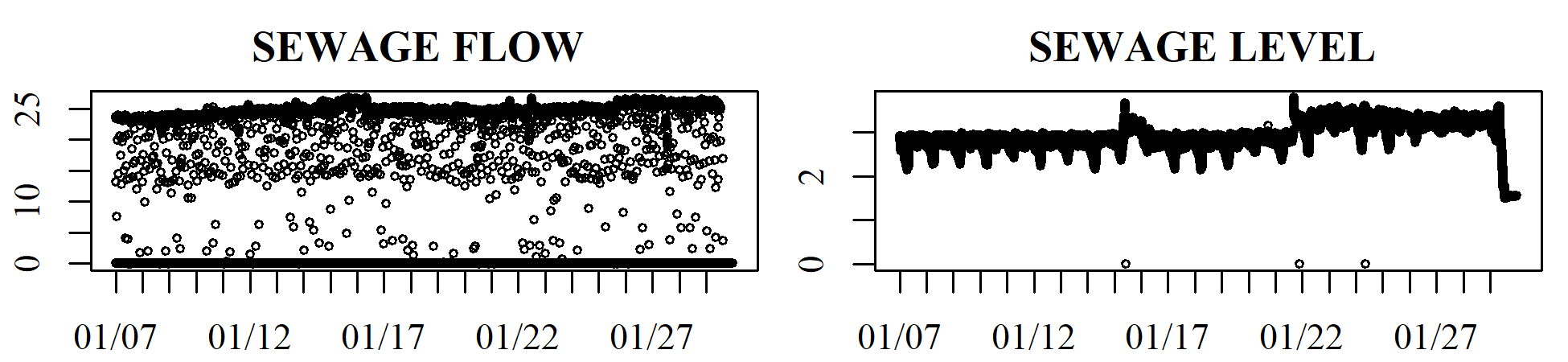
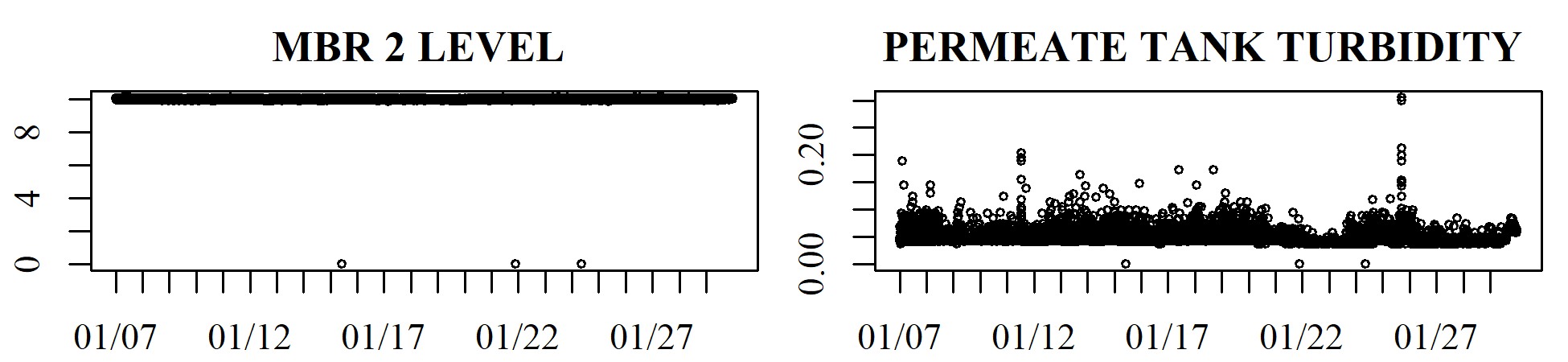
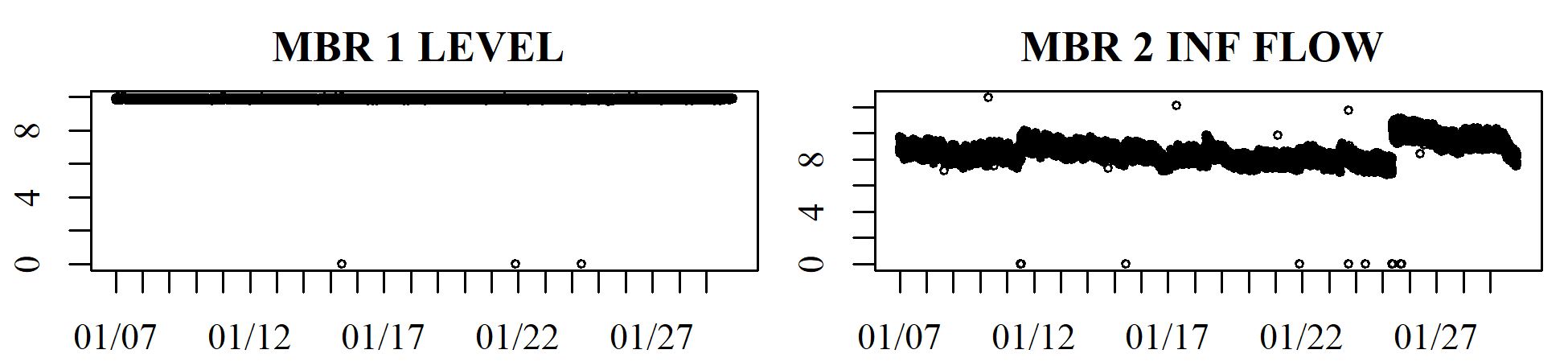
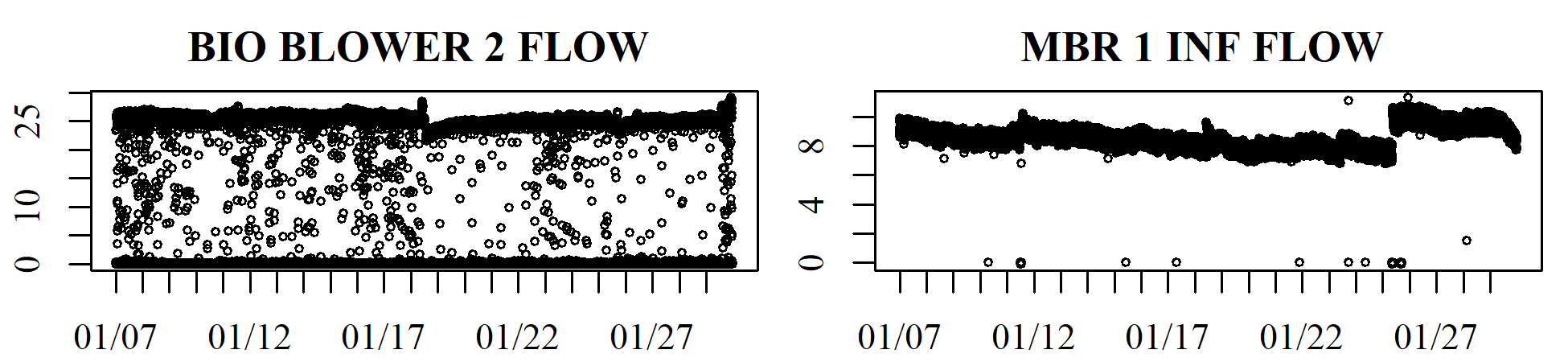
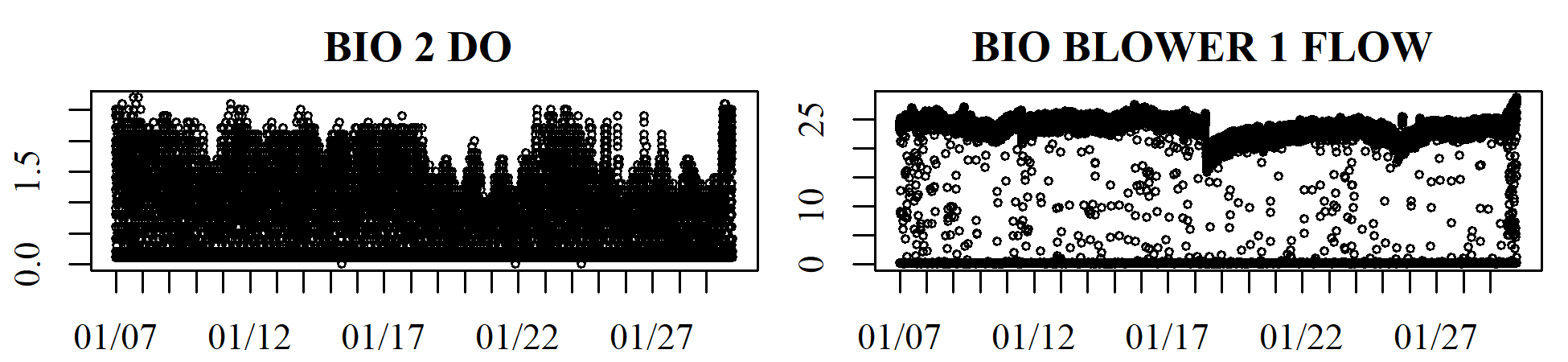
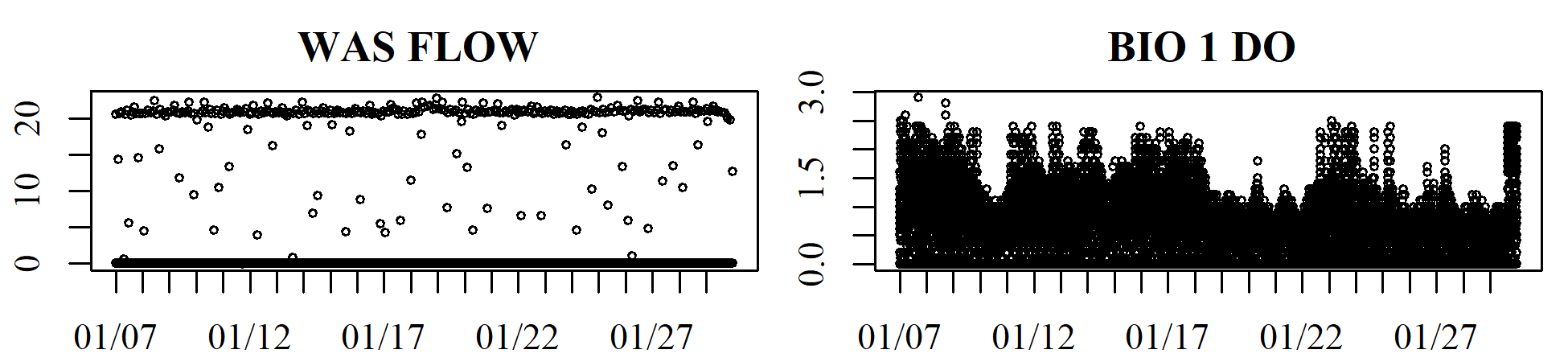
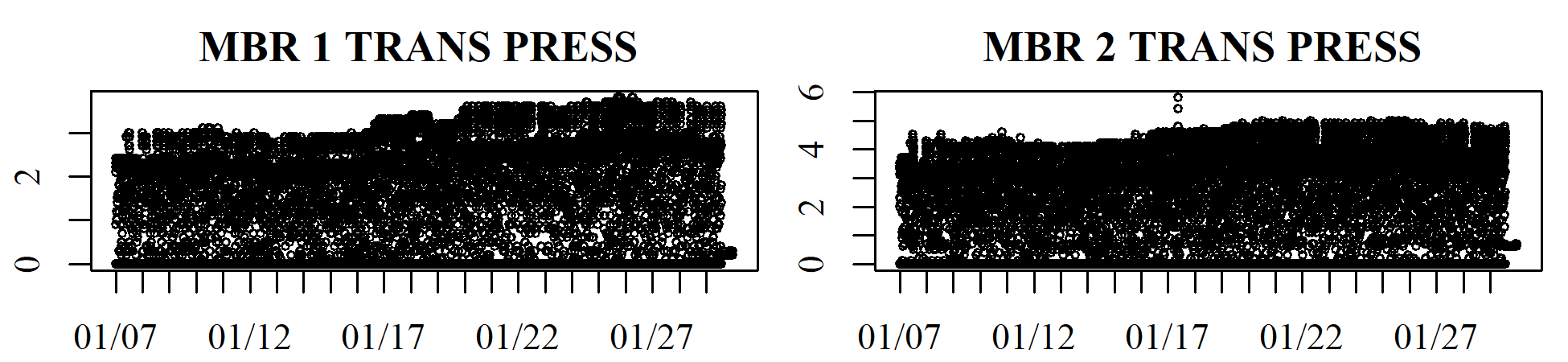
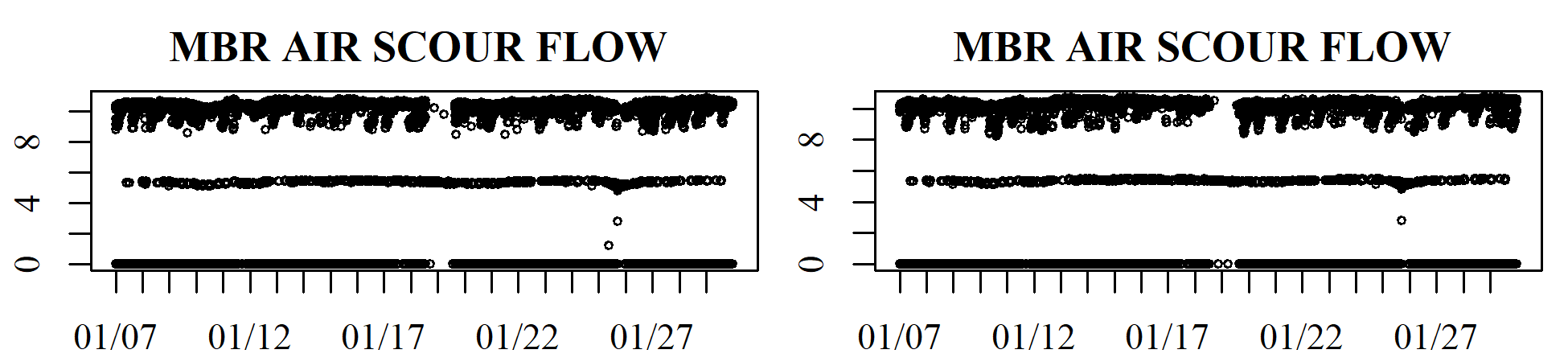
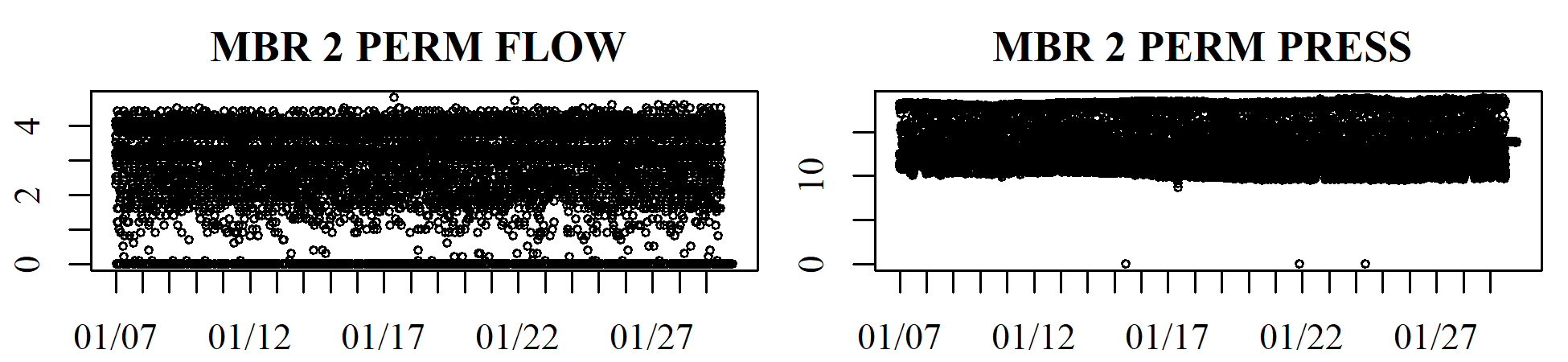
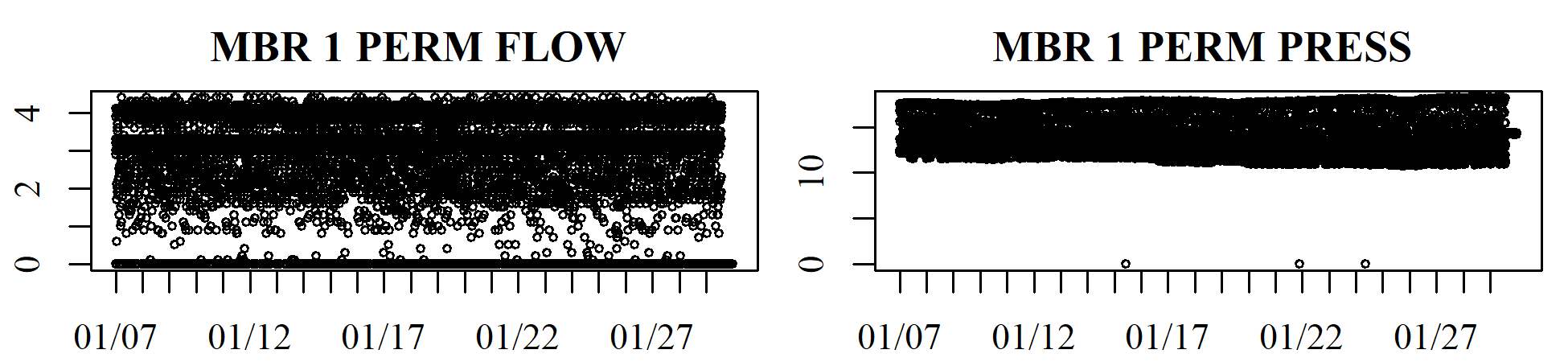
# Results

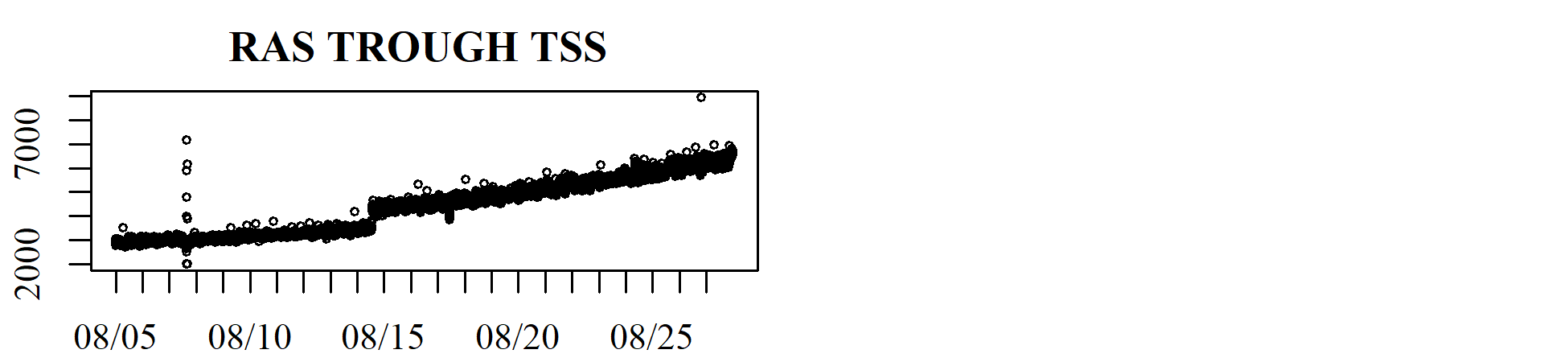
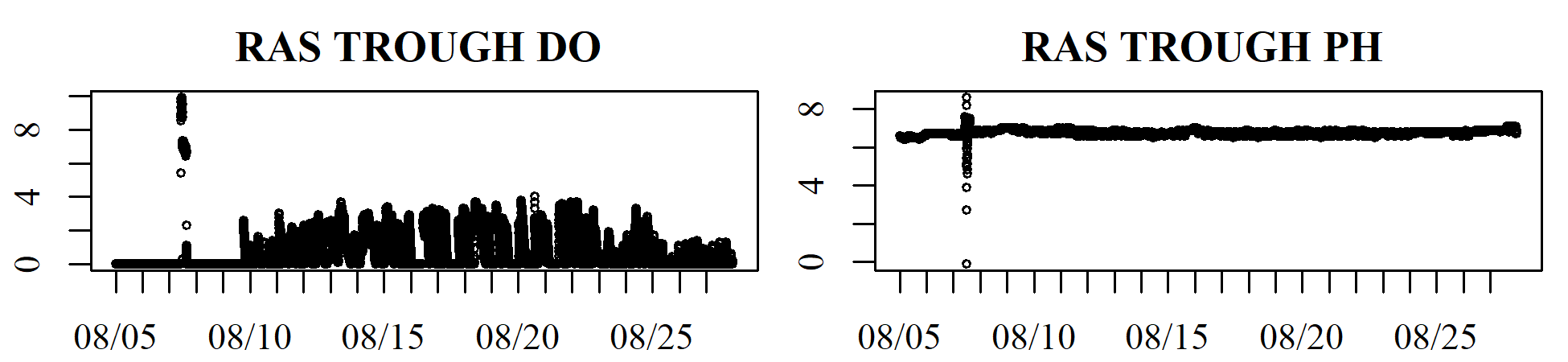
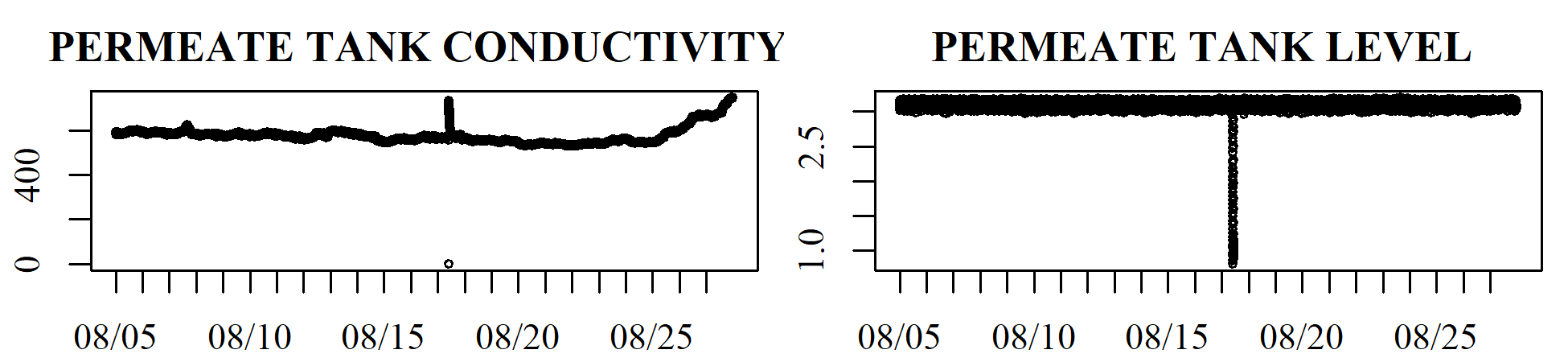
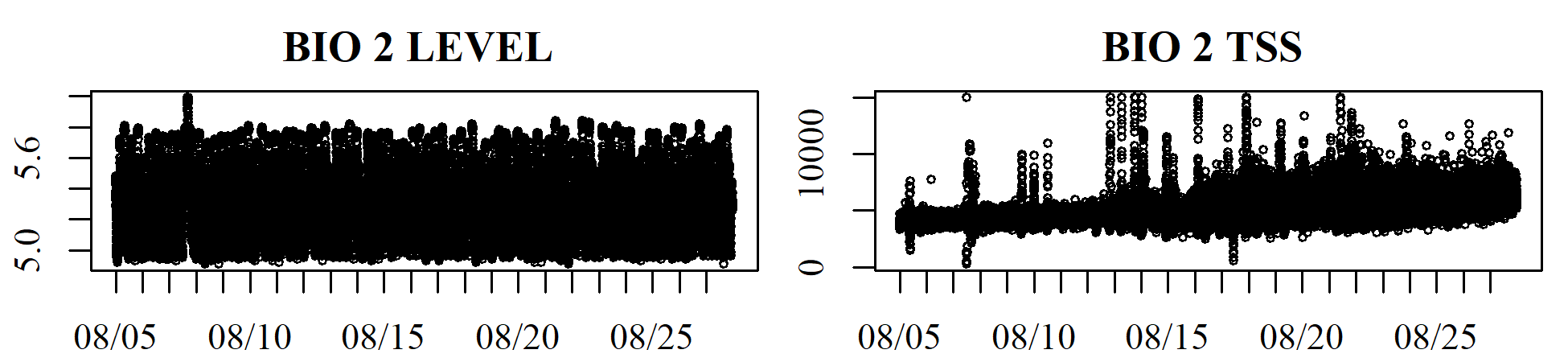
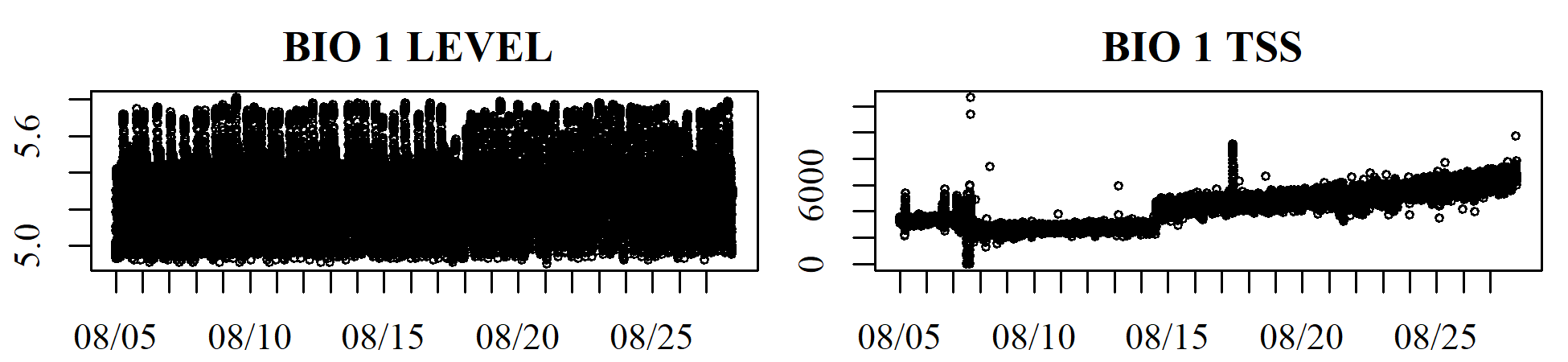
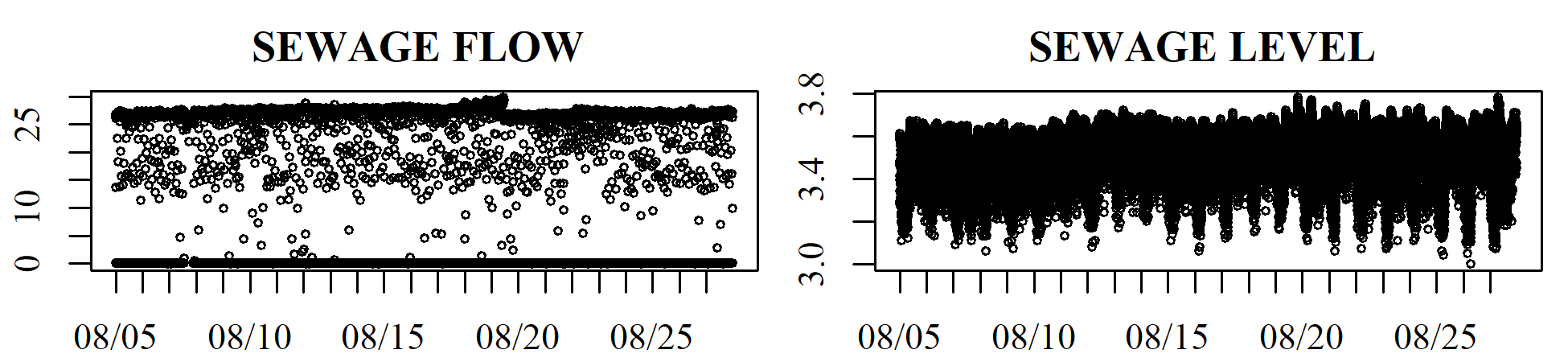
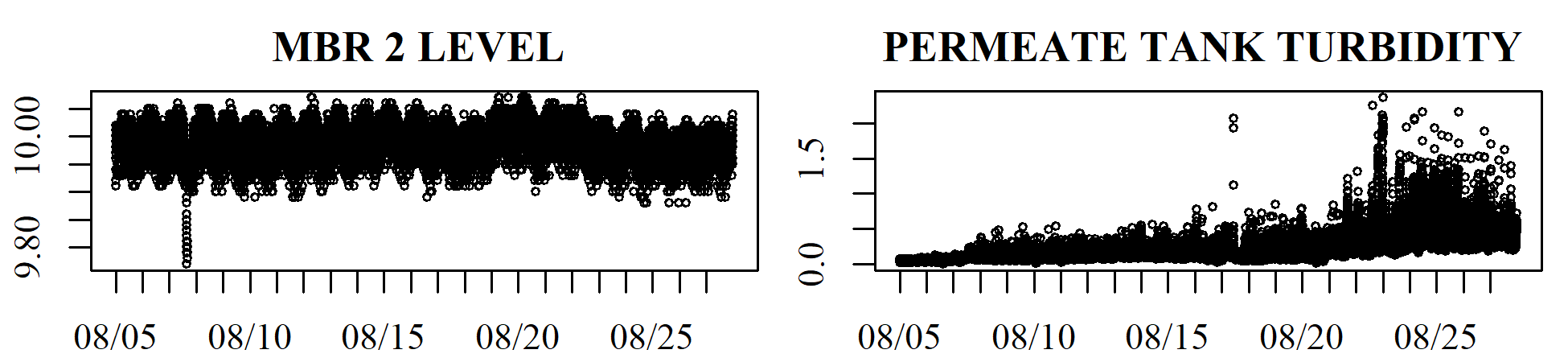
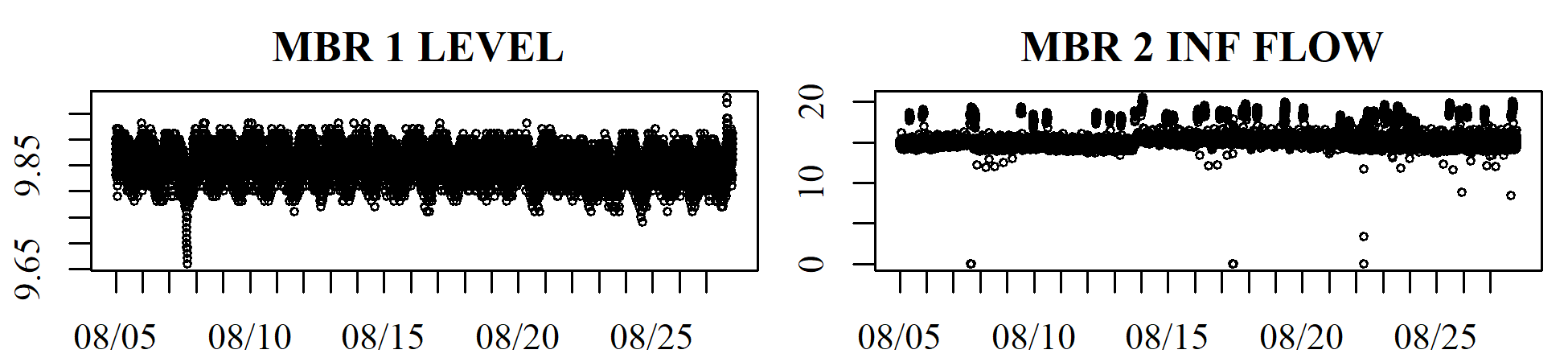
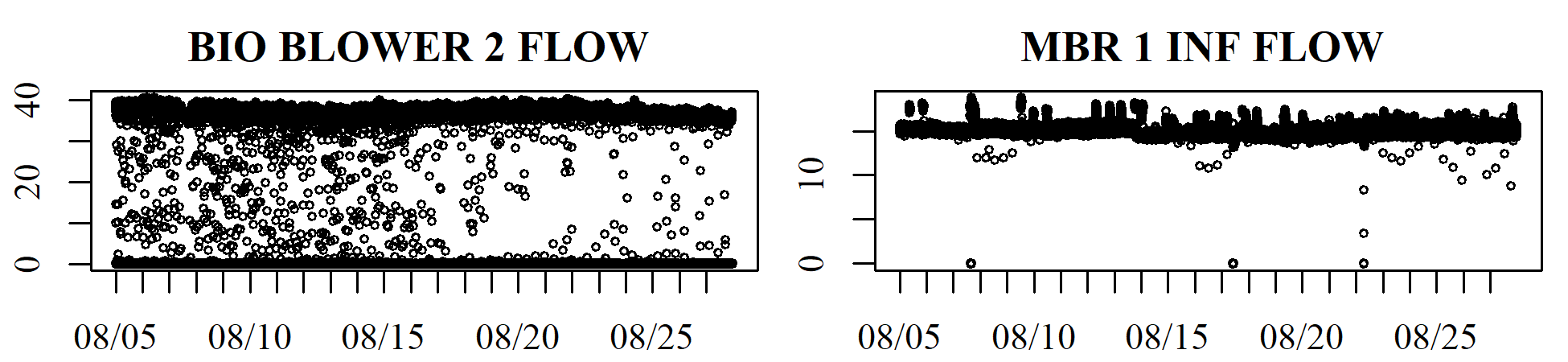
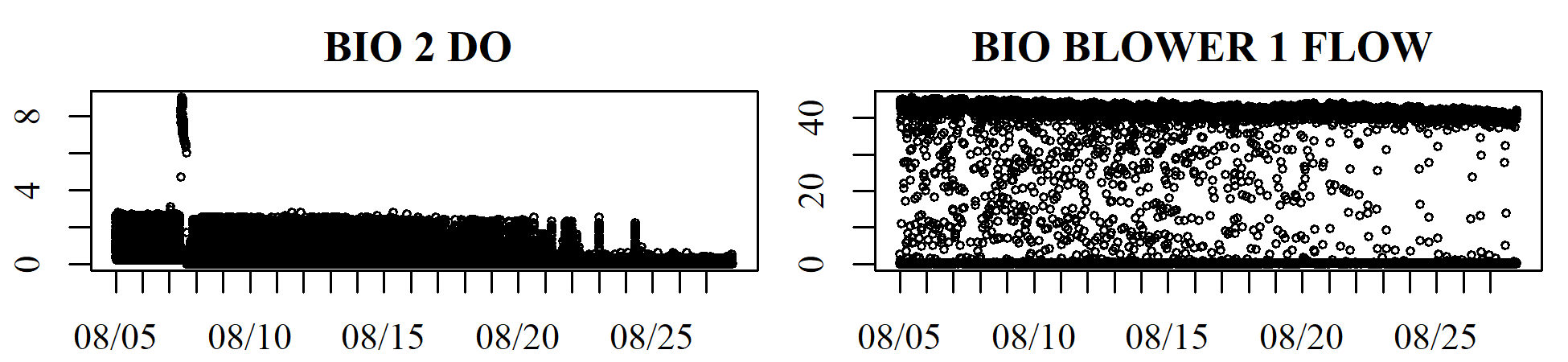
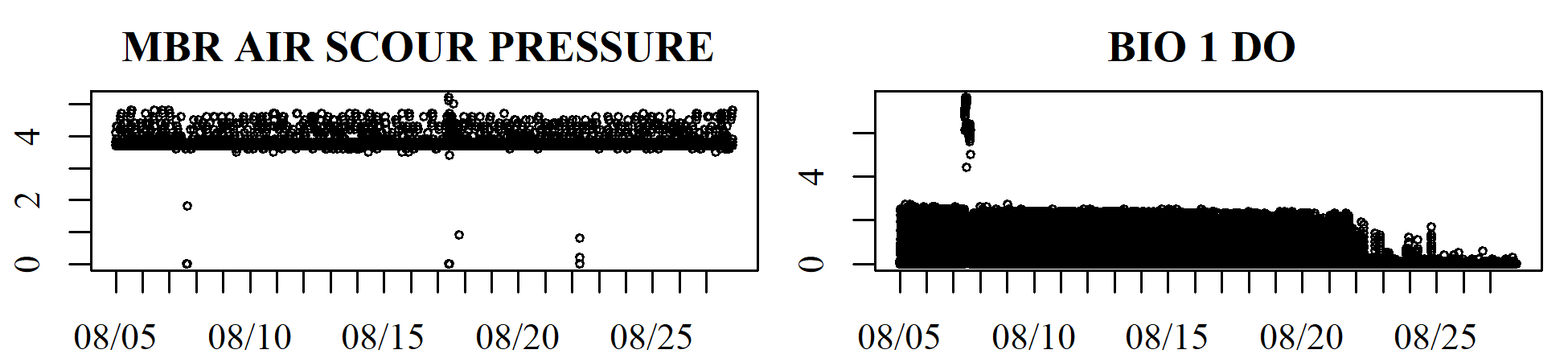
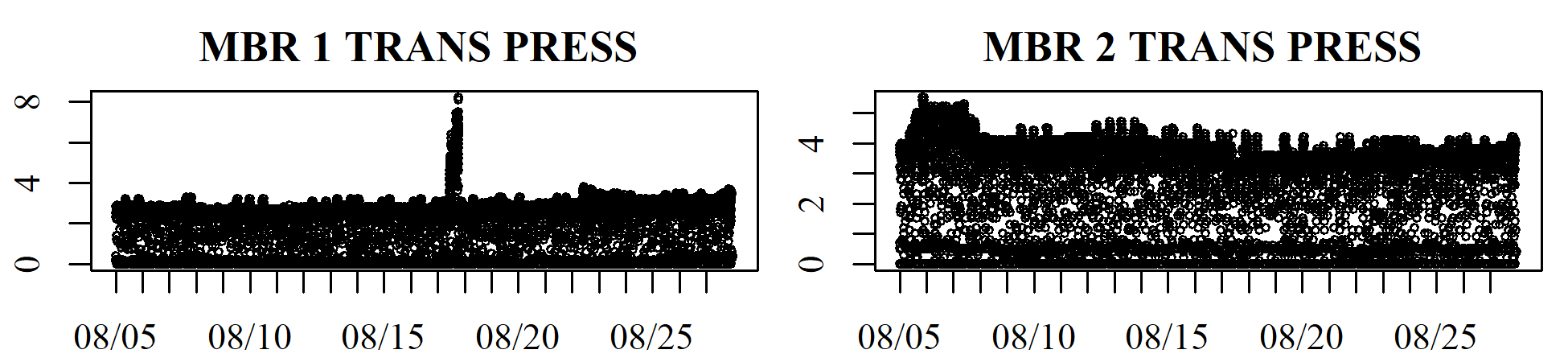
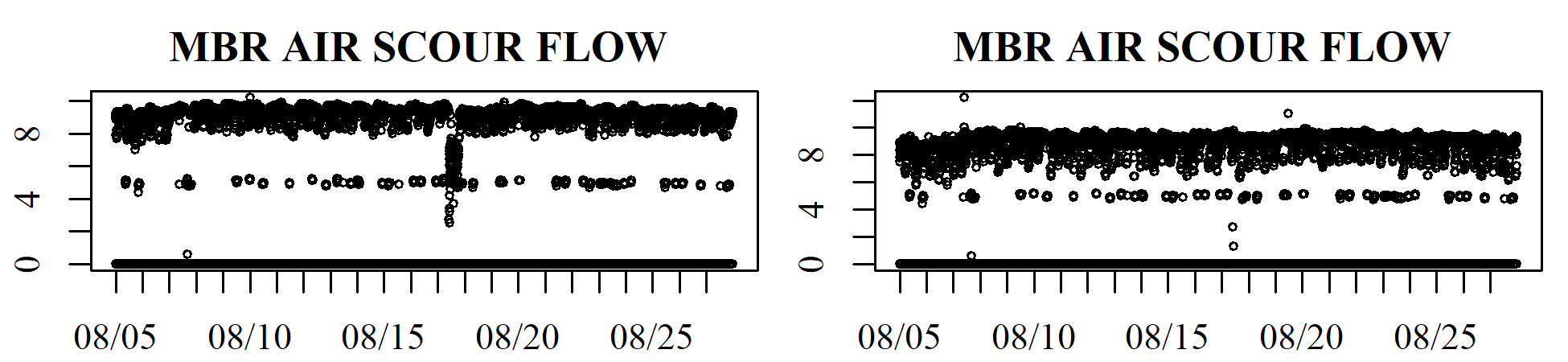
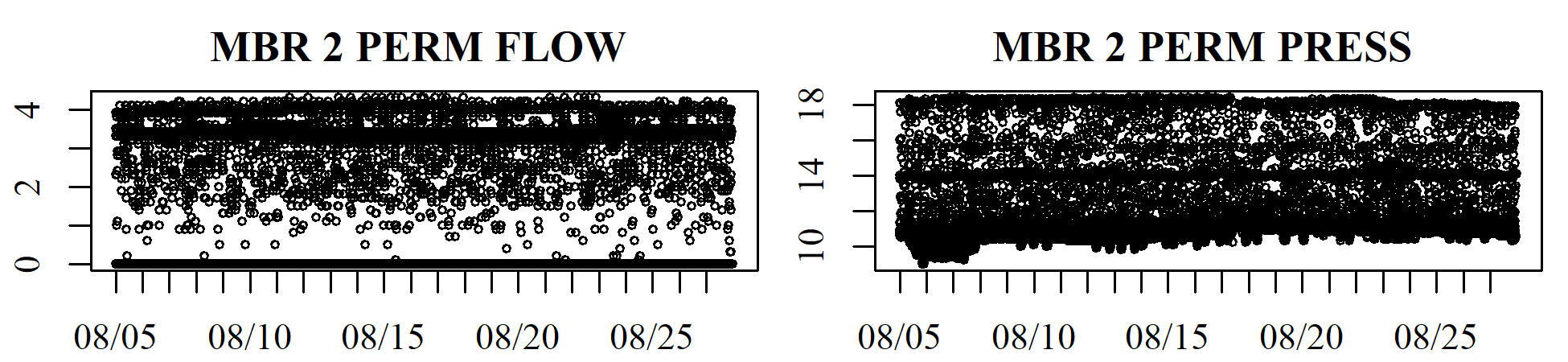
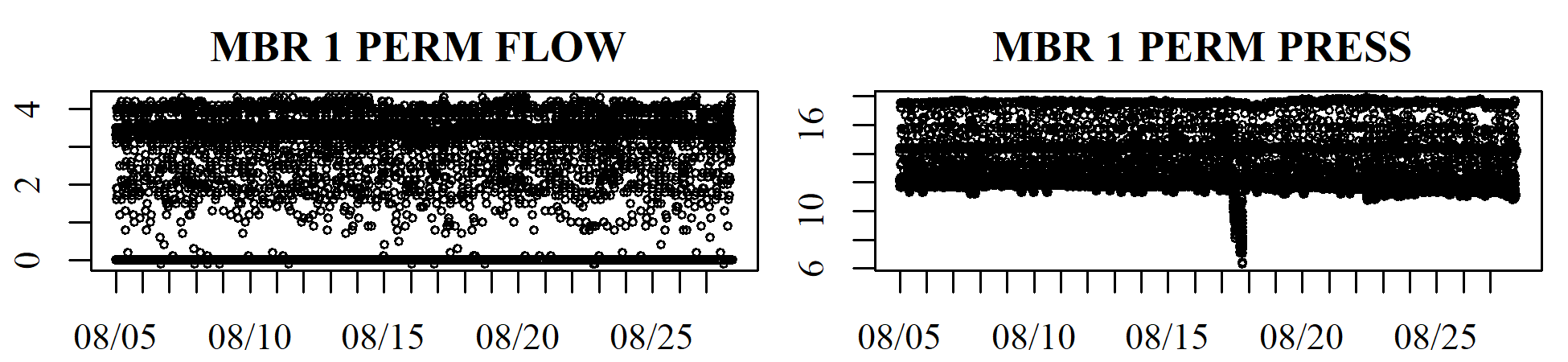
Multiple instances of process faults occurred at the Mines Park WWTP between January 2018 and August 2018. Here, we highlight select instances to illustrate a range of different type of faults (e.g., isolated, sustained, transient, and drift faults, see Newhart, Holloway, Hering, & Cath, [2019](#ref-newhart2019)) which have varying degrees of consequences for the operation of the SB-MBR. For the ideal scenarios, Tables 3-5 summarize the number of observations for each study, the TDR, and the FDR. These studies are referred to as “ideal” because the largest initial training window contained no OC observations.

Table 2. Total number of observations used in the initial training and testing/re-training, and the total number of faulty observations for each study.

# Appendix

## Timeseries

 Figure app.fig<-app.fig+1;app.fig. Pump failure timeseries

 Figure app.fig<-app.fig+1;app.fig. Influent change 1 timeseries

Alves, C., Henning, E., Samohyl, R., Cristina Konrath, A., & Formigoni Carvalho Walter, O. (2013). Application of Multivariate Control Charts for Monitoring an Industrial Process. *Tecno-Lógica*, *17*(2), 101–107. <https://doi.org/10.17058/tecnolog.v17i2.3636>

Baggiani, F., & Marsili-Libelli, S. (2009). Real-time fault detection and isolation in biological wastewater treatment plants. *Water Sci. Technol.*, *60*(11). <https://doi.org/10.2166/wst.2009.723>

Banadda, N., Nhapi, I., & Kimwaga, R. (2011). A review of modeling approaches in activated sludge systems. *African Journal of Environmental Science and Technology*, *5*(6), 397–408.

Berthouex, P. M. (1989). Constructing Control Charts for Wastewater Treatment Plant Operation. *J. Water Pollut. Control Fed.*, *61*(9/10), 1534–1551.

Choi, S. W., & Lee, I.-B. (n.d.). Multiblock PLS-based localized process diagnosis. *Journal of Process Control*, *15*(3), 295–306. <https://doi.org/https://doi.org/10.1016/j.jprocont.2004.06.010>

Corominas, L., Villez, K., Aguado, D., Rieger, L., Rosén, C., & Vanrolleghem, P. A. (2011). Performance evaluation of fault detection methods for wastewater treatment processes. *Biotechnology and Bioengineering*, *108*(2), 333–344.

Dürrenmatt, D. J., & Gujer, W. (2012). Data-driven modeling approaches to support wastewater treatment plant operation. *Environmental Modelling & Software*, *30*, 47–56. <https://doi.org/10.1016/j.envsoft.2011.11.007>

Jackson, J. E. (1991). *A User’s Guide To Principal Components*. John Wiley & Sons, Inc.

Jałowiecki, Ł., Chojniak, J. M., Dorgeloh, E., Hegedusova, B., Ejhed, H., Magnér, J., & Płaza, G. A. (2016). Microbial Community Profiles in Wastewaters from Onsite Wastewater Treatment Systems Technology. *PLOS ONE*, *11*(1), e0147725. <https://doi.org/10.1371/journal.pone.0147725>

Kazor, K., Holloway, R. W., Cath, T. Y., & Hering, A. S. (2016). Comparison of linear and nonlinear dimension reduction techniques for automated process monitoring of a decentralized wastewater treatment facility. *Stochastic Environmental Research and Risk Assessment*, *30*(5), 1527–1544. <https://doi.org/10.1007/s00477-016-1246-2>

King, K. L., Wang, Z., & Kroll, D. J. (2006). *Patent No. 3723429343*.

Kourti, T., Lee, J., & Macgregor, J. F. (1996). Experiences with industrial applications of projection methods for multivariate statistical process control. *Computers & Chemical Engineering*, *20*, S745–S750. <https://doi.org/10.1016/0098-1354(96)00132-9>

Kruger, U., Zhou, Y., & Irwin, G. W. (2004). Improved principal component monitoring of large-scale processes. *Journal of Process Control*, *14*(8), 879–888. <https://doi.org/10.1016/j.jprocont.2004.02.002>

Ku, W., Storer, R. H., & Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *30*(1), 179–196. <https://doi.org/10.1016/0169-7439(95)00076-3>

Lee, C., Choi, S. W., & Lee, I. B. (2006). Sensor fault diagnosis in a wastewater treatment process. *Water Sci. Technol.*, *53*(1), 251–257. <https://doi.org/10.2166/wst.2006.027>

Lee, D. S., Lee, M. W., Woo, S. H., Kim, Y.-J., & Park, J. M. (2006). Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant. *Process Biochemistry*, *41*(9), 2050–2057. <https://doi.org/10.1016/j.procbio.2006.05.006>

Lee, D. S., & Vanrolleghem, P. A. (2004). Adaptive Consensus Principal Component Analysis for On-Line Batch Process Monitoring. *Environmental Monitoring and Assessment*, *92*, 119–135.

MacGregor, J. F., & Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, *3*(3), 403–414.

Mina, J., & Verde, C. (2007). Fault Detection for Large Scale Systems Using Dynamic Principal Components Analysis with Adaptation. *International Journal of Computers Communications & Control*, *II*(2), 185–194.

Newhart, K. B., Holloway, R. W., Hering, A. S., & Cath, T. Y. (2019). Data-driven performance analyses of wastewater treatment plants: A review. *Water Research*, *157*, 498–513. <https://doi.org/10.1016/j.watres.2019.03.030>

Odom, G. J., Newhart, K. B., Cath, T. Y., & Hering, A. S. (2018). Multistate multivariate statistical process control. *Applied Stochastic Models in Business and Industry*, *34*(6), 880–892. <https://doi.org/10.1002/asmb.2333>

Olsson, G., Nielsen, M., Yuan, Z., Lynggaard-Jensen, A., & Steyer, J.-P. (2005). *Instrumentation, control and automation in wastewater systems*. IWA publishing.

Pankavich, S., & Swanson, R. (2015). Principal Component Analysis: Resources for an Essential Application of Linear Algebra. *PRIMUS*, *25*(5), 400–420. <https://doi.org/10.1080/10511970.2014.993446>

Prieto, A., Vuono, D., Holloway, R., Benecke, J., Henkel, J., Cath, T., … Drewes, J. (2013). Decentralized Wastewater Treatment for Distributed Water Reclamation and Reuse: The Good, The Bad, and The UglyExperience from a Case Study. *ACS Symposium Series*, *1123*, 251–266. Citeseer.

Qin, X., Gao, F., & Chen, G. (n.d.). Wastewater quality monitoring system using sensor fusion and machine learning techniques. *Water Research*, *46*(4), 1133–1144. <https://doi.org/https://doi.org/10.1016/j.watres.2011.12.005>

Rosen, C., & Lennox, J. A. (2001). Multivariate and multiscale monitoring of wastewater treatment operation. *Water Research*, *35*(14), 3402–3410. <https://doi.org/10.1016/S0043-1354(01)00069-0>

Rosen, C., Röttorp, J., & Jeppsson, U. (2003). Multivariate on-line monitoring: Challenges and solutions for modern wastewater treatment operation. *Water Sci. Technol.*, *47*(2), 171–179.

Schraa, O., Tole, B., & Copp, J. B. (2006). Fault detection for control of wastewater treatment plants. *Water Sci. Technol.*, *53*(4-5), 375–382. <https://doi.org/10.2166/wst.2006.143>

Wallace, J., Champagne, P., & Hall, G. (2016). Multivariate statistical analysis of water chemistry conditions in three wastewater stabilization ponds with algae blooms and pH fluctuations. *Water Research*, *96*, 155–165. <https://doi.org/10.1016/j.watres.2016.03.046>

Wang, X., Ratnaweera, H., Holm, J. A., & Olsbu, V. (2017). Statistical monitoring and dynamic simulation of a wastewater treatment plant: A combined approach to achieve model predictive control. *Journal of Environmental Managemetn*, *193*, 1–7. <https://doi.org/10.1016/j.jenvman.2017.01.079>

Wise, B. M., & Gallagher, N. B. (1996). The process chemometrics approach to process monitoring and fault detection. *J. Proc. Cont.*, *6*(6), 329–348. <https://doi.org/10.1016/0959-1524(96)00009-1>

Wold, S. (1994). Exponentially weighted moving principal components analysis and projections to latent structures. *Chemometrics and Intelligent Laboratory Systems*, *23*, 149–161.