Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

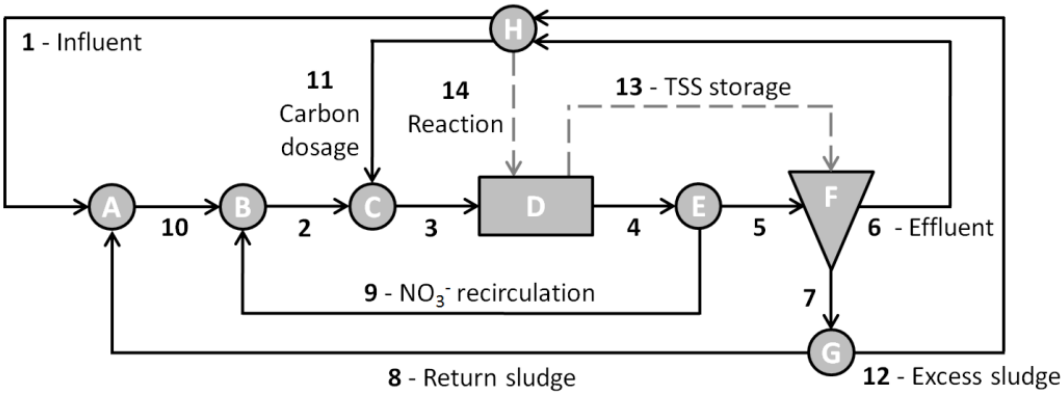
eawag
aquatic research

Master Thesis

Data Reconciliation and Gross Error Detection for Wastewater Treatment Processes

Urs Schönenberger

3rd August 2015



Head
Prof. Dr. Eberhard Morgenroth
Chair of Process Engineering in Urban Water Management, ETH Zürich

Supervisor
Dr. Kris Villez, eawag

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Data Reconciliation and Gross Error Detection for Wastewater Treatment Processes

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Schönenberger	Urs

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Uzwil, 02.08.2015	

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

Abstract

Sensors on wastewater treatment plants are often influenced by measurement errors. These errors can seriously affect the control and monitoring of the processes occurring in these facilities. In this thesis it is investigated to which extent mass balancing methods can be used to reduce random errors and to detect gross errors in the measurements of hydraulic flows and TSS concentrations of wastewater treatment plants.

Therefore, a bilinear data reconciliation technique is implemented and evaluated by means of the Benchmark Simulation model No. 1 – Long Term (BSM1_LT). The BSM1_LT is a wastewater treatment plant simulation platform which was developed to compare different error detection methods. In this thesis the following gross error detection methods are assessed using the BSM1_LT: Global test, univariate CUSUM charts and multivariate CUSUM charts.

The results show that bilinear data reconciliation is able to reduce random errors in the measurements of hydraulic flows and TSS component flows. CUSUM charts are found to show a much higher gross error detection power than the global test. Furthermore, it is revealed that accounting for TSS storage and TSS reaction processes is essential for the success of the implemented data reconciliation and gross error detection methods.

Acknowledgements

I would like to say a big ”thank you“ to the supervisor of my thesis, Dr. Kris Villez, for his valuable suggestions and for his great support during this time.

Furthermore, to Prof. Dr. Morgenroth, Dr. Nicolas Derlon and Dr. Joao Paulo Leitao I want to express my sincere gratitude for challenging me with their critical questions.

Finally, I'd like to thank Thomas Schönenberger and Ian Gardner for proofreading this report.

Contents

1 Introduction

In this section, first the background and the motivation for this thesis are illustrated. Afterwards the goals of the thesis are explained.

1.1 Motivation

Wastewater treatment plants and sewer systems play a fundamental role for the health of ecosystems and human beings by preventing pollutants from contaminating rivers, lakes and aquifers. However, continuous investments in these assets are needed. After Maurer & Herlyn (2006) the annual costs for sewage disposal and wastewater treatment in Switzerland equal 1.7 billion CHF. This amount is caused by the sewer system and the wastewater treatment plants (WWTPs) in equal shares. In this thesis the focus is set on WWTPs. Scientists, engineers and operators are putting much effort into improving the performance of these facilities and lowering their operational costs. These efforts depend strongly on reliable information about the processes occurring in these facilities. Therefore, on almost every WWTP process data are collected. After Spindler (2014) they reflect its performance and are used for "[...] monitoring, benchmarking and simulation, to adjust control strategies and to plan for process redesign or plant extension." The same author states the poor quality of process data to be the main obstacle for such tasks. For process control, Villez *et al.* (2013a) showed that the effect of a gross error in a certain sensor on the calculation of the solids retention time (SRT) can be dramatic. In absence of any corrective action this can lead to washout of the nitrifying biomass. Thus, in order to prevent such problems, efforts should be made on improving the quality of process data.

The errors in the process data originate from the harsh environment in which the sensors are installed. A proper maintenance and installation of the sensors plus plausibility checks might reduce the frequency of gross errors in the measurements. However, at least some of the gross errors as well as random errors cannot be avoided. To improve the data quality, different authors (e.g. Rieger *et al.* (2010); Spindler (2014); Meijer *et al.* (2002)) recommend to use data reconciliation in combination with gross error detection methods. These methods allow to reduce random errors, to detect gross errors and to isolate faulty sensors.

From the wide range of gross error detection methods in this thesis two main types of statistical tests were implemented: a) A test exploiting only the spatial redundancy (global test) and b) tests exploiting the spatial and the temporal redundancy (cumulative sum (CUSUM) charts).

In order to evaluate the performance of these methods the WWTP model "Benchmark Simulation Model No. 1 – Long Term (BSM1_LT)" (Gernaey *et al.*, 2014) was used. The model was developed to create an objective way to compare process monitoring and error detection methods and has already been applied by different authors for such comparisons (e.g. Corominas *et al.* (2011) or Villez *et al.* (2013a)).

1.2 Goals

The objectives of this thesis are defined as follows:

1. Implementation of a bilinear data reconciliation algorithm for the Benchmark Simulation Model No. 1 – Long Term. It should be able to handle varying degrees of structural observability, structural redundancy and measurement uncertainty while (i) reconciling random measurement errors, (ii) detecting gross errors and (iii) isolating faulty sensors.

2. Implementation of a technique based on CUSUM charts to enhance gross error detection and isolation accuracy.

3. Evaluation of the implemented techniques in terms of data reconciliation, fault detection and isolation accuracy.

2 Theory

In the following an introduction into the theory of errors, data reconciliation and gross error detection is given. These explanations are mainly based on the book of Narasimhan & Jordache (1999).

2.1 Errors

As mentioned in the introduction, a fundamental issue of working with process data is the occurrence of errors. The total error in a measurement can be divided into two types of errors: Random errors and gross errors (see equation 2.1).

$$\tilde{x} = x + \epsilon + \delta \tag{2.1}$$

$\tilde{x}$: Measured value
x: Real value of measured variable
ϵ: Random error
δ: Magnitude of gross error

Random errors can be caused by various reasons such as signal conversion noise or analog input filtering. Neither their magnitude nor their sign can be predicted with certainty, but they can be characterized by means of probability distributions. Although these errors cannot be completely eliminated, techniques like data reconciliation (see chapter 2.2) are able to reduce them.

In contrast to that *gross errors* are deterministic. That is, they have a specific magnitude and sign at a given time. They can result from inaccurate calibration of the instruments, wear and tear of sensors, biofouling, scaling or improper installation of the measurement devices. While fouling and scaling can be avoided by careful installation and maintenance of the sensors, other errors will remain anyhow. Thus, they must be detected and eliminated with help of gross error detection methods. These are explained in chapter 2.3.

Regarding the different types of gross errors (e.g. bias, failures, drifting), in this thesis the attention was directed at biases only. After Dunia *et al.* (1996) a bias corresponds to a constant deviation of the measured value from the real value. In the following, biases are assumed to occur randomly at any point in time. As soon as they appear they are supposed to persist during the rest of the time series.

2.2 Data reconciliation

Narasimhan & Jordache (1999) define *data reconciliation* as follows: "Data reconciliation is a technique [...] to improve the accuracy of measurements by reducing the effect of random errors in the data." In contrast to other error reduction methods, data reconciliation uses constraints to reach a reduction of random errors. These are either equality constraints

(e.g. mass balances, energy balances or equilibrium constraints) or inequality constraints (e.g. thermodynamic feasibility constraints or bounds). By definition, the estimates of process variables obtained by data reconciliation satisfy the imposed constraints. In this thesis only mass balance constraints were used. With help of the incidence matrix A, they can be written as in equation 2.2. The first chapter of Narasimhan & Jordache (1999) gives detailed explanations about how to deploy the incidence matrix.

$$\mathbf{A}\mathbf{x} = \mathbf{0} \tag{2.2}$$

A: Incidence matrix
x: Vector of real values of the measured variables

The above equation is not satisfied when measurement errors are present. With data reconciliation the measured values are adjusted such that they fulfil the mass balance equations. Thereby, it is solved for the solution leading to the smallest sum of squares of the adjustments in relation to the accuracy of the respective measurements. Thus, the data reconciliation problem can be written as in equation 2.3 (Sanchez & Romagnoli, 1996). To ensure that data reconciliation is effective, no gross errors should be present.

$$\begin{aligned} \min \ & (\hat{\mathbf{x}} - \tilde{\mathbf{x}})^T \Psi_x^{-1} (\hat{\mathbf{x}} - \tilde{\mathbf{x}}) \\ \text{s.t.} \ & \mathbf{A}\hat{\mathbf{x}} = \mathbf{0} \end{aligned} \tag{2.3}$$

$\tilde{\mathbf{x}}$: Vector of measured values
$\hat{\mathbf{x}}$: Vector of reconciled values
Ψ_x: Covariance matrix of the measurements
A: Incidence matrix

"Data reconciliation and gross error detection both achieve error reduction only by exploiting the [spatial] redundancy property of measurements." (Narasimhan & Jordache, 1999) Therefore spatial redundancy is a mandatory requirement for using these error reduction methods. A measurement is said to be *redundant* if there are more measurements than necessary to determine it. If a variable can just be determined with the measurements available it is said to be *observable*. However, if the variable cannot be dermined with the existing measurements, it is called *unobservable*.

If chemical reactions are occurring in the system they are violating the mass balance constraints. Even in absence of measurement errors the latter are not fulfilled. Crowe *et al.* (1983) added imaginary unmeasured flows to the system in order to account for this problem. The data reconciliation algorithm therefore is allowed to choose the extent of the imaginary flow to close the balance (equation 2.3). However, the addition of each imaginary

flow adds one degree of freedom to the system which can result in a loss of redundancy or observability of certain variables.

In the case of a wastewater treatment plant, usually two types of mass balance constraints equations are imposed: Those on hydraulic flow rates (Q) and those on component flow rates ($f = Q \cdot c$). For the reconciliation of the hydraulic flow rates the mass balances constraints are based on one variable (the hydraulic flow itself) only, which can directly be measured. Thus, a linear data reconciliation problem results. In contrast to that, for the reconciliation of the component flow rates the mass balance constraints are composed of the product of the measurements of hydraulic flow rates and concentrations. This leads to a bilinear reconciliation problem. An approach for the solution of such problems is pointed out in the following.

Solution of the data reconciliation problem A solution method for linear and bilinear data reconciliation was described by Crowe *et al.* (1983) and Crowe (1986), respectively. The method also considers unmeasured variables. It makes use of matrix projection in order to split the reconciliation problem in different subproblems. Sanchez & Romagnoli (1996) made adaptations to this method by using a Q-R decomposition (Goodall, 1993) to find the projection matrices needed. The solution method of these authors was implemented in this thesis and is explained here only very briefly. Detailed explanations can be found in the publication of Sanchez & Romagnoli (1996).

The solution method for the bilinear data reconciliation problem is splitted into the following steps:

1. The bilinear constraints are rewritten. Therefore the component flow rates are classified into different categories as indicated in table 1.
2. Unmeasured variables are eliminated using Q-R orthogonal transformations.
3. The measured variables (measured hydraulic flows and component flows of category 1 & 2) are estimated together with the observable unmeasured hydraulic flow rates in an iterative procedure. For the iterative procedure, suitable initial conditions have to be chosen.
4. Estimation of the observable component flow rates of category 3.

Table 1: Categories of component flow rates depending on the corresponding hydraulic flow and concentration measurements.

Category	Q	c
1	measured	measured
2	unmeasured	measured
3	measured/unmeasured	unmeasured

2.3 Gross error detection

Narasimhan & Jordache (1999) state that "gross error detection is a companion technique to data reconciliation [...]". To improve the measurement accuracy those two techniques are applied together. The procedure to eliminate gross errors is therefore performed as follows (see also figure A1).

First the measured datasets are reconciled. Subsequently, the gross error detection itself is performed, which (for this thesis) consists of three steps: 1) Error detection, 2) Error identification and 3) Sensor isolation. Finally, a second data reconciliation is performed on the measurements which are not concerned by a gross error. In the following, approaches for the three steps of the gross error detection are explained. They are based on the assumption, that at only one of the measurements can be affected by a gross error simultaneously.

2.3.1 Error detection

The error detection step is performed in order to check if there's a gross error present in the measurements. In this thesis two types of tests were used: The global test and CUSUM charts. Both methods are based on hypothesis testing. This means that they make use of a test statistic which is compared to a certain threshold value. If the test statistic exceeds the threshold value, a gross error is said to be detected. Two types of errors can occur when hypothesis testing is used: If a gross error is detected although in reality the measurements are not affected by a gross error, this corresponds to a *Type I error* or *false alarm*. On the other hand, if an existing gross error is not detected, this is called a *Type II error*. The *power* of a test is equal to the percentage of gross errors detected with a given detection limit. An optimal test should therefore have maximal power along with a minimal false alarm rate.

In general it can be said, that the power of these test increases with the size of the gross error interfering the measurements.

Global test The global test is exploiting the spatial redundancy of the measurement. Its statistic γ is defined as in equation 2.4.

$$\gamma = \mathbf{r}^T \mathbf{V}^{-1} \mathbf{r} \tag{2.4}$$

$$\text{where } \mathbf{V} = \mathbf{A} \mathbf{\Psi}_x \mathbf{A}$$

$\mathbf{r}$: Balance residuals
$\mathbf{V}$: Covariance matrix of balance residuals
$\mathbf{A}$: Incidence matrix
$\mathbf{\Psi}$: Covariance matrix of the measurements

Under the null hypothesis H_0 no gross error is affecting the measurements and the global test statistic follows a χ^2-distribution with ν degrees of freedom. (The degree of freedom ν

corresponds to the rank of the incidence matrix A.) The test criterion for a certain confidence level α therefore equals $\chi^2_{1-\alpha,\nu}$. If the test statistic γ exceeds the test criterion, H_0 is rejected and a gross error is detected. After the explanations of Narasimhan & Jordache (1999) the probability of a false alarm of the global test is less than or equal to the confidence level α.

CUSUM charts Cumulative sum (CUSUM) charts are designed to detect either positive or negative shifts in the mean of any monitored process variable (Montgomery, 2009). Therefore, they are not only using the spatial redundancy of the measurements, but also their temporal redundancy by integrating the test statistics of the examined variables over time.

Spindler & Vanrolleghem (2012) applied *univariate CUSUM charts* in order to detect gross errors in mass balance residuals. For normalized data they defined the CUSUM statistic as in equation 2.5. This type of CUSUM chart is named two-sided since it uses one test statistic each for positive and negative deviations in the monitored variable. It signals a bias if its value exceeds a certain detection limit h_{univ} for the positive chart or $-h_{univ}$ for the negative chart, respectively.

$$\begin{aligned} C_t^+ &= \max(0, C_{t-1}^+ - k + x_t) \\ C_t^- &= \min(0, C_{t-1}^- + k + x_t) \end{aligned} \tag{2.5}$$

$$\text{with } C_0 = 0$$

C_t^+/C_t^-: Positive respectively negative CUSUM chart at time t
k: Reference value
x_t: Monitored variable (for monitoring the balance residuals r_t: $x_t = r_t/\sigma_r$)

In comparison to the univariate CUSUM charts, for multivariate processes as the WWTP considered in this thesis, *multivariate CUSUM charts* are said to perform better in gross error detection. From the most widely used multivariate CUSUM charts (MCUSUM and MC1) Mahmoud & Maravelakis (2013) recommended to use the multivariate CUSUM chart No. 1 (MC1) proposed by Pignatiello & Runger (1990), which is defined in equation 2.6. In contrast to the univariate charts, for this method the data are not normalized. The MC1 chart signals a bias, if its value increases above the detection limit h_{MC1}.

$$MC1_t = \max(0, \sqrt{\mathbf{Z}_t^T \, \mathbf{V}^{-1} \, \mathbf{Z}_t} - k l_t) \tag{2.6}$$

$$\text{with} \quad \mathbf{Z}_t = \textstyle\sum_{j=t-l_t+1}^{t} (\mathbf{X}_t - \boldsymbol{\mu}_0)$$

$$\text{and} \quad l_t = \begin{cases} l_{t-1} + 1 & \text{if } Y_{t-1} > 0 \\ 1 & \text{otherwise} \end{cases}$$

$\mathbf{V}$: Covariance matrix of balance residuals

k: Reference value

$\mathbf{X}_t - \boldsymbol{\mu}_0$: Difference between the vector of the monitored variable at time t and its mean. For monitoring the vector of balance residuals $\boldsymbol{r}_t$: $\mathbf{X}_t - \boldsymbol{\mu}_0 = \mathbf{r}_t$

Both types of CUSUM charts are calibrated by means of the calibration parameters h and k. These are selected such, that they perform optimal in matters of the average run length (ARL). The ARL is defined as the average time elapsed until the chart signals. Therefore, if no gross error affects the measurements, the unbiased run length (referred to as ARL_0) should be as high as possible. On the other hand, if a bias exists in the measurements, this error is desired to be detected as fast as possible. Thus, the corresponding run length (ARL_μ) is desired to be low.

In order to reach an optimal ARL performance, Lucas & Crosier (2000) recommended to use a value of $k = \frac{\delta}{2}$ as reference value. δ is referred to as non-centrality parameter and corresponds to the square root of the statistical distance resulting from a bias of a certain size in the process mean (equation 2.7). The CUSUM chart will therefore perform in an optimal way in matters of a bias corresponding to the one used for the calculation of δ. Nevertheless, it will also be able to detect biases of other sizes.

$$k = \frac{\delta}{2} = \frac{\sqrt{(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \mathbf{V}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)}}{2} \tag{2.7}$$

$\mathbf{V}$: Covariance matrix of balance residuals

$\boldsymbol{\mu} - \boldsymbol{\mu}_0$: Bias in the mean of the monitored variable for which the chart is designed for. For monitoring balance residuals: Balance residuals bias resulting from a certain measurement bias.

After the selection of the parameter k, the detection limit h is calibrated. In order to make different types of CUSUM charts comparable, usually this parameter is selected such that the same ARL_0 result for each chart.

2.3.2 Error identification

Once a gross error is detected by one of the methods described above, the next step is to identify the affected sensor. If a global test is used for the detection of gross errors, Narasimhan & Jordache (1999) recommended to use the serial elimination procedure for error identification. During the serial elimination procedure, in a first step each measurement is deleted in turn. Then, data reconciliation is performed again without the respective measurement and the global test statistic γ is recalculated. After the explanations of Narasimhan & Jordache (1999), it was illustrated by Ripps (1965) that the gross error can be identified in the measurement whose deletion results in the highest decrease of the global test statistic. Narasimhan & Jordache (1999) also described that for specific sensor layouts the same

reduction in the global test statistic can be obtained for the deletion of different sensors. Therefore, if one of these sensors is affected by a gross error, the serial elimination procedure is not able to identify the exact sensor. It is only capable of identifying a set of sensors from which one is affected by a gross error. This is referred to as *equivalent set of gross errors*.

2.3.3 Sensor isolation

If the gross error identification was successful, a sensor isolation can be performed. During this procedure the measurement which was identified to contain a gross error is deleted and a data reconciliation is conducted. The resulting dataset corresponds to a reconciled dataset free of the detected gross errors. A sensor isolation can only be executed if the serial elimination procedure results in the identification of a single sensor, but not if a set of measurements is obtained.

2.4 Solids Retention Time

The solids retention time (SRT) is one of the most important parameters for the design and operation of activated sludge plants. It corresponds to the average time during which the activated sludge stays in the system. It is calculated from hydraulic flows and TSS concentrations as indicated in equation 2.8.

$$SRT = \frac{V_{reactor} \cdot c_{TSS,reactor}}{Q_{eff} \cdot c_{TSS,eff} + Q_{ex} \cdot c_{TSS,ex}} \tag{2.8}$$

SRT: Solids retention time (d)
$V_{reactor}$: Volume of the activated sludge reactor (m^3)
$c_{TSS,eff}$: TSS concentration in the effluent (kg m^{-3})
$c_{TSS,ex}$: TSS concentration in the excess sludge (kg m^{-3})
Q_{eff}: Hydraulic flow of the effluent (m^3 d^{-1})
Q_{ex}: Hydraulic flow of the excess sludge (m^3 d^{-1})

3 Material and Methods

In the first part of this section an introduction to the Benchmark Simulation Model No. 1 – Long Term (BSM1_LT) is given. Afterwards it is illustrated how data reconciliation and different gross error detection methods are implemented for this model. Finally, it is explained which alternative sensor layouts were tested on the BSM1_LT.

3.1 Benchmark Simulation Model No. 1 – Long Term

To compare the different error detection methods the Benchmark Simulation Model No. 1 – Long Term (BSM1_LT) (Gernaey *et al.*, 2014) is used. This simulation platform provides a realistic model of a simple WWTP consisting of an activated sludge tank with five compartments and a secondary clarifier as depicted in figure 1 (top).

For this thesis, data of the hydraulic flows and the total suspended solids (TSS) concentrations from such a dynamic simulation performed by Villez *et al.* (2013a) were available. They include real hydraulic flows and concentrations of the BSM1 plant without any influence of measurement errors. The simulation incorporates a duration of 609 days with a sampling interval of 15 minutes. During the whole period the influent data are dynamic by means of hydraulic flows, concentrations and temperature. The solids retention time for dry weather conditions is controlled at around 9 days. Prior to the dynamic simulation, a simulation of 200 days with constant influent properties is run in order to reach steady state in the system. As recommended by Gernaey *et al.* (2014), the first 63 days of the dynamic simulation are not used since this initial phase is meant to allow the system to reach a dynamic pseudo steady-state. The remaining 546 days are splitted up in a calibration period (182 days) and a validation period (364 days).

As pointed out by Villez *et al.* (2013b), the BSM1 plant layout can be drawn differently in a more graph-theoretical way (figure 1, bottom). Therefore the model is represented by 8 nodes and 12 streams. (Plus two imaginary streams which will be explained later on.) Node H is an imaginary node, called the environmental node. Although the activated sludge reactor consists of five different compartments, it is modelled by one node (D). Only in this node reactions (growth or decay of TSS) are possible. Additionally, together with the settler (F), the reactor is the only node in the BSM1_LT which has a volume and therefore is able to store components – which, for this thesis, means TSS. For the hydraulic flows, obviously, no storage processes or reactions are possible.

Sensor model For each of the 12 streams one sensor measuring hydraulic flow and one measuring TSS concentrations is modelled. In order to simulate the sensor behaviour, random and gross errors are added to the real variables as described in the following. As recommended in Gernaey *et al.* (2014), normally distributed random numbers are used as random errors. They have a mean of zero and a standard deviation corresponding to 2.5 % of the measurement range of the sensor. It is assumed that the measurement range of a sensor is equal to a value slightly higher than the maximum value of the corresponding real state variable observed during the calibration and validation period. The exact values used can be found in table A1 in the appendix. The sensors are assumed to have no delays and no upper or lower bounds are implemented for the measured values.

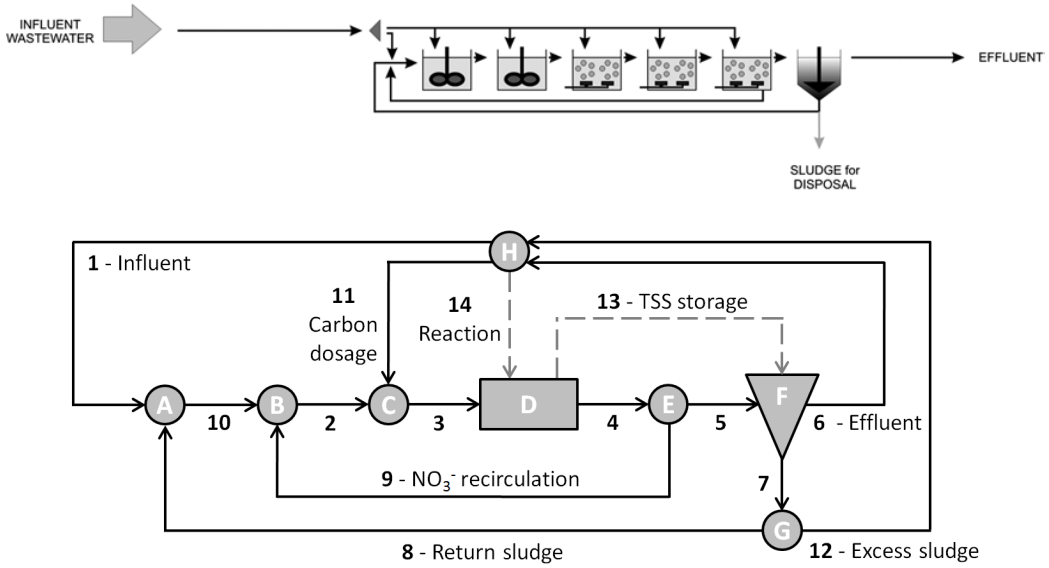


Figure 1: Plant layout of the Benchmark simulation model No. 1 after Gernaey *et al.* (2014) (top) and drawn in a graph-theoretical way (bottom).

In addition to the random errors, also gross errors of a certain magnitude can be added to the sensor signals. For this thesis only measurement biases are considered. It is assumed, that biases occur only for one sensor at a time.

3.2 Data reconciliation

The data series from the BSM1_LT simulation are reconciled by implementing the solution method for bilinear data reconciliation using Q-R decomposition (as described in chapter 2.2) in MATLAB. The mass balance constraints imposed include a balance over each of the real nodes (A-G) and over the overall plant, represented by the environmental node (H). For each node one hydraulic flow balance as well as one TSS balance is deployed, resulting in 16 balance equations. In a first analysis, all of the hydraulic flow and TSS concentration variables are assumed to be measured. Like this, the maximum reconciliation performance possible for the BSM1 plant is assessed. Since no calibration is necessary for data reconciliation, both, calibration and validation period are used for its assessment. The evaluation of other sensor layouts with only a part of the variables measured is described in chapter 3.4.

3.2.1 Modelling storage and reactions

The TSS storage processes and the TSS reaction mentioned above are violating the mass balance constraints. In order to deal with these processes, four different modifications of the data reconciliation model are tested.

- *Model A*: In this model, the errors caused by TSS storage and reaction are assumed to be insignificant and are therefore neglected. This corresponds to the model as

described above, with a mass balance over all nodes using the streams 1 to 12.

- *Model B*: As proposed by Crowe *et al.* (1983), the reaction is modelled by adding an imaginary component flow (flow 14) between the environmental node and the reactor. To account for the storage processes, the same approach using imaginary component flows is applied. Since the reactor and the settler are the only two nodes of the BSM1 plant with the ability to store components, all components which are stored in the reactor must be missing in the settler and vice versa. Thus, the mass balance errors caused by storage processes can be compensated by simply adding one imaginary component flow (flow 13) between the reactor and the settler. Flow 13 and 14 are depicted in figure 1 (bottom).
- *Model C*: In this model, only the reactions are considered and only flow 14 is added to the layout of model A.
- *Model D*: As in model C, only the reactions are considered. Instead of adding the component flow 14 as an unmeasured imaginary flow, it is modelled as a constant reaction. Therefore, it is implemented as a measured variable with a constant component flow rate of $1500\,\mathrm{kg_{TSS}\,d^{-1}}$. The measurement uncertainty is set to $10^{-12}\,\mathrm{kg_{TSS}\,d^{-1}}$.
- *Model E*: Only the storage processes are considered in this model. Thus, only flow 13 is added to the layout of model B.

3.2.2 Covariance matrices

For solving the data reconciliation problem and performing the gross error detection, the covariance matrix of the measurements is needed (see chapters 2.2 and 2.3). If it is assumed, that the random errors of the measurements are uncorrelated, the covariance matrix for the hydraulic flows can simply be defined by inserting the measurement variances of the sensors into the diagonal of the matrix. For the component flows, the variances are composed of the variances of the hydraulic flow sensor and the concentration sensor. Thus, the variances of the component flows are estimated with help of a linear Taylor expansion as shown in equation 3.1. In theory, the real hydraulic flows and concentrations should be used for this calculation. But – since only measured values are available – those are used as an approximation for the real values.

$$\boldsymbol{\Psi}_f = \boldsymbol{\sigma}_Q^2 \cdot (\tilde{\mathbf{c}} \circ \tilde{\mathbf{c}})^T + \boldsymbol{\sigma}_c^2 \cdot (\tilde{\mathbf{Q}} \circ \tilde{\mathbf{Q}})^T \tag{3.1}$$

$\boldsymbol{\Psi}_f$: Covariance matrix of the component flows
$\tilde{\mathbf{Q}}$: Vector of measured hydraulic flows
$\tilde{\mathbf{c}}$: Vector of measured concentrations
$\boldsymbol{\sigma}_Q^2, \boldsymbol{\sigma}_c^2$: Vector of the variances of the vectors Q and c
$\circ$: Hadamard product (elementwise multiplication)

3.2.3 Initial conditions

If some of the hydraulic flows are unmeasured but still observable, they are reconciled in an iterative procedure. Therefore estimates of the unmeasured hydraulic flows and of the covariance matrix are required as initial conditions. For this, the average hydraulic flows of the calibration and validation period are used. During each iteration step, a new value for the hydraulic flow and the covariance matrix is obtained and used for the following step. The iteration is stopped, as soon as convergence is achieved.

3.3 Gross error detection

As for data reconciliation, also for the gross error detection all variables are assumed to be measured in order to assess the highest possible gross error detection performance for the BSM1 plant. The global test and two types of CUSUM charts (see section 2.3.1) are tested on the different types of models as specified in the following.

3.3.1 Global test

Based on the data of the calibration period, for the models B and C the global test is calibrated on an average false alarm rate of 1 % by changing the confidence level α. Since such a low false alarm rate could not be reached for the models A and D using the confidence level α as calibration parameter, they are calibrated by changing the value of the test criterion $\chi^2_{1-\alpha,\nu}$ directly. Afterwards, a validation of the models is performed. This means, that it is checked if the false alarm rate obtained during calibration stays at the same level when using the calibrated model for the validation period. If for a model the change in the false alarm rate between the two periods exceeds $\pm 25\,\%$, it is said to fail the validation.

In order to assess the global test performance, positive biases between 0 and 50 % of the sensor measurement range are added to each of the sensor signals for the whole validation period. Afterwards, the corresponding test power is calculated.

3.3.2 CUSUM charts

The two types of CUSUM charts which are introduced in chapter 2.3.1 are implemented as follows. *Univariate CUSUM charts*: For each of the sixteen balance equations of the BSM1 model a positive and a negative CUSUM chart is created. Reference values k for each of the balance equations are calculated using equation 2.7 for a bias with a magnitude of half the measurement standard deviation in each of the sensors. For each balance therefore the same number of reference values results as streams are contributing to the balance. From those the highest reference value obtained for a balance is used for the corresponding positive and negative CUSUM chart. All the univariate charts are drawn together in one chart which is called a multiple univariate CUSUM chart. For convenience, in the following this type of chart will just be called "univariate CUSUM chart". By means of the data of the calibration period, the latter is calibrated to an average unbiased run length (ARL_0) of two years (730 days) by changing the detection limit h_{univ}. The calibration goal for the average unbiased run length is selected such, since this is expected to be a good choice in

order to gain the confidence of operators if these methods are used in practice. Too many false alarms would make the method unreliable.

As for the univariate chart, the *multivariate CUSUM chart* is also calibrated to perform optimally for a bias with a magnitude of half the measurement standard deviation. The detection limit h_{MC1} is calibrated as well to the same ARL_0. Afterwards, for both types of CUSUM charts a validation is performed. The validation is conducted similar to the one for the global test. However, instead of checking for a change in the false alarm rate, it is assessed if a change in the ARL_0 is obtained between calibration and validation. As a validation criterion as well a maximal change in the ARL_0 of $\pm 25\,\%$ is defined. If this value is exceeded, the models are said to fail the calibration.

With a length of only 182 days the calibration period is much shorter than the ARL_0 goal. Thus, it is actually not possible to calibrate the CUSUM charts to an ARL_0 goal of two years. This problem is solved by simply stringing together several calibration periods. Subsequently, the calibration is run over this string of periods. If, for example, a calibration run signals a false alarm on day 100 of the third calibration period of the string, this is reported as a run length of 464 days. It has to be mentioned that each time when a new calibration period is reached within the string (every 182 days), the CUSUM chart statistic is reset to zero. As a consequence, it is expected that the calibration overestimates the real ARL_0 to a certain degree. Also for the validation period the same procedure is used. The validation period has twice the length of the calibration period, which means that only half of the resets are performed. Thus, it is expected that the ARL_0 resulting for the validation is slightly shorter than the one of the calibration.

Due to the random errors to which the measurements are exposed, each run resulted in a different ARL_0. Additionally – as mentioned above – repeated runs are needed to account for the fact that the calibration and validation period are shorter than the ARL_0 goal. Thus, as proposed by Mahmoud & Maravelakis (2013) Monte Carlo simulations with several thousand runs are performed to calibrate and validate each model. Since this requires enormous computational efforts, the average run lengths could only be calibrated approximately to the desired value.

As for the Global Test, also for the CUSUM charts the bias size in each of the sensors is altered between 0 and 50 % of the sensor measurement range in order to calculate the average biased run length (ARL_μ). The point in time at which the bias is occurring is therefore altered randomly between the start of the validation period and 50 days before its end. For run lengths longer than this amount of time, it is possible that the point at which a bias would be detected exceeds the end of the validation period. Thus, only run lengths shorter than 50 days are reported correctly. All run lengths longer than this value are set to 50 days and are not reported in detail. To assess the average biased run length for a certain bias size in a sensor, six runs are performed each and their average was calculated.

3.3.3 Error identification

In this thesis the principle of serial elimination as described for the global test in chapter 2.3.2 is used analogously for the gross error identification by means of multivariate CUSUM charts. Each time the CUSUM chart signals a gross error a serial elimination is performed. The time frame during which the serial elimination is executed is defined as follows: It

ends at the point in time where the gross error detection method signaled a gross error. Its starting point is specified as the last point in time for which the CUSUM statistic equaled zero before the chart signaled a gross error. In turn, each measurement is deleted and the CUSUM statistic without the respective measurement is calculated for the serial elimination time frame. The measurement whose deletion results in the largest decrease of the CUSUM statistic at the endpoint of the serial elimination time frame is then identified as the measurement containing the gross error. If equivalent sets of gross errors are present, the procedure identified a set of measurements of which one is expected to contain the gross error.

The error identification performance is assessed for model B only. For this purpose alternately a bias of 10 % is added to each of the sensors and a gross error detection is performed using the multivariate CUSUM chart. Afterwards an error identification is performed as described above. For each sensor 20 such simulations are performed and the fractions of true and false positives are calculated.

3.3.4 Sensor isolation

In this thesis sensor isolation is implemented as described in chapter 2.3.3 but not investigated in detail. Though, an example of how gross error detection, error identification and sensor isolation could work together to improve solids retention time control is given: For model B a bias of 10 % is added to the concentration measurement of the outflow of the activated sludge reactor at day 70 of the validation period. Under the premise that gross error detection and sensor isolation were successful in identifying the sensor containing the bias, a sensor isolation is performed. The solids retention time is then calculated after equation 2.8 in three different ways: Based on the real system variables, on the biased measurements and on the reconciled estimates of the measurements after performing the gross error detection procedure including a sensor isolation. The resulting SRTs are smoothed by a moving average of $\pm$ 3.5 days.

3.4 Alternative sensor layouts

If only some of the hydraulic flows and concentrations of the BSM1 plant are measured, the performance of the bilinear data reconciliation and the gross error detection methods is expected to decrease. To assess this effect, a reduced sensor layout is developed for the models A and B. Thereby, the constraint is imposed that the reduced layouts must (at least theoretically) be able to detect errors in the variables used for the calculation of the solids retention time (see chapter 2.4). This means that the variables Q6, Q12, C4, C6 and C12 need to be redundant.

For a plant like the BSM1_LT some of the variables are expected to be measured usually: The hydraulic flow of the influent (Q1) and the effluent (Q6) as well as the concentrations of the flow leaving the activated sludge reactor (C4), the effluent (C6) and the excess sludge (C12). Furthermore it is assumed, that the hydraulic flow rates of the return sludge (Q8), the NO_3^--recirculation (Q9) and the excess sludge (Q12) are either measured as well or can be derived from the performance of the pumps installed. Additionally, for the carbon dosage the corresponding hydraulic flow (Q11) and the concentrations (C11) are also known.

In the reduced sensor layout of model A the above-mentioned flows and concentrations are implemented as measured variables. Additionally also the influent concentration (C1) is defined as measured. The resulting sensor layout of model A is indicated with circles in table 2. Since for this layout all measured variables are also redundant, gross errors in variables used for the SRT calculation can be detected theoretically. Since model B includes two imaginary component flows, the layout used for model A is not sufficient to obtain redundant measurements for all desired variables. Thus, additional measurements are added to the reduced sensor layout of model B. These are indicated in table 2 by means of plus signs.

For these two models data reconciliation is performed the same way as for the models with all variables measured (see chapter 3.2). Additionally, univariate and multivariate CUSUM charts are calibrated and validated as described in chapter 3.3.2. Subsequently, the performance of the charts is assessed.

Table 2: Measured variables for the reduced sensor layouts of models A (circles only) and B (circles and plus signs).

Sensor	1	2	3	4	5	6	7	8	9	10	11	12
Hydraulic flows	o			+	+	o	+	o	o		o	o
Concentrations	o			o	+	o	+	+	+		o	o

4 Results

In this section the results of this thesis are shown. First, the dataset which was produced by the BSM1_LT is presented. Subsequently, it is shown how the bilinear data reconciliation and the different gross error detection methods performed for this dataset. Finally, the results of the error identification procedure and an example of sensor isolation are presented.

4.1 Benchmark Simulation Model No. 1 – Long Term

The mean hydraulic flows, TSS concentrations and TSS component flows for all stream of the BSM1_LT simulation are listed in table 3. On an average day 20'600 m^3 of wastewater is treated by the BSM1 plant. The streams with the lowest TSS concentrations are the influent ($200\,\mathrm{g\,m^{-3}}$) and the effluent ($14.5\,\mathrm{g\,m^{-3}}$). With $300\,\mathrm{m^3\,d^{-1}}$, the lowest hydraulic flow is observed in the excess sludge. During the whole simulation period no carbon was dosed to the inflow of the reactor. Thus, the flows in stream 11 will not be considered in the following. The fluctuations of the influent characteristics over the simulation period are shown in figure 2 together with the different analysis periods.

More TSS is entering the system through the influent than what is lost through the effluent and the excess sludge. Thus, a net decay of TSS is prevailing in the activated sludge reactor which equals approximately 60 % of the TSS inflow rate. Figure 3 shows the TSS balance over the activated sludge reactor for six days of the calibration period, including inflow and outflow rates as well as the decay rate (indicated as reaction rate). Additionally, also the TSS storage is shown, with a positive value corresponding to TSS detention in the reactor. In contrast to the reaction rate, the storage rate integrated over a longer timespan is approximately equal to zero.

4.2 Data reconciliation

In order to assess the performance of the bilinear data reconciliation for the different models the focus is set on the measurement errors before the reconciliation and on the remaining

Table 3: Mean hydraulic flows, TSS concentrations and TSS flows for the whole simulation period of the BSM1_LT plant.

Stream		1	2	3	4	5	6
Q	$(m^3\,d^{-1})$	20'594	88'591	88'591	88'591	39'040	20'294
c_{TSS}	$(g\,m^{-3})$	200	3'824	3'824	3'806	3'806	14.5
f_{TSS}	$(t\,d^{-1})$	4.1	333	333	332	147	0.32
Stream		7	8	9	10	11	12
Q	$(m^3\,d^{-1})$	18'746	18'446	49'550	39'040	0.0	300
c_{TSS}	$(g\,m^{-3})$	7'807	7'807	3'806	3'881	0.0	7'807
f_{TSS}	$(t\,d^{-1})$	146	144	185	148	0.0	2.3

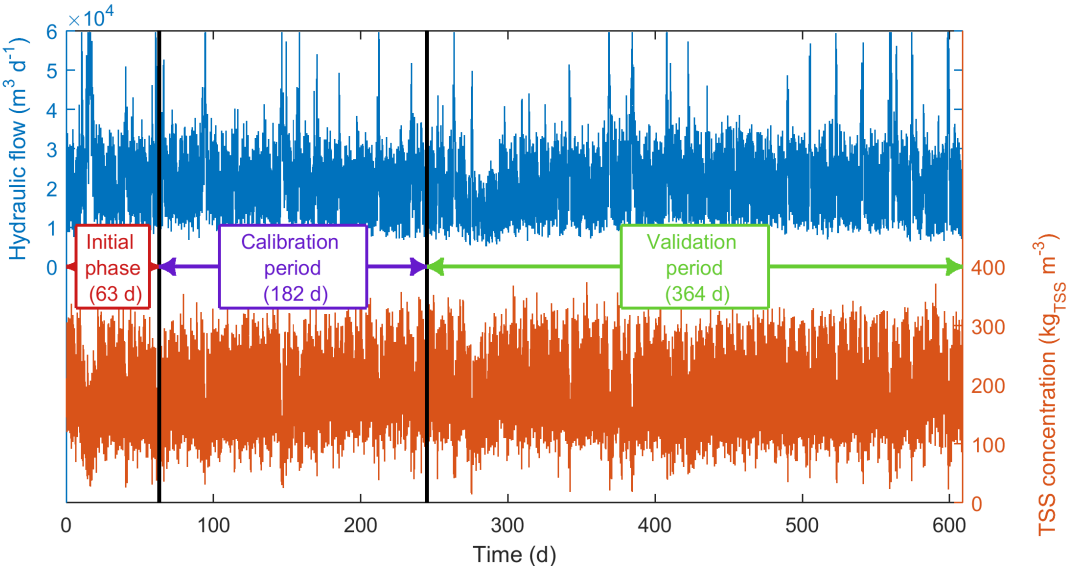


Figure 2: Influent characteristics of the BSM1_LT simulation for the different periods.

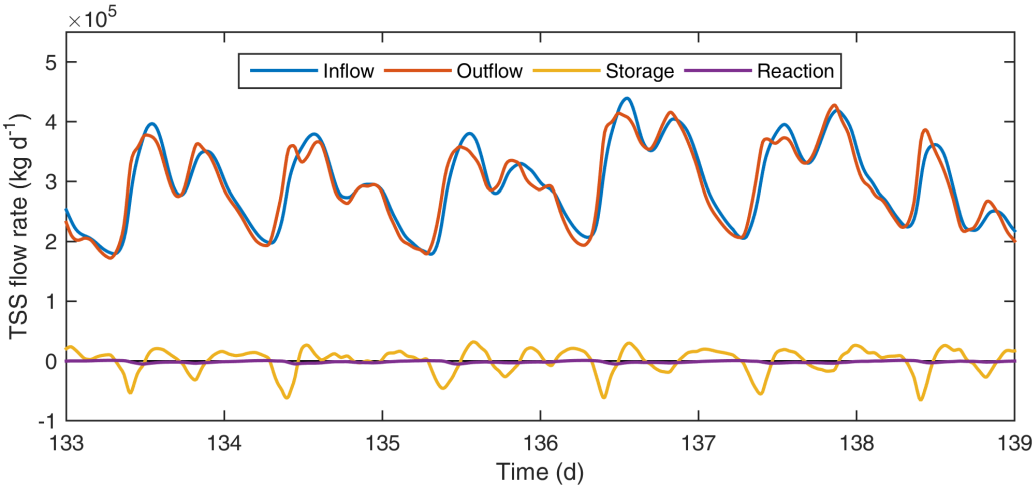


Figure 3: TSS balance over the activated sludge reactor.

errors after it. Provided that gross errors are absent, the means of both, the measurement errors and (in case of a successful data reconciliation) the errors of the reconciled estimates are approximately zero. The standard deviation on the other hand should decrease when reconciling the data.

For all of the models A–E, the relative means of the errors in the hydraulic flows are smaller than 10^{-4}. This is also true for the relative means of the errors in the measured concentrations and component flows. However, for the reconciled estimates of the concentrations and the component flows deviations from the mean are observed in all models. While these are small ($< 1\,\%$) for model B and C, the models A, D and E show strong deviations in the means of some concentration measurements. The highest deviations are observed in the reconciled estimates of the concentrations and component flows of stream 1 (up to around $30\,\%$) and stream 12 (up to $4\,\%$). The relative error means and standard deviations before and after data reconciliation for all of the models can be found in tables B2 to B10 in the appendix.

In figure 4 the proportion between the standard error of the reconciled estimates and the measurements for the models A (left) and B (right) are shown. The corresponding figures for models C–E as well as tables with the values of the relative standard deviations can be found in the appendix B.1. For all models a reduction of the standard errors in all of the hydraulic flows is reached by data reconciliation – except for flow 12, where the standard errors stay the same. When looking at the component flows, only in model B the standard errors of all streams are either decreasing or staying constant. For the other models some of the standard errors in the component flows increase strongly due to data reconciliation. Mostly, the streams 1, 5, 6 and 12 are affected, with the highest increase occurring in the streams 1 (up to a factor of 4.0) and 12 (up to a factor of 2.1). For the concentration measurements all models show an increase in the standard deviations of certain measurements. For the concentrations in flow 9, the proportion between the standard errors amount up to a factor of 32. For model B the smallest number of streams with an increased standard deviations is obtained.

The high standard errors in the reconciled estimates of the concentrations in stream 9 are found to result from the phase between day 270 and 310 of the dataset. During this timespan the whole plant is in a low flow phase and the hydraulic flows of stream 9 are zero or close to zero. The concentrations in stream 9 during this phase show short-term peaks with huge magnitudes in the positive and negative direction. Some of the obtained concentration estimates even are below zero.

When aggregating the reconciliation performance of the different models for hydraulic flows, concentrations and component flows, it can be said, that model B is the one showing the best results followed by models C, D and A.

4.3 Gross error detection

4.3.1 Global test

As mentioned in chapter 3.3.1 the global test is calibrated to a false alarm rate of $1\,\%$. Thus, after the explanations of Narasimhan & Jordache (1999) the confidence levels α are expected to be greater than or equal to a value of $1\,\%$. However, they are found to be

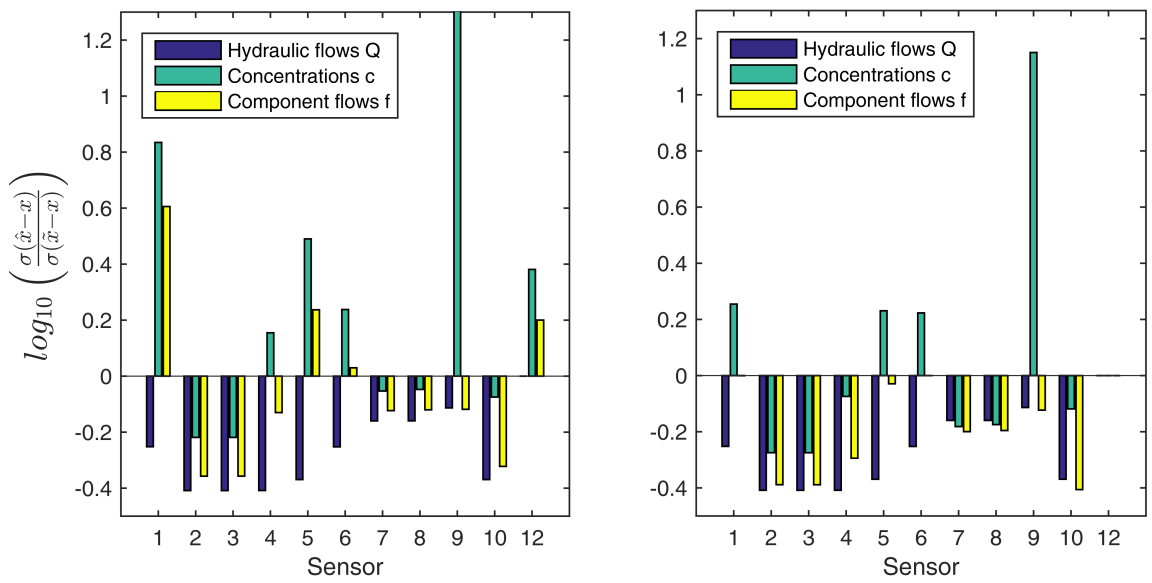


Figure 4: Proportion between the standard error of the reconciled estimates and the measurements for the models A (left) and B (right).

smaller than this value (e.g. 0.2 % for model B or $6 \cdot 10^{-6}$ for model C). For the models A and D they even are smaller than 10^{-16}. This means, that they fall below the accuracy limit of the MATLAB function "chi2inv" which was used to calculate the test criterion $\chi^2_{1-\alpha,\nu}$. Thus, the calibration of the global tests of these two models is found to be only possible by calibrating the test criterion directly.

The false alarm rates for the global test obtained from calibration and validation are shown in table 4 for models A to D. Based on the validation criterion (see chapter 3.3.1) all of the four models are successfully calibrated to a 1 % false alarm rate. While for models B and C the false alarm rate remain on the same level during the validation, a small decrease in the false alarm rate is observed for model A. For model D, the rate increases slightly.

Figure 5 shows the power of the global test applied on model B, depending on the relative size of the bias in each of the different sensors. The results for the other models can be found in chapter B.2 in the appendix. As expected, the power of the test increases with bias size.

It can be said in general, that the relative bias required to reach a certain test power is smaller for biases in the hydraulic flow measurements than in the concentration measurements. Furthermore, Model B results in the highest global test power for all measurements, except for the measurement C5. For this sensor the highest detection power was found using model C. Biases in certain measurements (e.g. Q2, Q3, Q4, C7 and C8) generally result in a relatively high test powers and are therefore easy to detect. For other sensors (e.g. Q12, C1 and C12) even high relative biases are not leading to a significant increase in the test power. Gross errors in these measurements are almost undetectable. Measurement C6 is not redundant for this model and therefore shows no increase in test power. Nevertheless, for all of the measurements – even for the measurements with a high detectability – quite large biases are needed to reach a suitable test power. For example, to reach a 95 % detection

power for the measurement Q2 using model B, a bias of 15 % is needed.

Table 4: False alarm rates for the global test calibration and validation. The uncertainty indicated corresponds to one standard deviation.

Model	A	B	C	D
Calibration	$1.02 \pm 0.018\%$	$0.96 \pm 0.036\%$	$1.01 \pm 0.026\%$	$0.98 \pm 0.015\%$
Validation	$0.90 \pm 0.005\%$	$0.95 \pm 0.007\%$	$1.01 \pm 0.012\%$	$1.21 \pm 0.017\%$

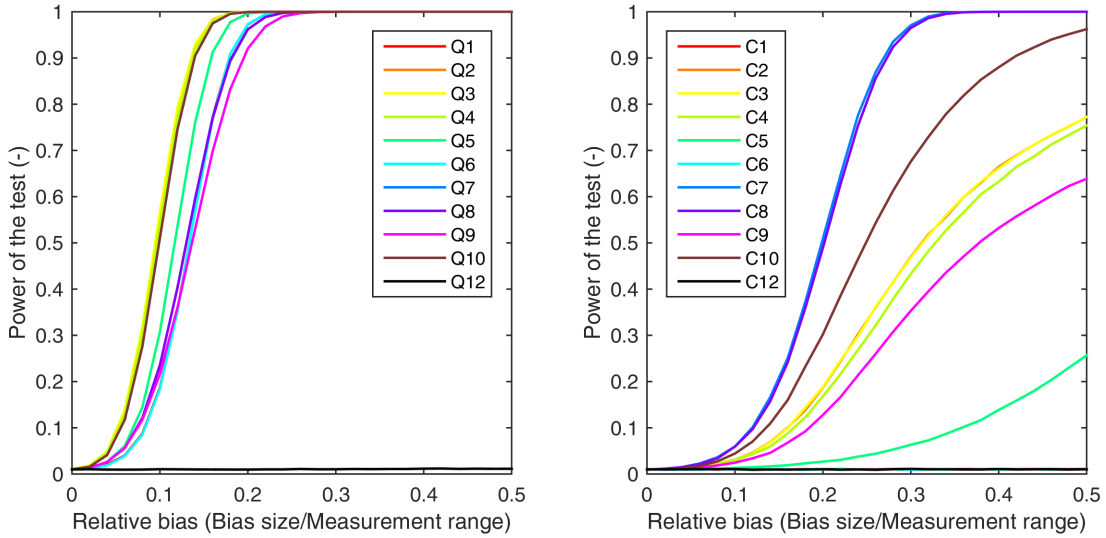


Figure 5: Power of the global test for model B depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

4.3.2 CUSUM charts

In table 5 the average unbiased run lengths (ARL_0) which are actually obtained for the calibration of the univariate and multivariate CUSUM charts are shown for the different models. Below, the average run lengths resulting for the validation period are listed. For most of the models, the ARL_0 of the validation period are significantly lower than the value they were calibrated for, which means that they fail the validation. Exceptions are the univariate and multivariate CUSUM charts of model B as well as the univariate CUSUM chart of model D. For those three only a small decrease (less than 20 %) in the ARL_0 is observed and their validation is accepted.

In order to identify the reasons for the strong reduction in the ARL_0 which resulted for some models, the different CUSUM charts are inspected visually. (Examples of the behaviour of the CUSUM charts of all models for the calibration and the validation period can be found in figures B10 to B19 in the appendix.)

Table 5: Average unbiased run lengths (ARL_0) in days for the univariate and multivariate CUSUM charts for the different models during calibration and validation. The uncertainty indicated corresponds to one standard deviation.

CUSUM chart	Period	ARL_0 (days)				
		A	B	C	D	E
Univariate	Calibration	715 ± 41	784 ± 71	764 ± 39	698 ± 26	722 ± 34
	Validation	93.8 ± 3.6	656 ± 11	199 ± 4.3	590 ± 25	106 ± 1.1
Multivariate	Calibration	727 ± 36	773 ± 87	718 ± 62	707 ± 34	862 ± 7.5
	Validation	227 ± 0.6	693 ± 8.0	355 ± 5.9	26.3 ± 0.1	217 ± 0.8

For the models which didn't result in a significant decrease of the ARL_0, the CUSUM charts show a random behaviour during the calibration and the validation period. This means, that for each Monte Carlo simulation peaks occur at different points in time. In contrast to that, the models which failed the validation show a systematic behaviour. Their peaks are found mainly at the same few points in time. Therefore, also most of the false alarms are signaled in a systematic way. For the all of the univariate models which failed the validation, the same scheme responsible for the decrease in the ARL_0 is detected: The magnitude of peaks increases when changing from the calibration to the validation period. Therefore, more false alarms are produced and a shorter average unbiased run length results. It has to be mentioned that a systematic behaviour is also found for the CUSUM statistics of some single balance nodes of the univariate CUSUM chart of model D (which passed the validation). However, this only leads to small peaks which aren't triggering a false alarm.

Referring to the multivariate CUSUM charts of the models which failed the validation, shorter average unbiased run lengths occur due to different reasons: For model C, as for the univariate charts, the magnitude of peaks increases. On the other hand, model D systematically signals an error on day 26 of the validation period which is the start of the low flow period of the BSM1 plant. The CUSUM statistic of models A and E show peaks of a high frequency which occur in a systematic way. In addition, over longer time scales a systematic increase of the CUSUM statistic is observed. For each validation run this trend is interrupted by the end of the low flow period on day 49 of the validation period (day 294 of the whole dataset). For these two charts, the average unbiased run length therefore strongly depends on the length of the analysed period. Thus, the decrease in the ARL_0 between the calibration and the validation period is mainly caused by the fact that the validation period has twice the length of the calibration period.

In the following the focus is only set on the models which show a constant ARL_0 behaviour for the calibration and validation period. The average biased run lengths (ARL_μ) for the univariate CUSUM chart of model B are shown in figure 6. (The same graphs for the multivariate CUSUM chart of model B and the univariate chart of model D can be found in figures B6 & B7 in the appendix. Furthermore, figure B9 shows two example charts for model B if one sensor is affected by a bias.) The standard deviations of the ARL_μ lie between 10 % and 90 % of the denoted value, depending on the sensor and on the bias size. For small biases, the relative errors are found to be higher than for large biases. For all

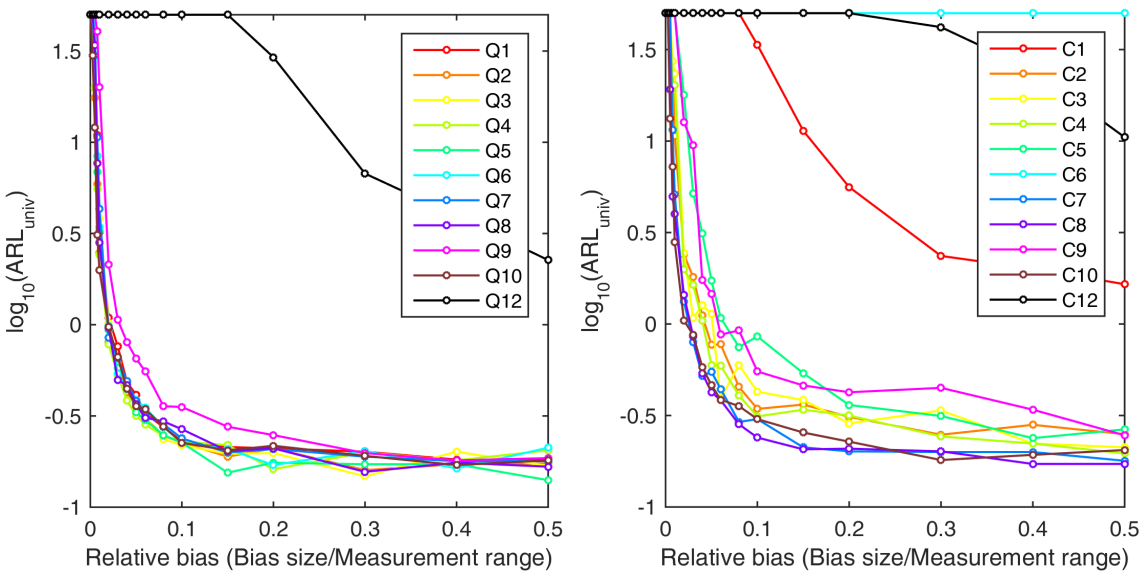


Figure 6: Average biased run lengths (ARL_μ) of the univariate CUSUM chart for model B depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

sensors the ARL_μ is decreasing with increasing size of the relative bias. Thus, as expected, the detectability of biases increases with their size. Nevertheless, due to the mentioned uncertainties some single points in the figure are violating this rule (e.g. in measurements C3, C5, C7 or C9). As for the global test, biases in the sensors Q12, C1 and C12 are much more difficult to detect than biases in other sensors. Furthermore, since measurement C6 is not redundant for this model, the CUSUM chart does not respond at all to the addition of a bias in this measurement.

To compare the performance of the univariate and the multivariate CUSUM charts for model B, in figure 7 the ratio of the average biased run lengths of the two chart types depending on the relative bias size is shown. For biases of more than 10 % in the hydraulic flow measurements the multivariate chart exhibits much shorter run lengths than the univariate one. (Except for flow 12, where no general trend is observed.) For biases smaller than 10 %, some of the biases in the hydraulic flow measurements are detected faster with the univariate charts, others with the multivariate ones. If the concentration measurements are affected by biases of more than 10 %, the multivariate CUSUM chart is performing slightly better for most of the sensors. Exceptions are the sensors C1, C5, C6 and C12. Concentration biases smaller than 2 % result in much shorter run lengths for the univariate chart (except for measurement C9).

A comparison of the run lengths of the univariate CUSUM charts of model B and D is shown in figure B8 in the appendix. For biases higher than 5 % model D generally shows either equal or shorter average run lengths compared to model B. The only exception is the concentration measurement C1, where model B has shorter run lengths.

When comparing all three CUSUM charts, the multivariate chart of model B is found to perform best for biases higher than 10 % . For smaller biases the picture is not so clear,

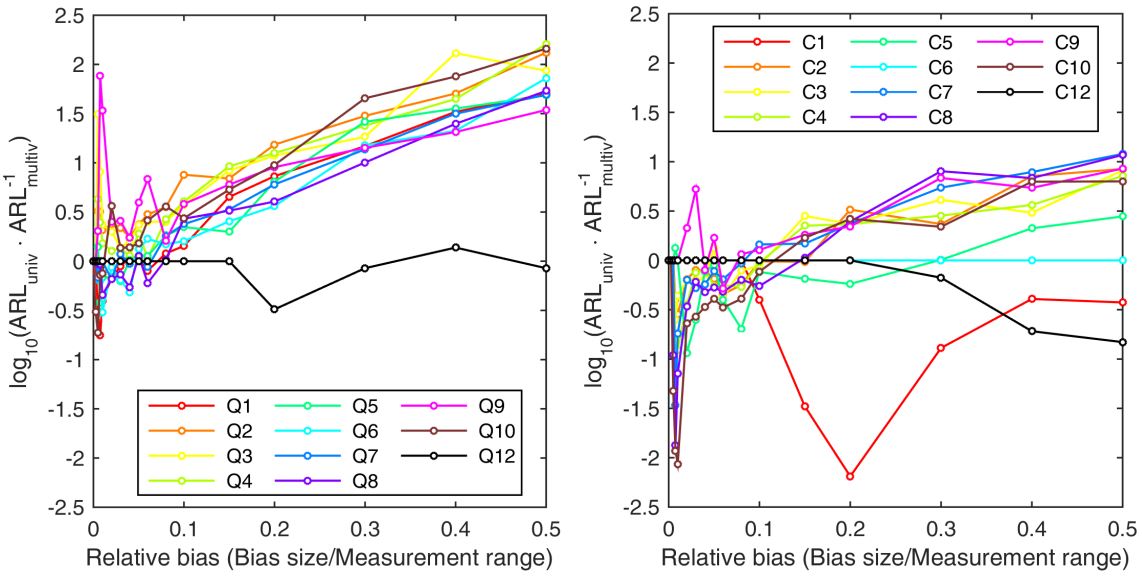


Figure 7: Ratio of the average biased run lengths (ARL_μ) of the univariate and the multivariate CUSUM chart for model B depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

except for biases in the concentration measurements which are smaller than 2 %. There, the univariate chart of model B produces the best results.

Comparison to global test As mentioned in chapter 4.3.1, for model B the detection power of the global test is 95 % for a bias of 15 % in measurement Q2. The detection power of the CUSUM charts strongly depends on the simulation duration. As it can be seen in figure 6, on average it takes around 0.2 days for the univariate CUSUM chart to detect a bias of 15 % in measurement Q2. Thus, if a simulation continues for 4 days after a bias occurred, a 95 % detection power is reached. If the simulation continues for more than this amount of time, a higher detection power results. The false alarm rate for the global test is around 1 %. For a 15 min sampling interval – as it is used here – this corresponds on average to one false alarm every 25 hours. In contrast to that the CUSUM charts are calibrated to only produce a false alarm every two years. Similar results are obtained for other sensors or models, as well as for the multivariate CUSUM chart. Thus, in general the CUSUM charts are found to perform better than the global test.

4.3.3 Error identification

The fractions of true and false positive identifications of the error identification procedure for model B are shown in table B12 in the appendix. A true positive identification is indicated if the error identification method found either the correct measurement or – if equivalent sets of gross errors are present – the correct set of measurements. Generally, if the CUSUM chart of a measurement shows a short average biased run length, the fraction of correct identifications is high. For these measurements (Q1-Q10, C2-C5 & C7-C10) the percentage

of true positive identifications ranged between 50 % and 100 %. The measurements Q11, Q12, C1, C11 and C12 on the other hand produce false positive identifications only. Since measurement C6 is not redundant, no gross error detection and as a consequence also no error identification is possible for this sensor.

4.3.4 Sensor isolation

The solids retention times which are calculated using the real system variables, the measurements and the estimates after sensor isolation are shown in figure 8. While the measured SRT show only small deviations from the real SRT during the time before the bias occurred, for the following timespan a significant deviation of around 1.2 days is observed. In contrast to that, the SRT calculated on the basis of the estimates resulting after sensor isolation only exhibits small deviations for the whole period.

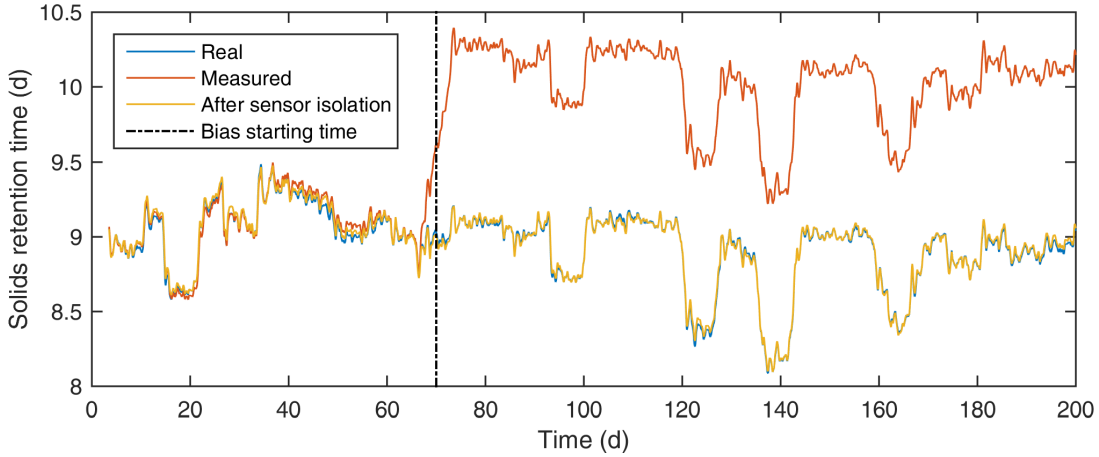


Figure 8: Comparison between the real solids retention time and the one resulting from measurements and after sensor isolation using a moving average of 7 days.

4.4 Alternative sensor layouts

4.4.1 Data reconciliation

As for the models with all sensors measured (original models) also for the alternative sensor layouts of model A and B the data reconciliation performance is assessed by calculating the relative means and standard deviations of the measurement errors and the reconciled datasets. These values can be found in tables B13 to B16 in the appendix. In comparison to the original models, less variables are observable and less measurements are redundant for the models with a reduced sensor layout. The means of the errors of both, measurements and reconciled estimates only show small deviations from the values which are obtained for the original models.

The proportions of the standard deviations of the errors in the reconciled estimates and the measurements of the models with reduced sensor layouts are shown in figure B20 in the appendix. For the reduced layout of model A, data reconciliation only leads to a reduction

of the standard errors of the hydraulic flows. However, all concentration and component flow measurements of this model show an increase in their standard errors due to data reconciliation. On the other hand, for model B the accuracy of all hydraulic and component flow measurements is increasing when a data reconciliation is performed.

Generally, the data reconciliation performance of the models with a reduced sensor layout is found to be inferior to the one of the original models.

4.4.2 CUSUM charts

In table 6 the average unbiased run lengths of the CUSUM charts obtained for the calibration and the validation period are shown for the two models with a reduced number of sensors. The results look similar to the ones obtained for the original models: For both types of CUSUM charts model A shows a strong reduction in the ARL_0 when the calibrated parameters are tested on the validation period. It therefore fails the validation. In contrast to that, with run lengths staying approximately constant, model B passes the validation. Thus, only this model is evaluated further.

For most of the measurements of the reduced sensor layout of model B both types of CUSUM charts are performing similar to the ones of the original models (see figures B21 and B22 in the appendix). Exceptions are the measurements Q1, Q6, C1 and C6 which are showing no or almost no decrease in their average biased run length. For the concentration measurements this is obvious, since both of them are not redundant for the reduced sensor layout of model B. However, for the hydraulic flow measurements this was not expected.

Table 6: Average unbiased run lengths (ARL_0) in days for the univariate and multivariate CUSUM charts for model A and B with a reduced number of sensors during calibration and validation. The uncertainty indicated corresponds to one standard deviation.

CUSUM chart	Period	ARL_0 (days)	
		A	B
Univariate	Calibration	664 ± 21	718 ± 48
	Validation	198 ± 0.0	694 ± 41
Multivariate	Calibration	742 ± 77	733 ± 52
	Validation	214 ± 1.2	723 ± 36

5 Discussion

In this section the results presented in the previous chapter as well as the assumptions and the methods used for this thesis are reflected critically. Additionally, recommendations for future studies are given.

5.1 Data reconciliation

The results of the data reconciliation show that the implemented bilinear data reconciliation method is capable of improving the accuracy of the hydraulic flow measurements by diminishing the effect of random errors. The standard deviations of the measurement errors are reduced while their means are kept approximately equal to zero.

While the hydraulic flow is a conservative quantity which cannot be stored, TSS is reactive and can be stored in the reactor and in the settler of the BSM1_LT plant. Because of that, the results obtained for the component flows are more complex. Model B, which is the only model considering TSS reaction as well as TSS storage processes in its layout, clearly shows a much better data reconciliation performance than the other models. Using this model, as already observed for the hydraulic flows, the standard deviations of the errors in the component flows are reduced or remain constant. Meanwhile, the means of the errors are not showing large deviations from zero. From the models which are not considering both of the processes, the one accounting for the reactions by means of an imaginary flow (model C) is also showing no major change in the error means when reconciling the data. However, for some sensors an increase in the standard error of the component flow is observed. All other models show a significant change the in mean values of at least some of the component flow variables and are therefore not successful in reconciling the data. This shows that it is necessary to account for TSS storage and TSS reactions to increase the measurement accuracy by data reconciliation. The implementation of imaginary TSS flows is observed to be a suitable method to do that.

The standard errors of some concentration measurements are found to increase due to data reconciliation even though the standard deviations of their corresponding hydraulic and component flows decreased. Accordingly, the bilinear data reconciliation is not successful for these measurements. This is due to the fact that – as an intensive quantity – the concentrations cannot be reconciled directly but are rather a "by-product" of the reconciliation of the two extensive quantities hydraulic and component flow. For example, the increase in the standard errors of the concentrations in stream 9 are found to be induced by a period during which the hydraulic flows are close to zero. The reconciled estimates of the concentrations are calculated from the quotient of the reconciled component flows and the reconciled hydraulic flows. Thus, if the hydraulic flows are close to zero, small changes in the latter can lead to huge deviations in the estimated concentrations and therefore result in an increase of their standard errors.

Since the peaks obtained for this phase are reaching enormous magnitudes, these errors could easily be avoided by using feasibility constraints to set upper or lower bounds for the respective variables (e.g. $c \geq 0$). However, bilinear reconciliation is not capable of handling inequality constraints (Narasimhan & Jordache, 1999). If one wants to implement such constraints, nonlinear reconciliation techniques would be the method of choice.

For some hydraulic and component flows the application of data reconciliation almost shows no effect at all, even though these flows are redundant (e.g. both types of flows in stream 12). These flows are only contributing by a minor amount to the overall balance of their respective nodes. Their standard deviations are much smaller than the ones of the other flows contributing to the balances of these nodes. Hence, the balances are closed by adjusting the variables with large standard deviations. The flows with small standard deviation are (almost) not adjusted at all. As a consequence, the mentioned measurements in stream 12 are redundant indeed but behave as practically nonredundant measurements (Narasimhan & Jordache, 1999). These practically nonredundant measurements are an inherent property of the treatment plant layout and can therefore not be avoided easily.

5.2 Gross error detection

5.2.1 Global test

The detection power of the global test is found to depend on the assumptions made on TSS storage and reactions. As for the data reconciliation, the models neglecting the effect of one or both of these processes (models A, C and D) show a worse performance than the one which is accounting for both of them (model B). For biases in the hydraulic flow measurements between 15 and 20 % the latter model shows a test power of 95 %, which is rated as a moderate performance. Biases of the same range in the concentration measurements lead to detection powers of 5 to 50 % only, which is a poor outcome. Biases in the measurements which were described as practically nonredundant in the previous chapter show no detection power at all even for large biases.

An unexpected finding results from the calibration of the global test. Although the confidence levels α are expected to be greater than or equal to the false alarm rate, they are found to be many orders of magnitude smaller for some of the models. For model B, however, the confidence levels were found to be only little smaller. The huge deviations found for the other models are expected to result because these models are neglecting the influence of TSS storage and reaction. Since these are non-random processes, evidently the balance residuals of the respective models are not normally distributed. From this it follows that the global test statistic is not following a χ^2-distribution. Thus, the null hypothesis is rejected much more frequently and a much larger false alarm rate is produced by the global test.

Nevertheless, the above conclusions are not able to explain that also for model B the confidence levels were smaller than the false alarm rate. A possible reason for this can be found in the way how the measurement covariance matrices of the component flows were estimated (see chapter 3.2.2). The estimation is subject to the following error sources: Firstly, the estimates are obtained from measured values. Thus, random and gross errors can affect the accuracy of the estimated covariance matrix. Secondly, a linear Taylor expansion is used to approximate the covariance matrices of the component flows. If this approximation is insufficient, deviations in the resulting matrices are expected. In further studies the approximation could be verified by using the real values instead of the measured ones (these are easily available in the BSM1_LT). Additionally, a more accurate equation could be used for the approximation of the covariance matrix (Goodman, 1960).

5.2.2 CUSUM charts

An important finding from the CUSUM chart validation is that most of the implemented charts exhibited a strong decrease in the average unbiased run length between the calibration period and the validation period. Because of this they failed the validation. This subject is discussed in the following. First, the focus is set on those charts which passed the validation.

In contrast to the global test both, univariate and multivariate CUSUM charts reveal a significantly higher gross error detection performance. For example, most of the hydraulic flow and concentration measurements biases of 5 % can be detected on average within one day or less. Furthermore, Villez *et al.* (2013a) evaluated a gross error detection method based on weighted sum of squared residuals (WSSR) in combination with nonlinear data reconciliation on the BSM1_LT plant. They found that this method (which is like the global test only using the spatial redundancy) exhibited a low performance. These results clearly show that the gross error detection performance can be increased significantly by exploiting the temporal redundancy in addition to the spatial redundancy. Therefore, in future studies also other methods which make use of the temporal redundancy should be tested, for example exponentially weighted moving average (EWMA) control charts (Montgomery, 2009).

The comparison of the univariate and the multivariate CUSUM charts shows that the multivariate chart is generally performing better for large biases. The univariate chart on the other hand results in shorter average run lengths for small biases in the concentrations measurements.

For the validation period, without exception, all models show a decrease in the average unbiased run length (ARL_0) in comparison to the calibration period. A certain decrease in the ARL_0 is expected due to the effect explained in chapter 3.3.2. It is assumed that this effect is responsible for the small ARL_0 decrease in the models which passed the validation. In order to avoid this effect for future studies, it is recommended to calibrate the CUSUM charts to an ARL_0 which is shorter than the length of the calibration and the validation period. As a positive side effect this would lead to a lower computational effort for the calibration and the validation of the CUSUM charts.

The models which failed the validation systematically showed peaks at the same points in time of the calibration and the validation period. This shows, that the false alarms of these charts are not triggered by random errors. It is reasonable that these charts are in fact responding on the balance errors caused by the neglect of reaction and storage processes. Furthermore, some of the charts are very sensitive to the operational state of the treatment plant. For example, they are clearly responding on the low flow phase which was occurring at the beginning of the validation period. Since gross error detection methods preferably should perform equal for all operational states, the usage of such charts should be avoided. However, in the univariate CUSUM chart of model D it is not accounted for the storage processes and the reactions are only modelled by a constant value, this model still passes the validation test. A visual inspection of this CUSUM chart reveals a systematic behaviour for the CUSUM statistics of some of the considered balances. But – in contrast to the other models – this behaviour is not leading to peaks which are large enough to trigger a false alarm. Furthermore, this chart is even reveals a slightly better gross error detection than the univariate CUSUM chart of model B. Hitherto, the reasons for this could not be

distinguished, but it is expected that further investigations on this would be worthwhile.

The multivariate CUSUM charts of the models A and E exhibit a positive trend in their statistic over time. Thus, for these models the decrease of the ARL_0 between the calibration and the validation period is mainly caused by this trend (see explanations in chapter 4.3.2). It is assumed that this trend comes from the residuals which are induced by the neglect of the TSS reactions. However, if a larger value for the calibration parameter k is selected, this would most likely eliminate this trend. Therefore, further analyses should be carried out to check if a larger value for k would make these charts pass the validation.

5.2.3 Error identification

The results of the error identification show that the serial elimination strategy is on average capable of identifying the correct sensor in 70 % of the cases if a bias of 10 % is added to the measurents. Although biases in some sensors are identified with true positive rates of up to 95 %, biases in other sensors produced false positive identifications only.

If equivalent sets of gross errors are existent, instead of single measurements, sets of measurements are obtained by the serial elimination procedure (see chapter 3.3.3). If the procedure identifies the correct set of measurements, this is reported as true positive identification in the confusion matrix. However, it is actually not clear which of the measurements in the set is affected by the gross error. Especially for the reduced sensor layouts this might be a crucial point, since for these layouts equivalent sets of gross errors are frequent.

The error identification performance of the applied serial elimination strategy is rated as insufficient for a bias of 10 %. In further studies it should therefore be analysed how the performance changes if smaller or larger biases are affecting the measurements. Furthermore, other methods for error identification could be tested, for instance the principal component measurement test as stated by Narasimhan & Jordache (1999).

5.3 Alternative sensor layouts

If only less sensors are available for data reconciliation, this leads to a decrease in the data reconciliation performance of these sensors (see chapter 4.4). This is expected, since the reduced layouts can only use less balances for the reconciliation. This is making the optimization method less efficient. For the reduced layout of model B an appropriate reduction in the standard errors is still reached for the redundant measurements of the hydraulic flows and the component flows.

The CUSUM charts of the reduced layout of model B yield in a performance which – for most of the redundant measurements – is comparable to the performance of the charts of the original layout. Exceptions are the CUSUM statistics of the hydraulic flow measurements of the influent and the effluent. They show practically no reaction on the biases imposed. This is not expected, since these measurements are both redundant and are contributing to a major part to the hydraulic flow balance of the environmental node. Therefore, errors in these measurements should be easily detected. Since no other explanation is found for the behaviour of the CUSUM charts of these two measurements, it is hypothesized that an error may be present in the programs written to this end. This should be examined further though.

However, the results obtained for the data reconciliation and CUSUM chart performance still show that these methods are also functional with less sensors installed. Future studies should therefore aim on finding the minimum number of sensors necessary to reach a certain performance in data reconciliation and gross error detection. Additionally, it should be analysed how the location of a sensor influences the performance of those methods.

6 Conclusions

In this thesis it is shown that the measurement accuracy of hydraulic flows and TSS component flows can be improved by means of bilinear data reconciliation. If TSS storage and TSS reaction processes are neglected, this reduces the efficiency of the used data reconciliation method in a decisive way. The usage of imaginary TSS flows is found to be a useful method to account for these processes and to enhance the reconciliation performance. For the TSS concentrations, however, only for some sensors of the BSM1_LT plant an improvement in measurement accuracy is reached by data reconciliation while for other sensors the accuracy is deteriorated.

CUSUM charts are observed to be much more efficient than the global test in detecting gross errors. This follows from the fact that CUSUM charts are, in contrast to the global test, not only exploiting the spatial redundancy of the measurements, but also their temporal redundancy.

It is found to be beneficial for the performance of both types of the investigated gross error detection methods to use imaginary component flows to account for TSS storage and TSS reaction processes: The test statistic of the global test is strongly affected if these processes are neglected. Furthermore, most of the CUSUM charts which are not considering these processes are failing the validation.

However, one of the examined univariate CUSUM charts is found to pass the validation and is even showing a good gross error detection performance, although this chart is not accounting for TSS storage and is modelling the TSS reaction by a constant rate only.

Finally, the comparison between the univariate and the multivariate CUSUM charts shows that for large bias sizes the multivariate charts generally reach a better gross error detection performance than the univariate charts. On the other hand, the univariate charts are better in detecting small biases in the concentration measurements.

References

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

Crowe, C. M. 1986. Reconciliation of process flow rates by matrix projection. Part II: The nonlinear case. *AIChE Journal*, **32**(4), 616–623.

Crowe, C. M., Campos, Y. A. Garcia, & Hrymak, A. 1983. Reconciliation of process flow rates by matrix projection. Part I: Linear case. *AIChE Journal*, **29**(6), 881–888.

Dunia, Ricardo, Qin, S. Joe, Edgar, Thomas F., & McAvoy, Thomas J. 1996. Identification of faulty sensors using principal component analysis. *AIChE Journal*, **42**(10), 2797–2812.

Gernaey, K.V., Jeppsson, U., Vanrolleghem, P.A., & Copp, J.B. 2014. *Benchmarking of control strategies for wastewater treatment plants.* Tech. rept. International Water Association (IWA). Scientific and Technical Report No. 23.

Goodall, C. R. 1993. Computation using the QR decomposition. *Handbook of Statistics*, **9**, 467–508.

Goodman, Leo A. 1960. On the Exact Variance of Products. *Journal of the American Statistical Association*, **55**(292), 708–713.

Lucas, J.M., & Crosier, R.B. 2000. Fast initial response for CUSUM quality-control schemes: Give your CUSUM a head start. *Technometrics*, **42**(1), 102–107.

Mahmoud, M.A., & Maravelakis, P.E. 2013. The performance of multivariate CUSUM control charts with estimated parameters. *Journal of Statistical Computation and Simulation*, **83**(4), 721–738.

Maurer, M., & Herlyn, A. 2006. Zustand, Kosten und Investitionsbedarf der schweizerischen Abwasserentsorgung.

Meijer, S.C.F, van der Spoel, H., Susanti, S., Heijnen, J.J., & van Loosdrecht, M.C.M. 2002. Error diagnostics and data reconciliation for activated sludge modelling using mass balances. *Water Science and Technology*, **45**(6), 145–156.

Montgomery, D. C. 2009. *Introduction to Statistical Quality Control.* John Wiley & Sons.

Narasimhan, S., & Jordache, C. 1999. *Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data.* Elsevier Science.

Pignatiello, J. J., & Runger, G. C. 1990. Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, **22**(3), 173–186.

Rieger, L., Takacs, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P.A., & Comeau, Y. 2010. Data reconciliation for wastewater treatment plant simulation studies – Planning for high-quality data and typical sources of errors. *Water Environment Research*, **82**(5), 426–433.

Ripps, D.L. 1965. Adjustment of Experimental Data. *Chemical engineering progress symposium series*, **61**, 8–13.

Sanchez, M., & Romagnoli, J. 1996. Use of orthogonal transformations in data classification-reconciliation. *Computers and chemical engineering*, **20**(5), 483–493.

Spindler, A. 2014. Structural redundancy of data from wastewater treatment systems. Determination of individual balance equations. *Water Research*, **57**, 193–201.

Spindler, A., & Vanrolleghem, P.A. 2012. Dynamic mass balancing for wastewater treatment data quality control using CUSUM charts. *Water Science & Technology*, **65**(12), 2148–2153.

Villez, K., Vanrolleghem, P.A., & Corominas, L. 2013a. Sensor fault detection and diagnosis based on bilinear mass balances in wastewater treatment systems. *In: Proceedings of the 11th IWA conference on instrumentation control and automation (ICA2013).*

Villez, K., Vanrolleghem, P.A., & Corominas, L. 2013b. Structural observability and redundancy classification for sensor networks in wastewater systems. *In: Proceedings of the 11th IWA conference on instrumentation control and automation (ICA2013).*

A Appendix: Material and Methods

Table A1: Measurement range and corresponding noise for the hydraulic flow and concentration sensors.

	Hydraulic flows Q ($m^3\,d^{-1}$)		TSS concentrations c ($g\,m^{-3}$)	
Stream	Range	Noise δ_Q	Range	Noise δ_c
1	60'000	1'500	400	10
2	200'000	5'000	5'000	125
3	200'000	5'000	5'000	125
4	200'000	5'000	5'000	125
5	80'000	2'000	5'000	125
6	60'000	1'500	30	0.75
7	20'000	500	10'000	250
8	20'000	500	10'000	250
9	100'000	2'500	5'000	125
10	80'000	2'000	6'000	150
11	0	0	0	0
12	500	12.5	10'000	250

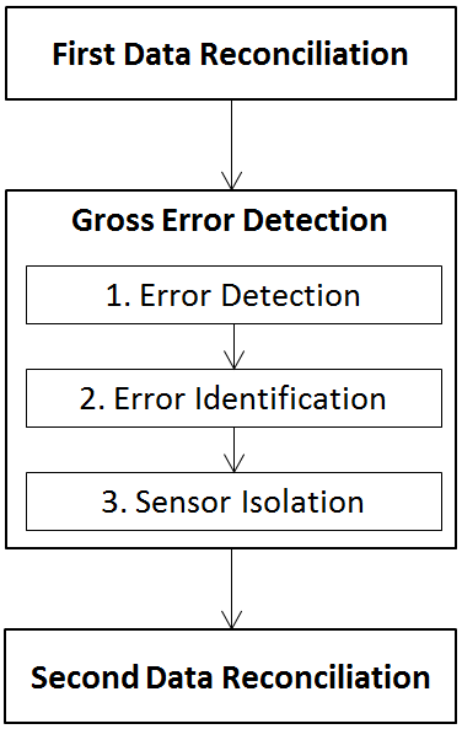


Figure A1: Gross error elimination procedure.

B Appendix: Results

B.1 Data reconciliation

In the following tables the results of the data reconciliation performance measures of the four different models are shown. Measured values are marked with a tilde symbol and a hat signals a reconciled estimate. If a variable has neither of these symbols, it corresponds to the real value.

Table B2: Data reconciliation performance model A: Relative means of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\mu(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\tilde{c}-c)}{\mu(c)}$	$\frac{\mu(\hat{c}-c)}{\mu(c)}$	$\frac{\mu(\tilde{f}-f)}{\mu(f)}$	$\frac{\mu(\hat{f}-f)}{\mu(f)}$
1	0.00 %	0.00 %	0.00 %	-27.00 %	0.00 %	-33.32 %
2	0.00 %	0.00 %	0.00 %	-0.51 %	0.00 %	-0.53 %
3	0.00 %	0.00 %	0.00 %	-0.51 %	0.00 %	-0.53 %
4	0.00 %	0.00 %	0.00 %	-0.03 %	0.00 %	-0.09 %
5	0.00 %	0.00 %	0.00 %	0.79 %	0.00 %	-0.14 %
6	0.00 %	0.00 %	0.00 %	1.39 %	0.01 %	1.93 %
7	0.00 %	0.00 %	0.00 %	-0.14 %	0.00 %	-0.15 %
8	0.00 %	0.00 %	0.00 %	-0.21 %	0.00 %	-0.22 %
9	0.00 %	0.00 %	0.00 %	0.09 %	0.00 %	-0.05 %
10	0.00 %	0.00 %	0.00 %	-1.17 %	0.00 %	-1.14 %
12	0.00 %	0.00 %	0.00 %	3.96 %	0.00 %	3.92 %

Table B3: Data reconciliation performance model B: Relative means of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\mu(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\tilde{c}-c)}{\mu(c)}$	$\frac{\mu(\hat{c}-c)}{\mu(c)}$	$\frac{\mu(\tilde{f}-f)}{\mu(f)}$	$\frac{\mu(\hat{f}-f)}{\mu(f)}$
1	0.00 %	0.00 %	0.00 %	-0.02 %	0.00 %	-0.02 %
2	0.01 %	0.00 %	0.00 %	-0.19 %	0.01 %	-0.19 %
3	0.00 %	0.00 %	0.00 %	-0.19 %	0.00 %	-0.19 %
4	0.00 %	0.00 %	0.00 %	-0.16 %	0.00 %	-0.15 %
5	0.00 %	0.00 %	0.00 %	-0.07 %	0.00 %	-0.07 %
6	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %
7	0.00 %	0.00 %	0.00 %	-0.14 %	0.00 %	-0.15 %
8	0.00 %	0.00 %	0.00 %	-0.14 %	0.00 %	-0.15 %
9	0.00 %	0.00 %	0.00 %	-0.28 %	0.00 %	-0.22 %
10	0.00 %	0.00 %	0.00 %	-0.11 %	0.00 %	-0.14 %
12	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %

Table B4: Data reconciliation performance model C: Relative means of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\mu(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\tilde{c}-c)}{\mu(c)}$	$\frac{\mu(\hat{c}-c)}{\mu(c)}$	$\frac{\mu(\tilde{f}-f)}{\mu(f)}$	$\frac{\mu(\hat{f}-f)}{\mu(f)}$
1	0.00 %	0.00 %	0.00 %	0.02 %	0.00 %	-0.01 %
2	0.00 %	0.00 %	0.00 %	-0.31 %	0.00 %	-0.27 %
3	0.00 %	0.00 %	0.00 %	-0.31 %	0.00 %	-0.27 %
4	0.00 %	0.00 %	0.00 %	-0.17 %	0.00 %	-0.27 %
5	0.00 %	0.00 %	0.00 %	0.63 %	0.00 %	-0.32 %
6	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.01 %
7	0.00 %	0.00 %	0.00 %	-0.32 %	0.00 %	-0.33 %
8	0.00 %	0.00 %	0.00 %	-0.33 %	0.00 %	-0.34 %
9	0.00 %	0.00 %	0.00 %	-0.21 %	0.00 %	-0.23 %
10	0.00 %	0.00 %	0.00 %	-0.46 %	0.01 %	-0.33 %
12	0.00 %	0.00 %	0.00 %	-0.01 %	0.00 %	0.00 %

Table B5: Data reconciliation performance model D: Relative means of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\mu(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\tilde{c}-c)}{\mu(c)}$	$\frac{\mu(\hat{c}-c)}{\mu(c)}$	$\frac{\mu(\tilde{f}-f)}{\mu(f)}$	$\frac{\mu(\hat{f}-f)}{\mu(f)}$
1	0.00 %	0.00 %	0.00 %	9.92 %	0.00 %	-1.27 %
2	0.00 %	0.00 %	0.00 %	-0.22 %	0.00 %	-0.27 %
3	0.00 %	0.00 %	0.00 %	-0.22 %	0.00 %	-0.27 %
4	0.00 %	0.00 %	0.00 %	-0.24 %	0.00 %	-0.27 %
5	0.00 %	0.00 %	0.00 %	0.58 %	0.01 %	-0.34 %
6	0.00 %	0.00 %	0.00 %	-1.43 %	0.00 %	-0.41 %
7	0.00 %	0.00 %	0.00 %	-0.34 %	0.00 %	-0.35 %
8	0.00 %	0.00 %	0.00 %	-0.29 %	0.00 %	-0.30 %
9	0.00 %	0.00 %	0.00 %	-0.33 %	0.00 %	-0.22 %
10	0.00 %	0.00 %	0.00 %	-0.34 %	0.00 %	-0.33 %
12	0.00 %	0.00 %	0.00 %	-3.43 %	0.00 %	-3.36 %

Table B6: Data reconciliation performance model E: Relative means of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\mu(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\tilde{c}-c)}{\mu(c)}$	$\frac{\mu(\hat{c}-c)}{\mu(c)}$	$\frac{\mu(\tilde{f}-f)}{\mu(f)}$	$\frac{\mu(\hat{f}-f)}{\mu(f)}$
1	0.00 %	0.00 %	0.00 %	-27.00 %	0.01 %	-33.32 %
2	0.00 %	0.00 %	0.00 %	-0.37 %	0.00 %	-0.42 %
3	0.00 %	0.00 %	0.00 %	-0.37 %	0.00 %	-0.42 %
4	0.00 %	0.00 %	0.00 %	-0.08 %	0.00 %	-0.06 %
5	0.00 %	0.00 %	0.00 %	-0.09 %	0.00 %	-0.10 %
6	0.00 %	0.00 %	0.00 %	1.38 %	0.00 %	1.92 %
7	0.00 %	0.00 %	0.00 %	0.08 %	0.00 %	0.08 %
8	0.00 %	0.00 %	0.00 %	0.02 %	0.00 %	0.02 %
9	0.00 %	0.00 %	0.00 %	-0.24 %	-0.01 %	-0.03 %
10	0.00 %	0.00 %	0.00 %	-0.77 %	0.00 %	-0.91 %
12	0.00 %	0.00 %	0.00 %	3.97 %	0.00 %	3.93 %

Table B7: Data reconciliation performance model A: Relative standard deviations of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\sigma(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\tilde{c}-c)}{\mu(c)}$	$\frac{\sigma(\hat{c}-c)}{\mu(c)}$	$\frac{\sigma(\tilde{f}-f)}{\mu(f)}$	$\frac{\sigma(\hat{f}-f)}{\mu(f)}$
1	7.28 %	4.08 %	5.00 %	34.16 %	9.38 %	37.82 %
2	5.65 %	2.20 %	3.27 %	1.98 %	6.75 %	2.97 %
3	5.65 %	2.20 %	3.27 %	1.98 %	6.75 %	2.97 %
4	5.64 %	2.20 %	3.28 %	4.69 %	6.75 %	5.00 %
5	5.12 %	2.19 %	3.28 %	10.15 %	6.23 %	10.75 %
6	7.40 %	4.14 %	5.19 %	8.98 %	8.72 %	9.34 %
7	2.67 %	1.85 %	3.20 %	2.84 %	4.18 %	3.15 %
8	2.71 %	1.88 %	3.20 %	2.87 %	4.21 %	3.19 %
9	5.04 %	3.89 %	3.28 %	105.37 %	6.36 %	4.84 %
10	5.12 %	2.19 %	3.86 %	3.25 %	6.66 %	3.17 %
12	4.16 %	4.16 %	3.20 %	7.70 %	5.32 %	8.43 %

Table B8: Data reconciliation performance model B: Relative standard deviations of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\sigma(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\tilde{c}-c)}{\mu(c)}$	$\frac{\sigma(\hat{c}-c)}{\mu(c)}$	$\frac{\sigma(\tilde{f}-f)}{\mu(f)}$	$\frac{\sigma(\hat{f}-f)}{\mu(f)}$
1	7.28 %	4.08 %	5.00 %	8.98 %	9.38 %	9.37 %
2	5.64 %	2.20 %	3.27 %	1.74 %	6.75 %	2.76 %
3	5.65 %	2.20 %	3.27 %	1.74 %	6.75 %	2.76 %
4	5.64 %	2.20 %	3.29 %	2.77 %	6.75 %	3.43 %
5	5.12 %	2.19 %	3.28 %	5.59 %	6.23 %	5.82 %
6	7.39 %	4.14 %	5.19 %	8.68 %	8.72 %	8.72 %
7	2.67 %	1.85 %	3.20 %	2.11 %	4.18 %	2.64 %
8	2.71 %	1.88 %	3.20 %	2.14 %	4.21 %	2.68 %
9	5.05 %	3.89 %	3.29 %	46.46 %	6.36 %	4.79 %
10	5.12 %	2.19 %	3.87 %	2.94 %	6.66 %	2.62 %
12	4.16 %	4.16 %	3.20 %	3.20 %	5.32 %	5.32 %

Table B9: Data reconciliation performance model C: Relative standard deviations of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\sigma(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\tilde{c}-c)}{\mu(c)}$	$\frac{\sigma(\hat{c}-c)}{\mu(c)}$	$\frac{\sigma(\tilde{f}-f)}{\mu(f)}$	$\frac{\sigma(\hat{f}-f)}{\mu(f)}$
1	7.29 %	4.07 %	5.00 %	9.01 %	9.38 %	9.37 %
2	5.65 %	2.20 %	3.27 %	1.97 %	6.75 %	2.96 %
3	5.65 %	2.20 %	3.27 %	1.97 %	6.75 %	2.96 %
4	5.64 %	2.20 %	3.28 %	4.71 %	6.75 %	5.02 %
5	5.12 %	2.19 %	3.28 %	10.17 %	6.23 %	10.77 %
6	7.39 %	4.14 %	5.19 %	8.67 %	8.71 %	8.72 %
7	2.67 %	1.85 %	3.20 %	2.81 %	4.18 %	3.13 %
8	2.71 %	1.88 %	3.20 %	2.85 %	4.21 %	3.18 %
9	5.05 %	3.88 %	3.29 %	43.59 %	6.36 %	4.83 %
10	5.12 %	2.19 %	3.87 %	3.11 %	6.66 %	3.09 %
12	4.16 %	4.16 %	3.20 %	3.20 %	5.32 %	5.32 %

Table B10: Data reconciliation performance model D: Relative standard deviations of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\sigma(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\tilde{c}-c)}{\mu(c)}$	$\frac{\sigma(\hat{c}-c)}{\mu(c)}$	$\frac{\sigma(\tilde{f}-f)}{\mu(f)}$	$\frac{\sigma(\hat{f}-f)}{\mu(f)}$
1	7.29 %	4.08 %	5.00 %	43.24 %	9.38 %	36.32 %
2	5.64 %	2.20 %	3.27 %	1.97 %	6.75 %	2.96 %
3	5.64 %	2.20 %	3.27 %	1.97 %	6.75 %	2.96 %
4	5.64 %	2.20 %	3.28 %	4.70 %	6.75 %	5.01 %
5	5.12 %	2.19 %	3.29 %	10.14 %	6.23 %	10.76 %
6	7.39 %	4.14 %	5.19 %	9.27 %	8.71 %	9.26 %
7	2.67 %	1.85 %	3.20 %	2.83 %	4.18 %	3.14 %
8	2.71 %	1.88 %	3.20 %	2.86 %	4.21 %	3.18 %
9	5.05 %	3.89 %	3.28 %	57.56 %	6.36 %	4.84 %
10	5.12 %	2.19 %	3.87 %	3.24 %	6.65 %	3.16 %
12	4.16 %	4.16 %	3.20 %	10.73 %	5.32 %	11.11 %

Table B11: Data reconciliation performance model E: Relative standard deviations of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\sigma(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\tilde{c}-c)}{\mu(c)}$	$\frac{\sigma(\hat{c}-c)}{\mu(c)}$	$\frac{\sigma(\tilde{f}-f)}{\mu(f)}$	$\frac{\sigma(\hat{f}-f)}{\mu(f)}$
1	7.28 %	4.07 %	5.00 %	34.16 %	9.38 %	37.82 %
2	5.65 %	2.20 %	3.27 %	1.76 %	6.76 %	2.78 %
3	5.65 %	2.20 %	3.27 %	1.76 %	6.75 %	2.78 %
4	5.65 %	2.20 %	3.28 %	2.77 %	6.75 %	3.43 %
5	5.13 %	2.19 %	3.29 %	5.59 %	6.23 %	5.83 %
6	7.39 %	4.13 %	5.19 %	8.97 %	8.71 %	9.32 %
7	2.67 %	1.85 %	3.20 %	2.13 %	4.18 %	2.65 %
8	2.71 %	1.88 %	3.20 %	2.15 %	4.21 %	2.69 %
9	5.05 %	3.89 %	3.28 %	68.82 %	6.36 %	4.80 %
10	5.12 %	2.19 %	3.87 %	3.09 %	6.66 %	2.75 %
12	4.16 %	4.16 %	3.20 %	7.70 %	5.31 %	8.43 %

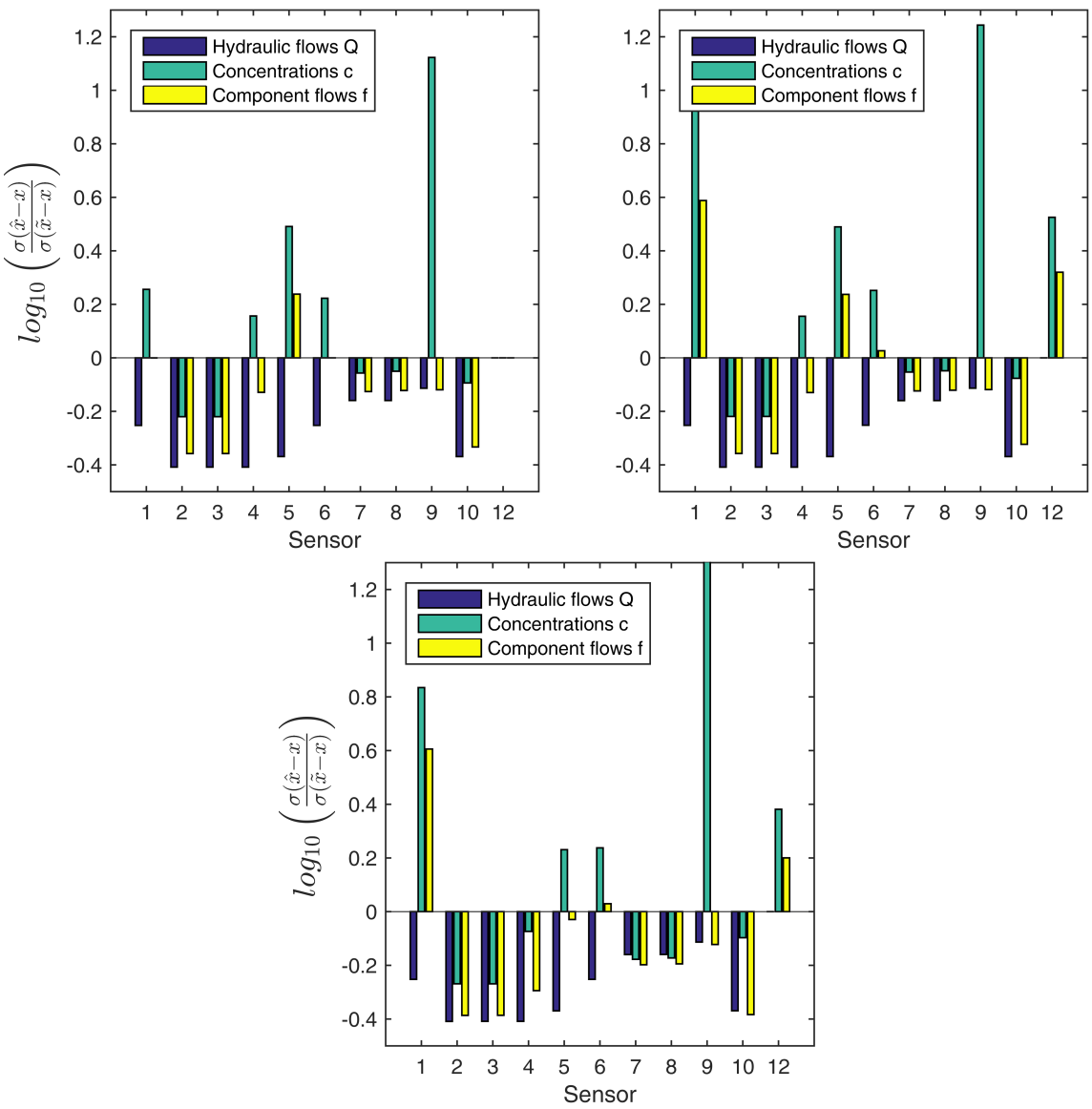


Figure B2: Proportion between the standard error of the reconciled estimates and the measurements for the models C (top left), D (top right) and E (bottom).

B.2 Gross error detection

B.2.1 Global test

The following figures show additional results of the global test.

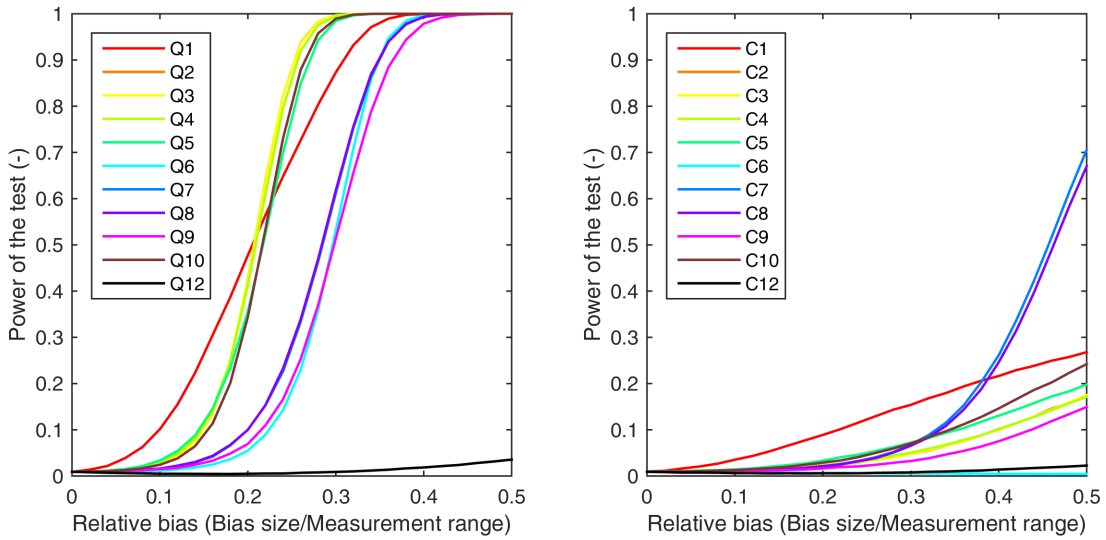


Figure B3: Power of the global test for model A depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

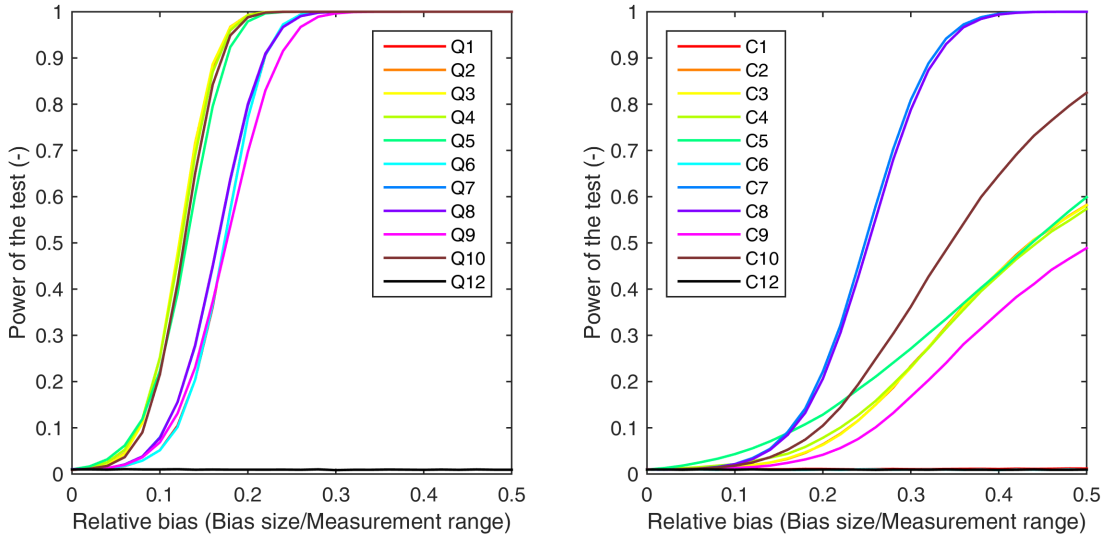


Figure B4: Power of the global test for model C depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

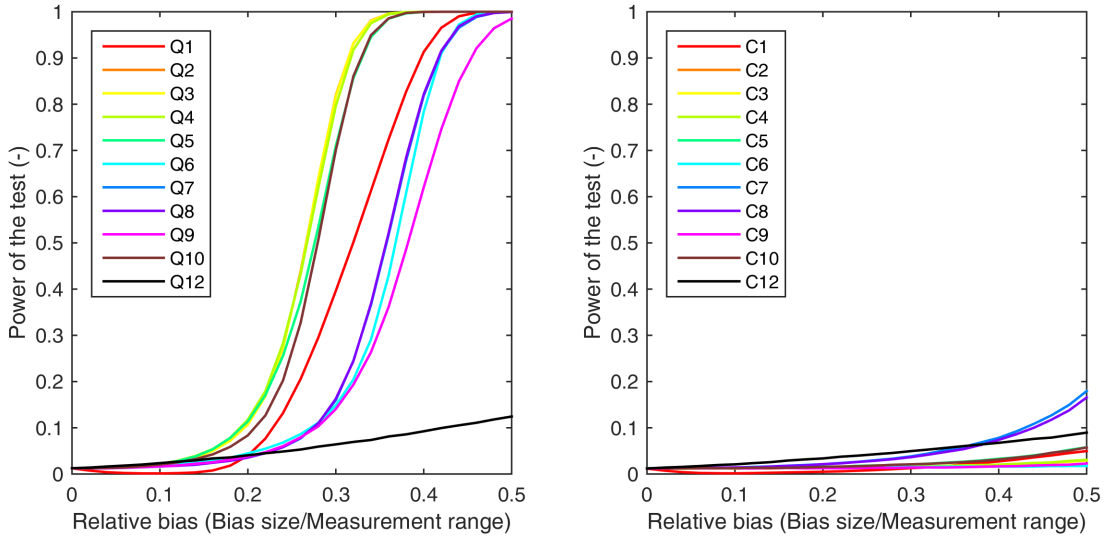


Figure B5: Power of the global test for model D depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

B.2.2 CUSUM charts

The following figures show additional results for the different types of CUSUM charts.

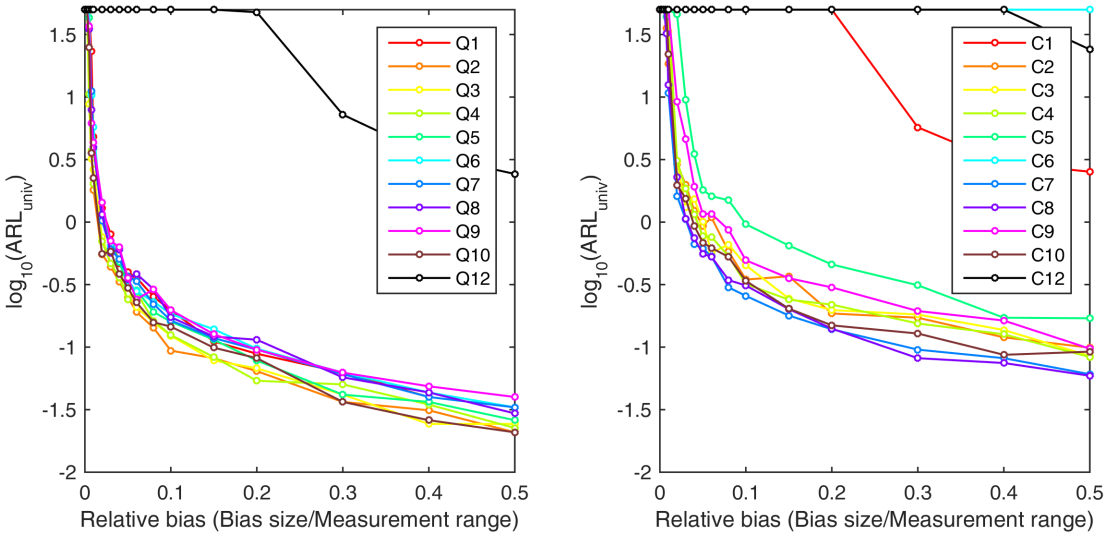


Figure B6: Average biased run lengths of the multivariate CUSUM chart for model B depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

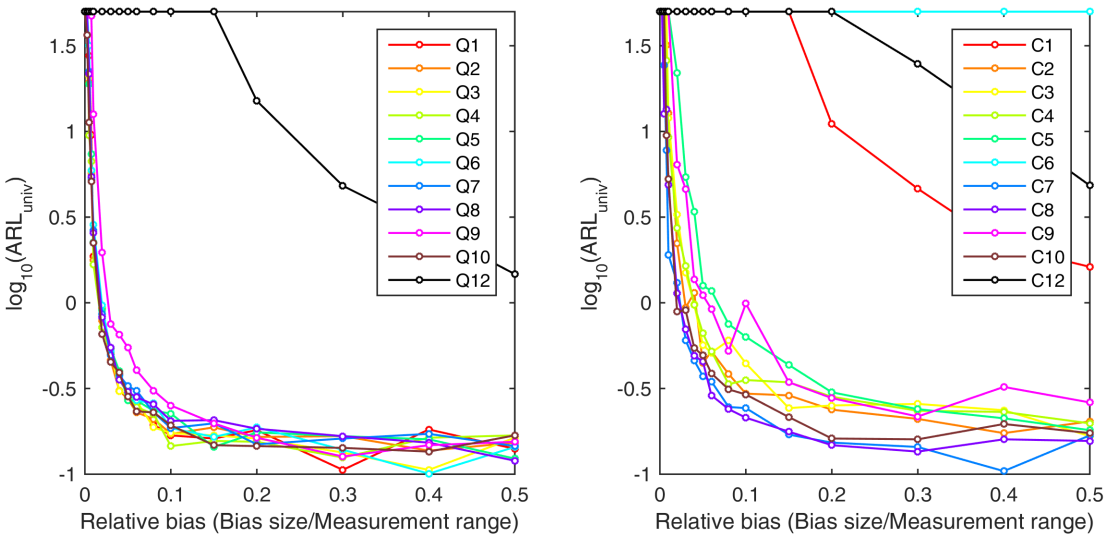


Figure B7: Average biased run lengths of the univariate CUSUM chart for model D depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

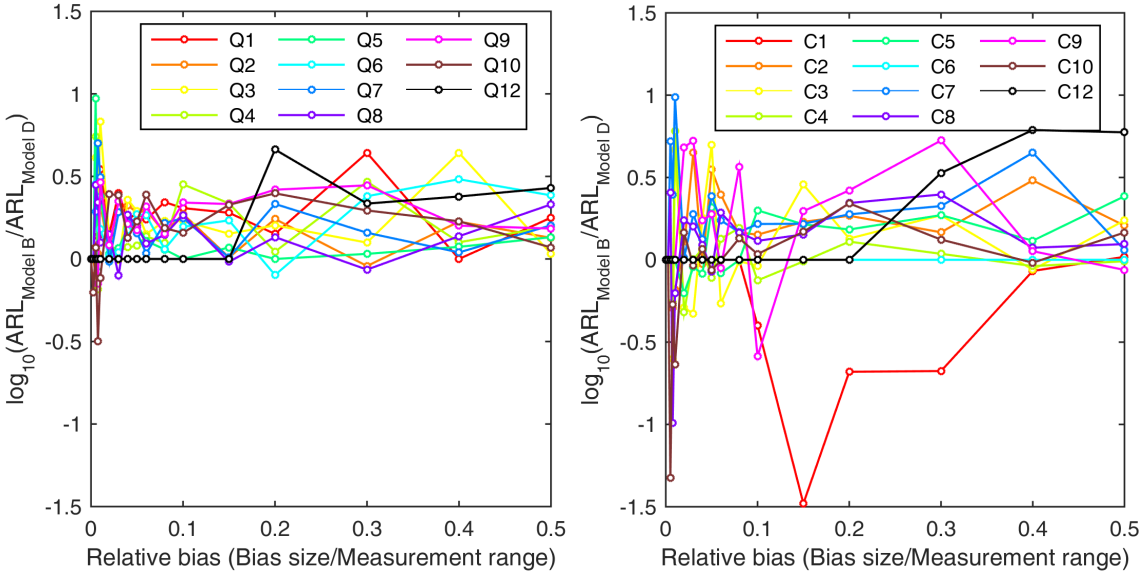


Figure B8: Ratio of the average biased run lengths (ARL_μ) of the univariate CUSUM charts of model B and D depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

Example CUSUM charts In the following, examples of CUSUM charts are shown. The colored lines of the univariate charts are standing each for an univariate chart of the residuals of one balance.

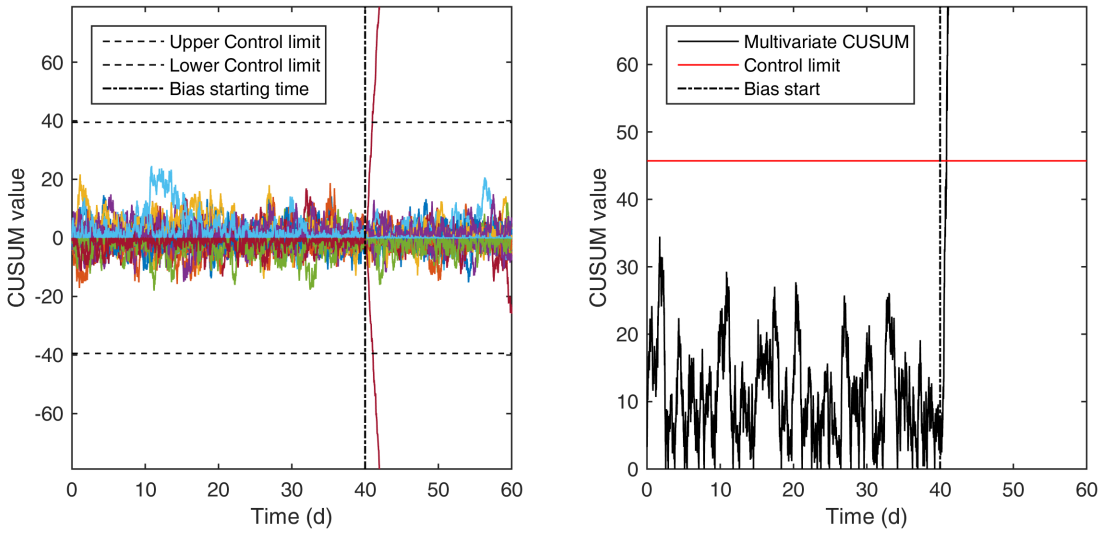


Figure B9: Examples of the behaviour of the univariate (left) and the multivariate (right) CUSUM chart of model B for a bias of 10 % affecting the concentration measurement of stream 4 on day 40 of the validation period.

In the following figures, examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 are shown for the different models and the two types of CUSUM charts. For these simulations no bias was affecting the measurements.

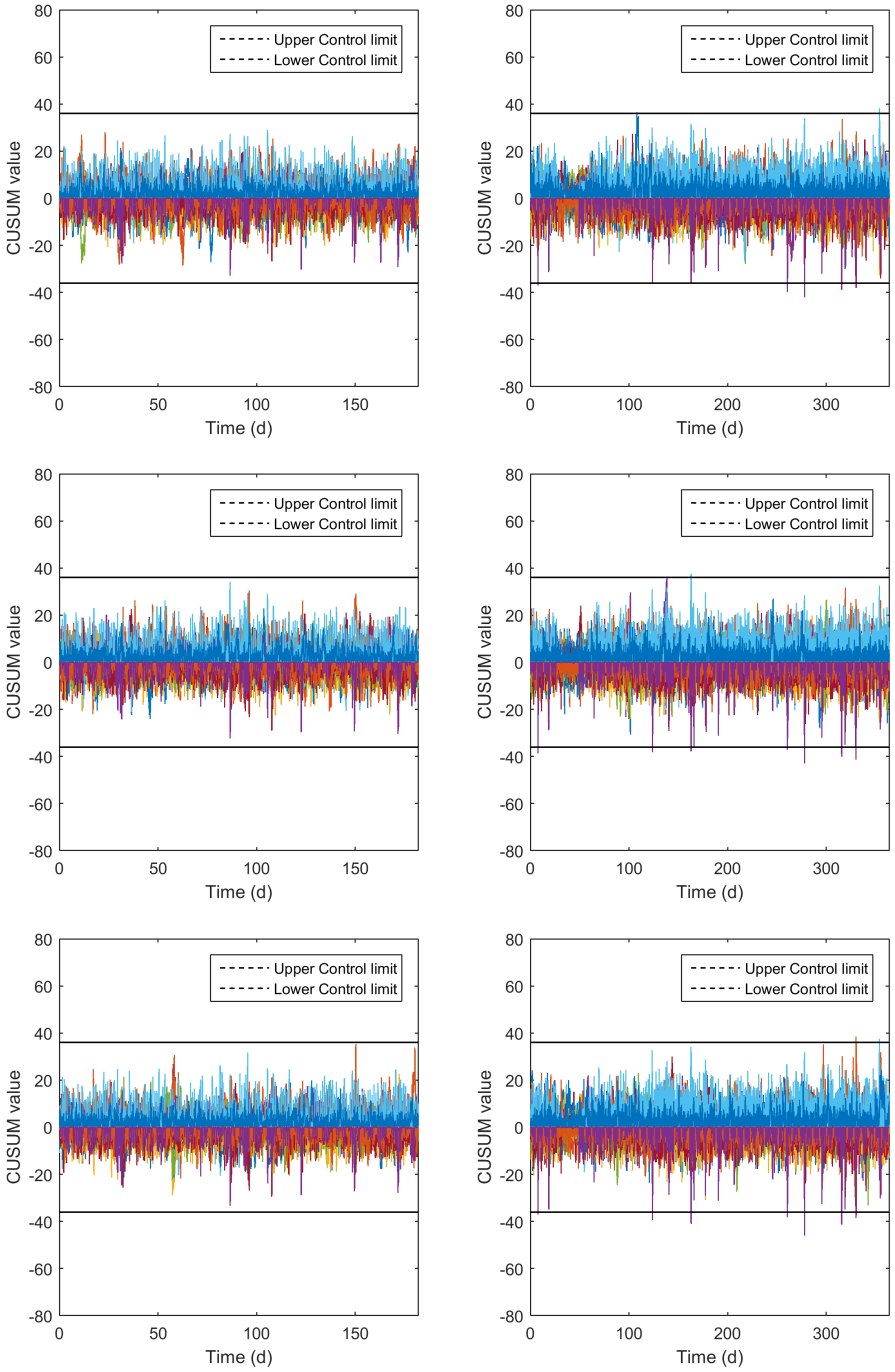


Figure B10: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the univariate CUSUM chart of model A. Calibration: left, Validation: right.

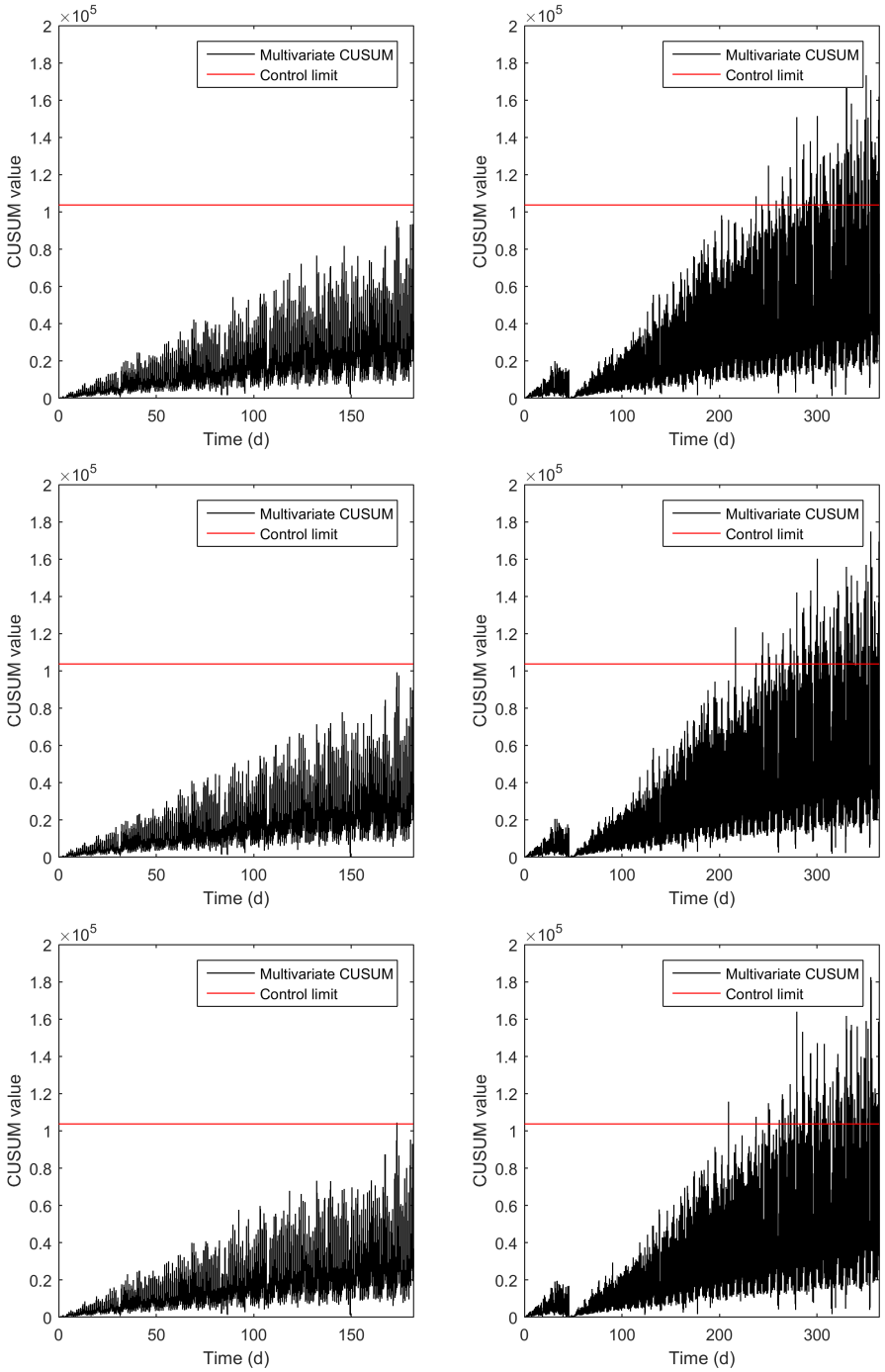


Figure B11: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the multivariate CUSUM chart of model A. Calibration: left, Validation: right.

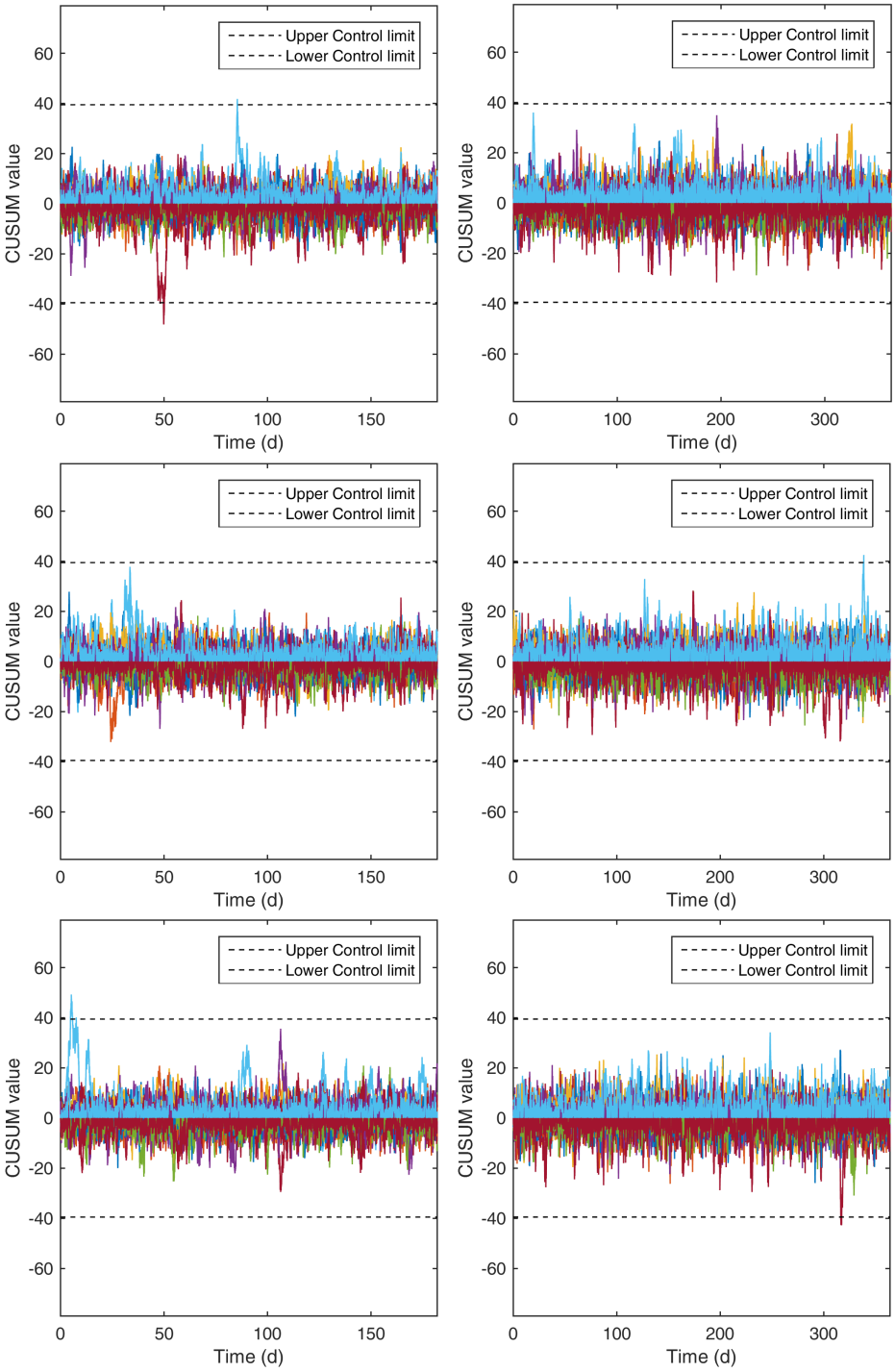


Figure B12: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the univariate CUSUM chart of model B. Calibration: left, Validation: right.

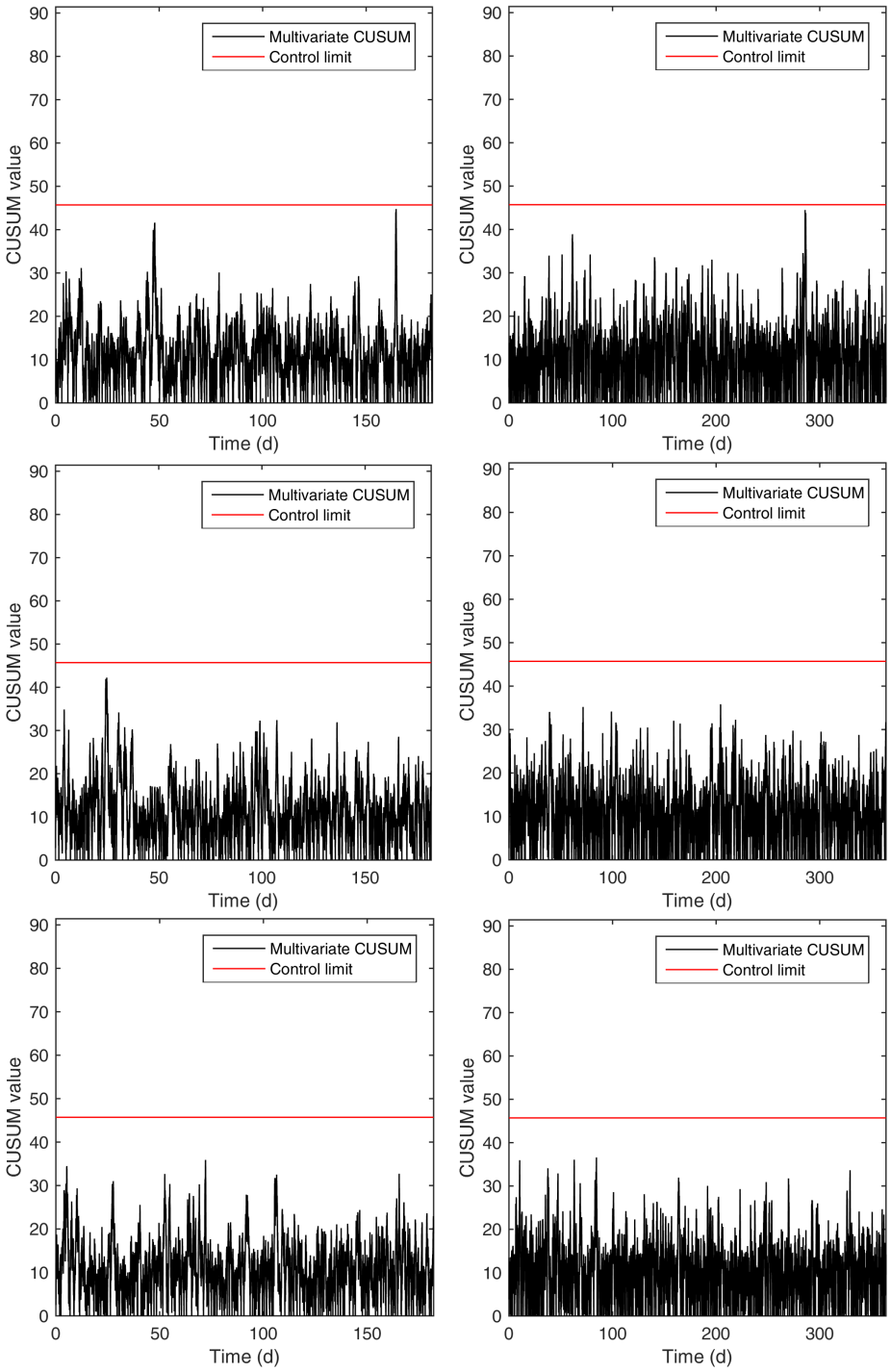


Figure B13: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the multivariate CUSUM chart of model B. Calibration: left, Validation: right.

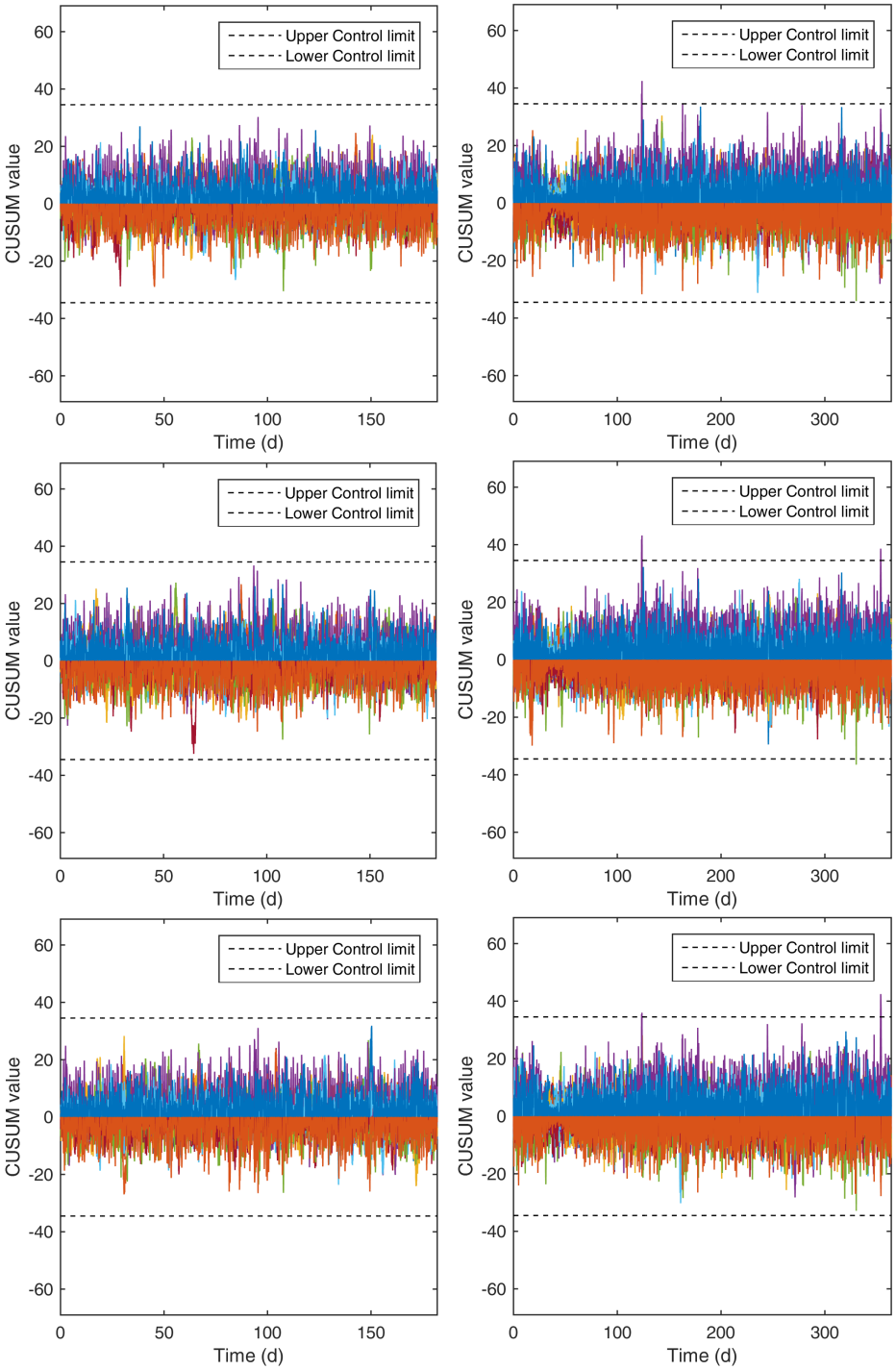


Figure B14: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the univariate CUSUM chart of model C. Calibration: left, Validation: right.

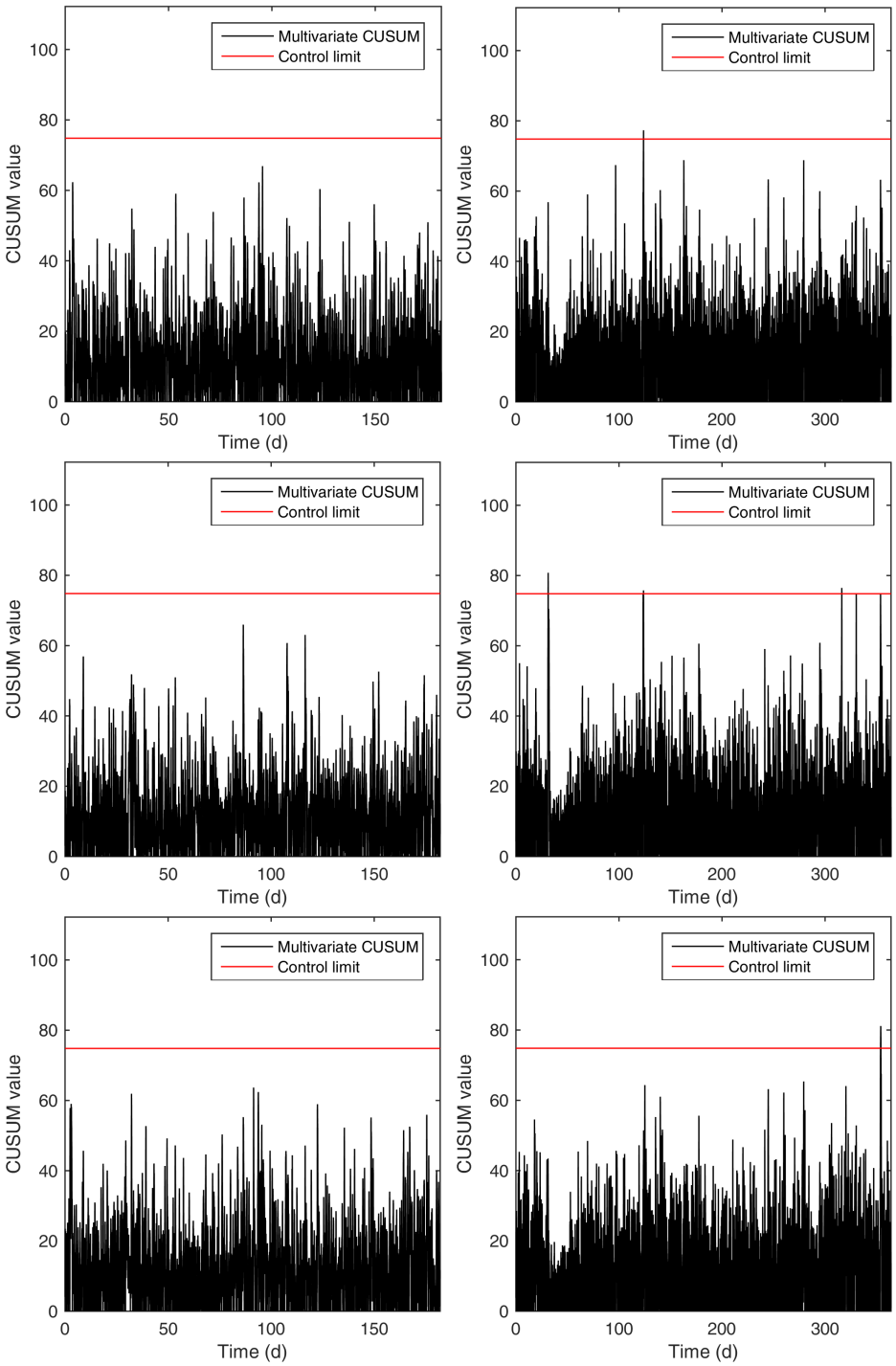


Figure B15: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the multivariate CUSUM chart of model C. Calibration: left, Validation: right.

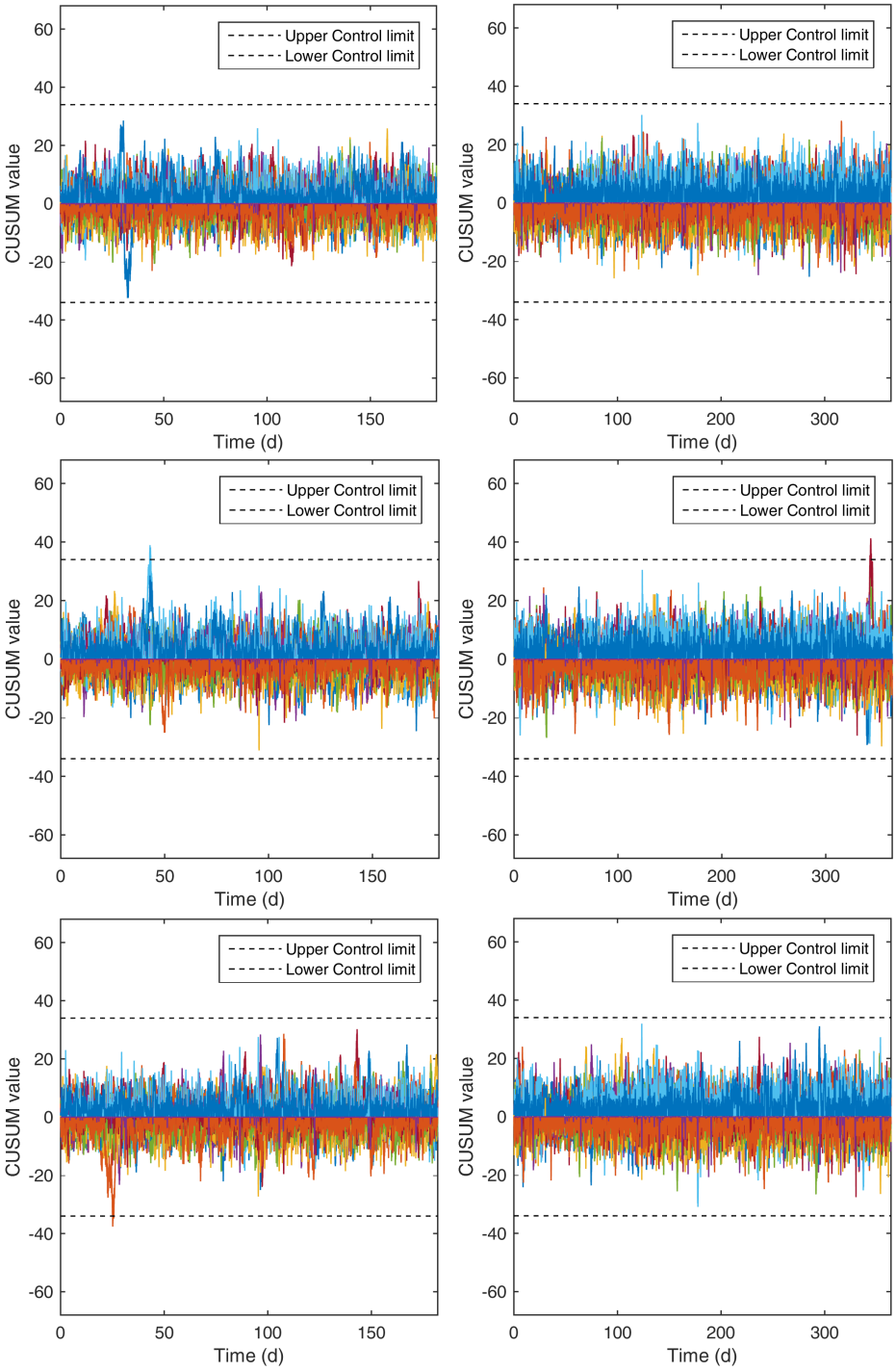


Figure B16: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the univariate CUSUM chart of model D. Calibration: left, Validation: right.

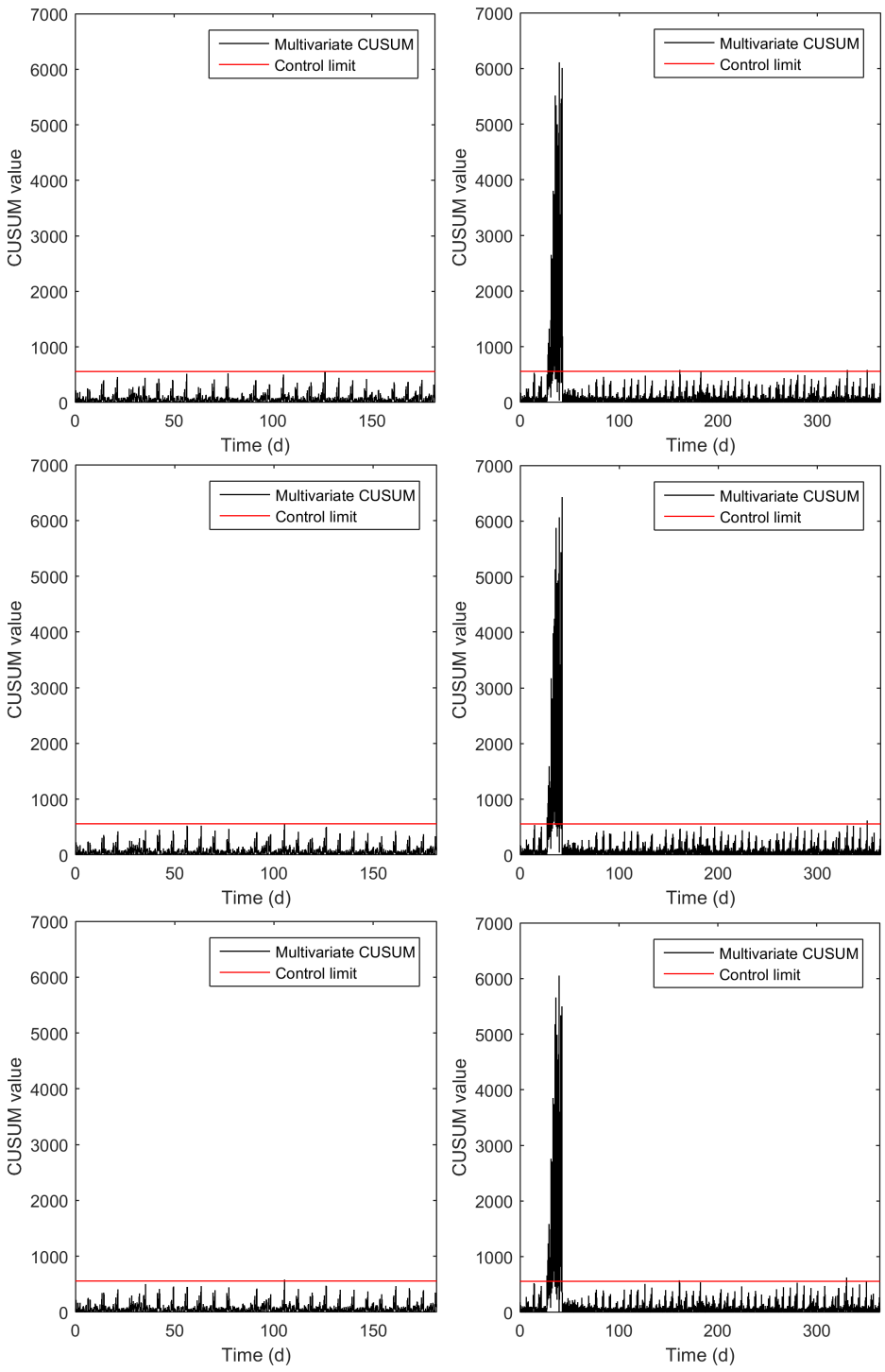


Figure B17: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the multivariate CUSUM chart of model D. Calibration: left, Validation: right.

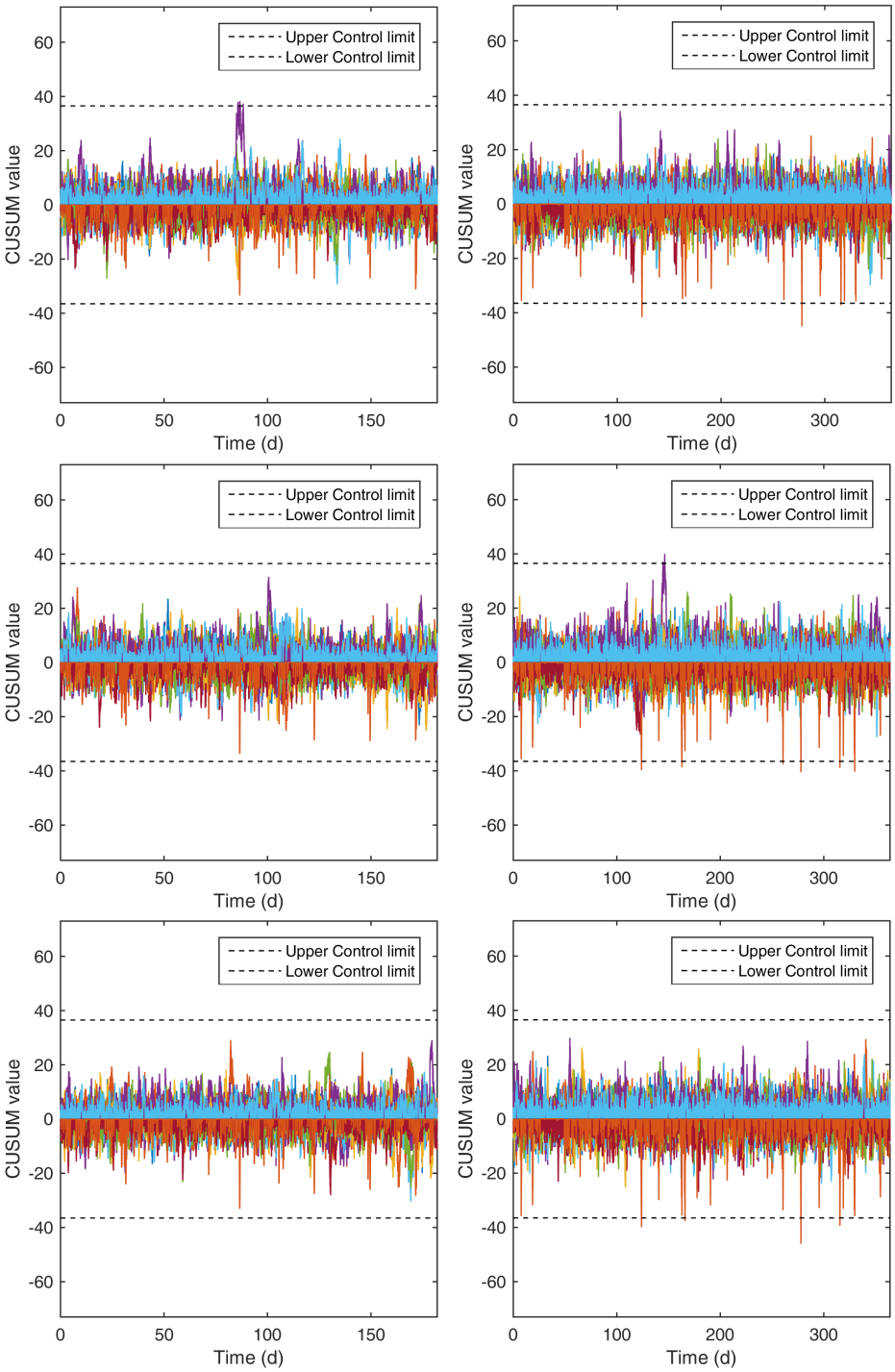


Figure B18: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the univariate CUSUM chart of model E. Calibration: left, Validation: right.

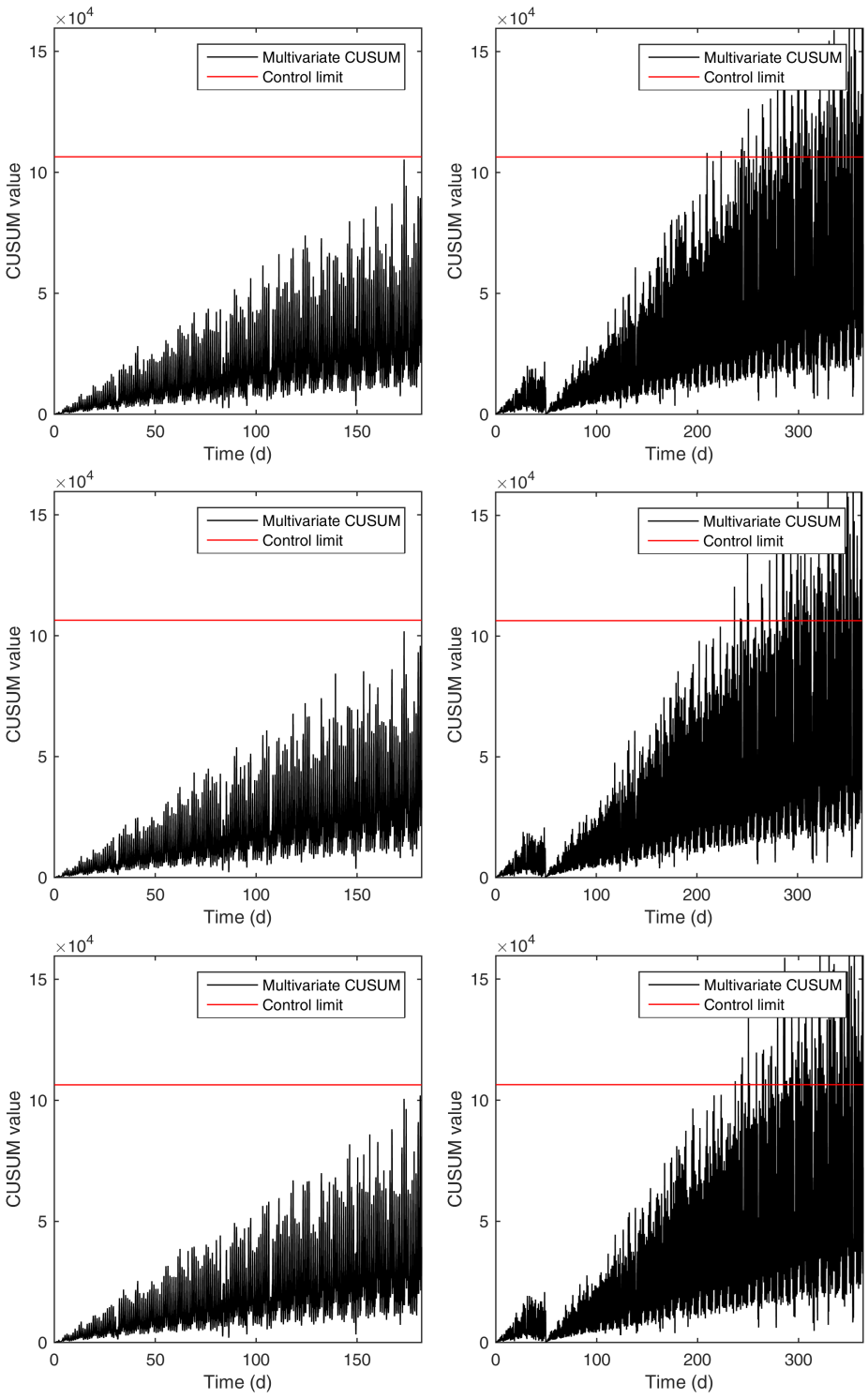


Figure B19: Examples of different Monte Carlo simulation runs which were performed to calibrate the ARL_0 for the multivariate CUSUM chart of model E. Calibration: left, Validation: right.

B.2.3 Error identification

Table B12: Confusion matrix with true and false positive identification rates of the error identification procedure for a 10 % bias in each of the sensors of model B. The rows correspond to the biased sensors, the columns to the sensors which were indicated to be biased by the error identification method. The true positive identifications are marked in bold.

		Sensor identified																						
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Sensor biased	Q1	**0.80**	0	0	0.05	0	0	0	0.05	0	0	0.05	0	0	0	0.05	0	0	0	0	0	0	0	0	0
	Q2	0	**0.95**	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0
	Q3	0	0	**0.85**	0.10	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Q4	0	0.05	0.05	**0.85**	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Q5	0	0.10	0.05	0	**0.65**	0	0	0	0	0.05	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0
	Q6	0	0	0	0.05	0	**0.85**	0	0	0	0.05	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0
	Q7	0	0.10	0.1	0.05	0	0	**0.50**	0	0.05	0.10	0	0.05	0	0	0	0.05	0	0	0	0	0	0	0	0
	Q8	0	0.05	0.1	0.05	0	0	0.05	**0.65**	0	0	0	0	0.05	0	0	0	0	0	0.05	0	0	0	0	0
	Q9	0	0	0	0.05	0	0	0	0	**0.9**	0.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Q10	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Q11	0	0.17	0.33	0.17	0	0	0	0	0	0	**0**	0.17	0.17	0	0	0	0	0	0	0	0	0	0	0
	Q12	0	0.14	0.07	0	0	0	0	0.07	0.14	0.43	0	**0**	0	0	0	0	0	0	0.07	0.07	0	0	0	0
	C1	0	0	0.25	0.13	0	0	0	0.13	0	0.13	0	0	**0**	0	0	0	0	0	0.13	0	0	0.25	0	0
	C2	0	0.05	0.05	0.05	0	0	0	0	0	0.05	0	0	0	**0.80**	0	0	0	0	0	0	0	0	0	0
	C3	0	0	0	0.05	0	0	0	0	0.05	0	0	0	0	0.05	**0.75**	0	0	0	0	0	0.05	0.05	0	0
	C4	0	0.05	0.05	0.05	0	0	0	0	0.05	0	0	0	0	0	0	**0.75**	0	0	0	0	0.05	0	0	0
	C5	0	0.10	0.05	0.05	0	0	0	0	0.10	0	0	0	0	0	0	0	**0.70**	0	0	0	0	0	0	0
	C6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	C7	0	0	0	0.05	0	0	0.05	0	0	0	0	0	0.05	0	0	0	0	0	**0.80**	0	0	0	0	0.05
	C8	0	0.05	0	0	0	0	0.05	0.05	0	0	0	0	0	0	0	0	0	0	0.05	**0.80**	0	0	0	0
	C9	0.05	0.05	0	0.10	0	0	0	0	0.15	0	0	0	0	0	0	0	0	0	0	0	**0.65**	0	0	0
	C10	0	0	0.05	0	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0	**0.90**	0	0
	C11	0	0.09	0.27	0.27	0	0	0	0.09	0.27	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0
	C12	0	0.20	0	0	0	0	0.20	0	0	0.20	0	0.20	0	0	0.20	0	0	0	0	0	0	0	0	**0**

B.3 Alternative sensor layouts

Table B13: Data reconciliation performance of model A with a reduced number of sensors: Relative means of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\mu(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\tilde{c}-c)}{\mu(c)}$	$\frac{\mu(\hat{c}-c)}{\mu(c)}$	$\frac{\mu(\tilde{f}-f)}{\mu(f)}$	$\frac{\mu(\hat{f}-f)}{\mu(f)}$
1	-0.01 %	-0.01 %	0.00 %	-26.9 %	-0.01 %	-33.3 %
2	-	0.00 %	-	-0.48 %	-	-0.45 %
3	-	0.00 %	-	-0.48 %	-	-0.45 %
4	-	0.00 %	0.00 %	0.00 %	-	0.00 %
5	-	0.00 %	-	-	-	-
6	-0.01 %	-0.01 %	0.00 %	1.37 %	-0.01 %	1.91 %
7	-	0.00 %	-	-	-	-
8	0.00 %	0.00 %	-	-	-	-
9	0.00 %	0.00 %	-	-	-	-
10	-	0.00 %	-	-	-	-
12	0.00 %	0.00 %	0.00 %	3.96 %	0.00 %	3.92 %

Table B14: Data reconciliation performance of model B with a reduced number of sensors: Relative means of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\mu(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\mu(\tilde{c}-c)}{\mu(c)}$	$\frac{\mu(\hat{c}-c)}{\mu(c)}$	$\frac{\mu(\tilde{f}-f)}{\mu(f)}$	$\frac{\mu(\hat{f}-f)}{\mu(f)}$
1	0.01 %	0.00 %	0.00 %	0.01 %	0.01 %	0.01 %
2	-	0.00 %	-	-0.09 %	-	-0.10 %
3	-	0.00 %	-	-0.09 %	-	-0.10 %
4	0.00 %	0.00 %	0.00 %	-0.10 %	0.01 %	-0.09 %
5	0.00 %	0.00 %	-0.01 %	-0.08 %	0.00 %	-0.08 %
6	-0.01 %	0.00 %	0.00 %	-0.01 %	-0.01 %	-0.01 %
7	0.00 %	0.00 %	0.00 %	-0.09 %	0.00 %	-0.09 %
8	0.00 %	0.00 %	0.00 %	-0.09 %	0.00 %	-0.09 %
9	0.01 %	0.01 %	0.00 %	-0.12 %	0.00 %	-0.10 %
10	-	0.00 %	-	-0.03 %	-	-0.08 %
12	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %

Table B15: Data reconciliation performance of model A with a reduced number of sensors: Relative standard deviations of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\sigma(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\tilde{c}-c)}{\mu(c)}$	$\frac{\sigma(\hat{c}-c)}{\mu(c)}$	$\frac{\sigma(\tilde{f}-f)}{\mu(f)}$	$\frac{\sigma(\hat{f}-f)}{\mu(f)}$
1	7.28 %	5.15 %	5.00 %	34.36 %	9.38 %	37.82 %
2	-	3.12 %	-	6.58 %	-	7.21 %
3	-	3.12 %	-	6.58 %	-	7.21 %
4	-	3.12 %	3.28 %	3.28 %	-	4.74 %
5	-	3.00 %	-	-	-	-
6	7.39 %	5.23 %	5.19 %	8.14 %	8.71 %	9.33 %
7	-	2.67 %	-	-	-	-
8	2.71 %	2.71 %	-	-	-	-
9	5.04 %	5.04 %	-	-	-	-
10	-	3.00 %	-	-	-	-
12	4.16 %	4.16 %	3.20 %	7.70 %	5.31 %	8.43 %

Table B16: Data reconciliation performance of model B with a reduced number of sensors: Relative standard deviations of the errors for hydraulic flows, concentrations and component flows.

	Hydraulic flows Q		Concentrations c		Component flows f	
	Meas.	Rec.	Meas.	Rec.	Meas.	Rec.
	$\frac{\sigma(\tilde{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\hat{Q}-Q)}{\mu(Q)}$	$\frac{\sigma(\tilde{c}-c)}{\mu(c)}$	$\frac{\sigma(\hat{c}-c)}{\mu(c)}$	$\frac{\sigma(\tilde{f}-f)}{\mu(f)}$	$\frac{\sigma(\hat{f}-f)}{\mu(f)}$
1	7.28 %	4.50 %	5.00 %	8.75 %	9.38 %	9.38 %
2	-	2.67 %	-	2.48 %	-	3.42 %
3	-	2.67 %	-	2.48 %	-	3.42 %
4	5.65 %	2.67 %	3.28 %	2.66 %	6.75 %	3.74 %
5	5.12 %	2.46 %	3.29 %	5.45 %	6.23 %	5.85 %
6	7.38 %	4.57 %	5.18 %	8.38 %	8.70 %	8.70 %
7	2.67 %	1.86 %	3.20 %	2.27 %	4.18 %	2.94 %
8	2.71 %	1.89 %	3.20 %	2.31 %	4.21 %	2.99 %
9	5.04 %	4.53 %	3.29 %	24.14 %	6.36 %	5.71 %
10	-	2.46 %	-	3.60 %	-	2.92 %
12	4.16 %	4.16 %	3.20 %	3.20 %	5.32 %	5.32 %

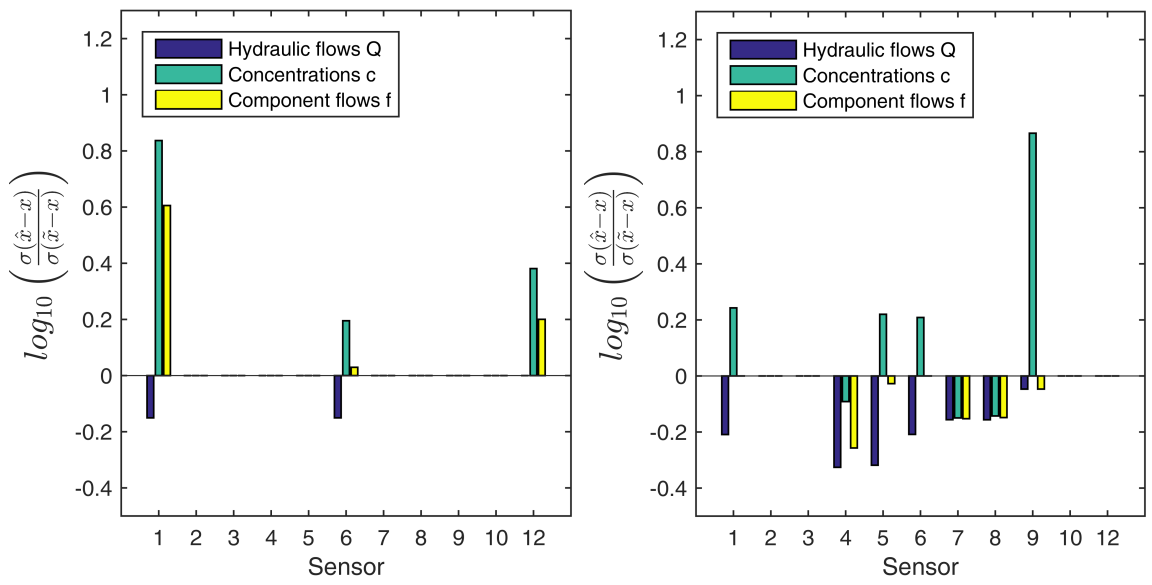


Figure B20: Proportion between the standard deviations of the errors of the reconciled estimates and the measurements for the models A (left) and B (right) with a reduced number of sensors.

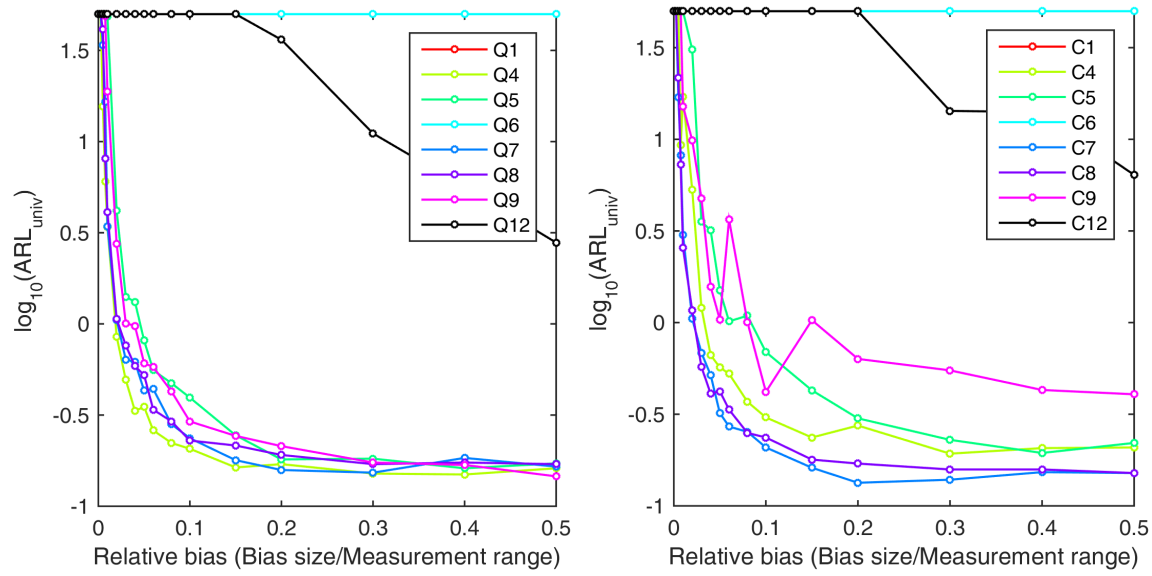


Figure B21: Average biased run lengths (ARL_μ) of the univariate CUSUM chart for the reduced sensor layout of model B depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

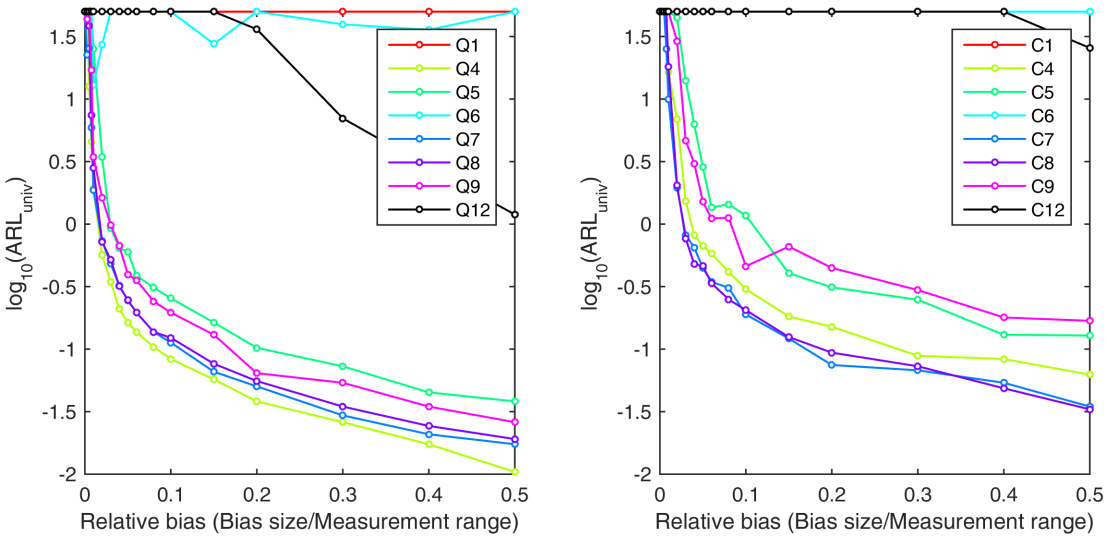


Figure B22: Average biased run lengths (ARL_μ) of the multivariate CUSUM chart for the reduced sensor layout of model B depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).

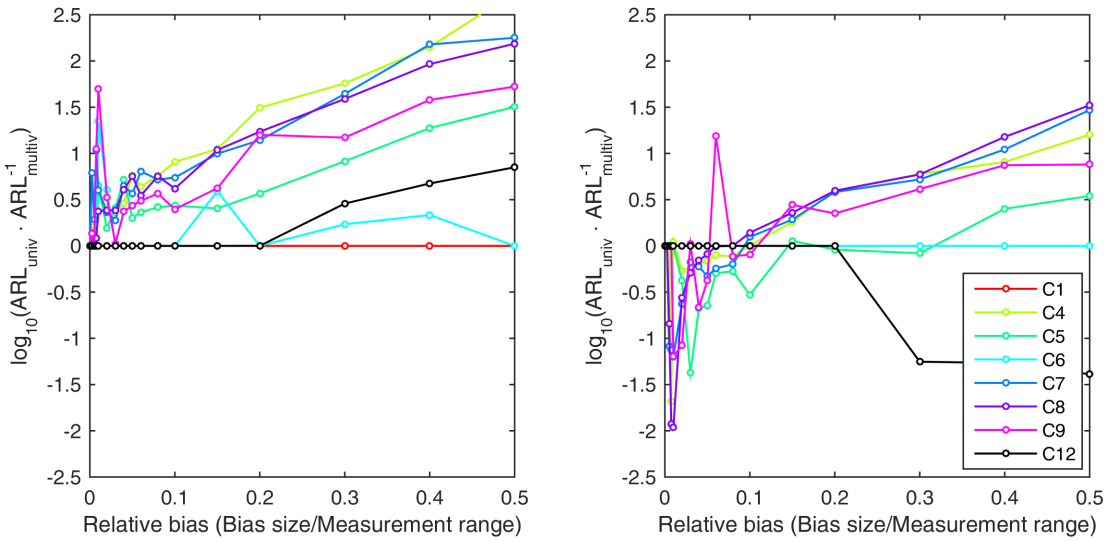


Figure B23: Ratio of the average biased run lengths (ARL_μ) of the univariate and the multivariate CUSUM chart for the reduced sensor layout of model B depending on the bias size in each of the hydraulic flow (left) and concentration sensors (right).