



Anomaly Analysis in Cleaning-in-Place Operations of an Industrial Brewery Fermenter

Yang, Jifeng; Jensen, Bo B.B.; Nordkvist, Mikkel; Rasmussen, Peter; pedersen, Bjarne; Kokholm, Anders; Jensen, Lars; Gernaey, Krist V.; Krühne, Ulrich

Published in:
Industrial & Engineering Chemistry Research

Link to article, DOI:
[10.1021/acs.iecr.8b02417](https://doi.org/10.1021/acs.iecr.8b02417)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Yang, J., Jensen, B. B. B., Nordkvist, M., Rasmussen, P., pedersen, B., Kokholm, A., Jensen, L., Gernaey, K. V., & Krühne, U. (2018). Anomaly Analysis in Cleaning-in-Place Operations of an Industrial Brewery Fermenter. *Industrial & Engineering Chemistry Research*, 57(38), 12871-12883. <https://doi.org/10.1021/acs.iecr.8b02417>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Anomaly Analysis in Cleaning-in-Place Operations of an Industrial Brewery Fermenter

Jifeng Yang, Bo B.B. Jensen, Mikkel Nordkvist, Peter Rasmussen, bjarne pedersen, Anders Kokholm, Lars Jensen, Krist V. Gernaey, and Ulrich Krühne

Ind. Eng. Chem. Res., **Just Accepted Manuscript** • DOI: 10.1021/acs.iecr.8b02417 • Publication Date (Web): 29 Aug 2018

Downloaded from <http://pubs.acs.org> on September 3, 2018

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.



Anomaly Analysis in Cleaning-in-Place Operations of an Industrial Brewery Fermenter

Jifeng Yang[†], Bo B.B. Jensen[‡], Mikkel Nordkvist[‡], Peter Rasmussen[§], Bjarne Pedersen[§], Anders Kokholm[§], Lars Jensen[⊥], Krist V. Gernaey[†], Ulrich Krühne^{†}*

[†] Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[‡] Alfa Laval Copenhagen A/S, Maskinvej 5, 2860 Søborg, Denmark

[§] Carlsberg Danmark A/S, Vestre Ringvej 111, 7000 Fredericia, Denmark

[⊥] Ecolab Denmark A/S, Høffdingsvej 36, 2500 Valby, Denmark

KEYWORDS

Cleaning-in-Place, Anomaly detection, Fermenter cleaning, Multivariate analysis, Data analysis

ABSTRACT

Analyzing historical data of industrial Cleaning-in-place (CIP) operations is essential to avoid potential operation failures, but is usually not done. This paper presents a three-level **approach** of

1
2
3 analysis based on the CIP case of a brewery fermenter to describe how to analyze the historical
4 data in steps for detecting anomalies. In the first level, the system is assessed before cleaning to
5 ensure that the selected recipe and system are able to accomplish the task. In the second level, a
6 multiway principal component analysis (MPCA) algorithm is applied to monitor the process
7 variables on-line or post cleaning, with the purpose of locally detecting the anomalies and
8 explaining the potential causes of the anomalous event. The third level analysis is performed
9 after cleaning to evaluate the cleaning results. The implementation of the analysis **approach** has
10 significant potential to automatically detect deviations and anomalies in future CIP cycles and to
11 optimize the cleaning process.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1. Introduction

Cleaning-in-Place (CIP) is commonly used in the food processing industries. The concept of CIP is to clean the components of a plant without dismantling or opening the equipment and with little or no manual involvement of the operator^{1,2}. CIP operations can generally be considered as batch processes with a predefined time schedule. The cleaning program is usually automatically controlled and accomplished according to a standard operating procedure (SOP). A cleaning validation guideline was released by the European Hygienic Engineering & Design Group (EHEDG) in 2016, which suggests the minimum records when establishing SOPs for CIP: (1) cleaning frequency and non-production times; (2) water quality; (3) cleaning and disinfecting agents; (4) cleaning process parameters (the equipment used, the concentration of chemicals, as well as the cleaning time, temperature, pressure and flow rate) and their recordings; (5) responsibilities and qualifications of the persons involved; and (6) monitoring of the cleaning process including calibration of the sensors³. Therefore, the documentation and interpretation of data play a critical role in the implementation of SOPs for CIP.

Modern CIP systems are normally equipped with a supervisory control and data acquisition (SCADA) or similar system. The monitoring variables typically include time, flow rate, temperature, conductivity, etc., with the purpose to ensure that the critical control points (CCPs) are obtained⁴. Even though the industrial processing plants have adopted automated control and most of the repeated CIPs can be performed in a similar profile, the risk of equipment failure or operation fault can still happen due to the complex process layout and a multitude of pumps and valves. Failure of CIP or faults in CIP procedures are costly, and potentially catastrophic to public health and operational safety^{5,6}. The companies usually have their own trouble-shooting guidelines in the event of a problem. For example, if the detergent strength is below the

1
2
3 requirement, the operator tends to stop the ongoing cleaning, take actions and restart the cleaning
4
5 program after solving the problem. Sometimes, the problem is not detected until the whole
6
7 cleaning procedure has been finished, meaning the CIP cycle needs to be repeated in order to
8
9 gain complete confidence in the results. In general, anomalous indicators appear prior to the
10
11 occurrence of failure or faults. The early detection of anomalies would allow to reduce the
12
13 cleaning downtime, water use and energy consumption.
14
15

16
17
18 Cleaning generates considerable historical records of data. One potential application of the
19
20 data is to establish forecasting models to identify the likely future CIP performance by
21
22 comparing the current observations with the previously stored experiences ⁷. Another application
23
24 is to evaluate and optimize the system by examining the current parameters or process. For
25
26 example, an expert could provide comments for improving cleaning efficiency or reducing
27
28 cleaning costs by checking if the cleaning duration or flow rate is insufficient or over demand.
29
30 However, this kind of analysis is generally performed manually by external experts, and not very
31
32 frequently (e.g. once per year). Industry users expect a more efficient and automatic approach to
33
34 evaluate a cleaning operation and detect the abnormal event right after cleaning or even
35
36 immediately when it occurs.
37
38
39
40

41
42 The current application of analyzing historical cleaning data is still experience-based, **which**
43
44 **will remain in the near future**. This study aims to provide an example showing a possible
45
46 approach to use the existing historical data to learn for future operations. The aforementioned
47
48 method to analyze the monitoring data has the objective to focus on only one parameter at a time,
49
50 which is called a univariate statistical approach ⁸. For complex processes, the method results in
51
52 large datasets containing many interacting variables. Therefore multivariate methods become of
53
54
55
56
57
58
59
60

growing interest, which allows monitoring and analyzing multiple variables simultaneously⁹. Both methods will be introduced and developed in the proposed study with different purposes.

This study introduces a three-level **approach** to analyze CIP monitoring data based on a case study from the brewery industry. The objective of the analysis is to provide a benchmark to evaluate the performance of the CIP system **in the cleaning of a brewery fermenter**, and detect and diagnose anomalies based on historical measurements and monitoring data.

2. Description of an industrial CIP case **in a brewery**

The presented research was performed based on a case study on fermenter cleaning at a brewery (Carlsberg, Fredericia, Denmark). The fermenter, mainly soiled by yeast, was cleaned by a rotary jet head device (Alfa Laval Denmark). There were 27 CIP operations executed for the same fermenter over a period of two years, allowing an analysis of this dataset.

The CIP procedure is described in Table 1. The collected data can be classified into two groups: quantitative and qualitative data. Quantitative data are measured on a numeric or quantitative scale. Qualitative data refer to categorical or discrete variables¹⁰.

Table 1. CIP procedure of the analyzed brewery fermenter (reproduced with permission of Carlsberg, Fredericia, Denmark)

No.	Steps	Approx. time (Unit: minute)
1	Alkaline without circulation	20
2	Alkaline circulation	35
3	Intermediate rinse	6

4	Acid circulation	20
5	Intermediate rinse	9
6	Disinfectant circulation	30
7	Final rinse	17
<hr/>		
Total	137 minutes (plus draining and pausing)	
<hr/>		

The cleaning process is monitored by a CIP diagnosis system. The available data are shown in Table 2. The challenges in dealing with the data include:

- The feed pump stops and restarts, or the flow rate varies during the steps. Such a dynamic flow pattern helps to accumulate liquid for cleaning the bottom pipe elements. Moreover, the generated burst effect when increasing the pump flow rate improves the cleaning performance and saves cleaning chemicals by allowing a holding time that is used to penetrate the soil and react¹¹. However, fluctuations of flow rates appear when the pump status is changed.
- The overall cleaning time is documented. However, the step time and interval need to be assessed manually. It is required to recognize the step time based on both valve and pump status. The step is considered as “start” when both the valve and pump are “on”. Meanwhile, the step is considered as “stop” when either the valve or pump is “off”.
- Even though the same recipe is used, the real step times are different on a minute scale from batch to batch. Such a non-uniform behavior can be caused by many situations, like the delay to reach the CCP, liquid filling in the transfer lines, pump failure, etc.

- There exist both quantitative and qualitative variables. All of the data are exported as discrete datasets. For the quantitative data, different variables are recorded with their own time series. The size of the database, time indices and intervals are different between each other.
- The cleaning is based on time. There is no measurement related to the cleaning quality. Therefore, the cleaning time is normally set over demand, assuming that the most difficult part of the system to cleaning can also be cleaned. This is a common problem for most of the industrial CIP operations.
- Not all modifications in the past two years are recorded. Some data related to the process performance are unknown, e.g. the tank liquid volume and the pump power.

Table 2. The monitoring process data by the CIP diagnosis system (reproduced with permission of Carlsberg, Fredericia, Denmark)

Type of data	Names	Units	Record frequency	Notes
Quantitative	Feed flow rate	m^3/h	ca. every second	
	Return flow rate	m^3/h	ca. every second	
	Conductivity in feed flow	mS/cm	ca. every second	
	Conductivity in return flow	mS/cm	ca. every second	
	Pump outlet pressure	bar	ca. every second	
Qualitative	Date and time	Dimensionless	Once per cleaning	

Recipe number	Dimensionless	Once per cleaning	
Pump status	Dimensionless	Once per change	Binary data (On and off)
Valve status	Dimensionless	Once per change	Binary data (On and off), including the chemical valves, drain valve and circulation valve

3. Three-level **approach** for anomaly detection and diagnosis

The CIP operation can be characterized as a batch process. Each batch also includes several steps, which are assumed to be independent of each other. A three-level **approach** of analysis was designed for the data analysis, forming the main highlight of the article.

The first level, the prior-to-batch analysis, was performed before the cleaning operation. The purpose was to ensure that the cleaning requirements could be achieved. The analyzed factors included but were not limited to:

- The interval to the last production or cleaning, to assess the soil age or if the cleaning was repeated;
- The recipe number, to ensure the right program was selected;
- The off-line measurement of chemical samples (i.e. pH and conductivity), to ensure the chemical was active;
- The accumulated equipment errors, to ensure that the system was effective.

1
2
3 In this study, the cleaning was assumed to be performed at the right time, meaning the cleaning
4 interval was correct. The procedure selection (recipe number) was confirmed to be correct as
5 well. The chemical solutions used for cleaning were measured and checked to be effective
6 enough for the cleaning tasks. Therefore, only the performance of the feed pump was analyzed,
7 which was one of the most important factors for a successful cleaning.
8
9

10
11
12
13
14
15 The second level, the process monitoring (i.e. flow rate, pressure, conductivity, temperature),
16 was to ensure that the cleaning program was under a state of control. Monitoring the process
17 variables includes end-of-batch monitoring and real-time on-line monitoring. The end-of-batch
18 monitoring was carried out when the whole trajectories were available. The real-time on-line
19 monitoring was the ultimate goal of the process monitoring, which provided the process
20 information while the deviation took place and served the real-time quality control ¹². In both
21 monitoring **methods**, a multivariate analysis approach was applied by simultaneously monitoring
22 the pressure, flow rate and conductivity measurements. It was optional to enable the recording of
23 more parameters into the model if the data were available, e.g. temperature.
24
25
26
27
28
29
30
31
32
33
34
35
36

37 The third level, the end-of-batch performance evaluation, was made after the cleaning
38 operation. A standard procedure was defined based on good batches (defined later), in terms of
39 the cleaning step time, water and chemical consumption, pump and valve status, etc. The purpose
40 of this level was to find if the cleaning operation was performed consistently. After cleaning, the
41 operator could compare the cleaning time or consumption with the reference and evaluate if the
42 cleaning was finished normally or abnormally.
43
44
45
46
47
48
49
50

51 The three-level **approach** only indicates the existence of anomalous situations and the potential
52 causes. In the current stage, the operators should be trained to identify the root cause of an
53
54
55
56
57
58
59
60

1
2
3 anomaly and solve the problem. Ideally, the anomalies can be automatically detected and solved
4
5 through the integration of the proposed **approach** with the available control systems, which,
6
7 however, is not a topic of this study.
8
9

10 11 4. Methods

12 13 14 4.1 Feed pump analysis

15
16
17 The feed pump output was dynamically adjusted during the cleaning period. There are short
18
19 periods when the flow rate was kept stable at the required level, though. Therefore, we captured
20
21 the working flow rate when the pump was operated at a static pressure and the startup transition
22
23 time to reach the steady state. Both characteristics form indicators of the pump performance. The
24
25 two-year results were divided into four groups with data collected during each half year as a
26
27 group. The average flow rate and transition time in each group were calculated and plotted
28
29 together for analysis.
30
31

32 33 34 4.2 Multivariate Monitoring Analysis

35
36
37 The multiway principal component analysis (MPCA) method is a common batch process
38
39 monitoring technique extending the ordinary principal component analysis (PCA) to batch
40
41 processes. It is equivalent to performing PCA on a large two-dimensional matrix constructed by
42
43 unfolding the three-way array. The implementation of MPCA or PCA requires a complex data
44
45 pre-processing treatment. In this work, MPCA was applied and the procedure was described in
46
47 Figure 1. All calculations were carried out in Matlab (The Mathworks, Inc., Natick, MA, USA).
48
49
50
51
52
53
54
55
56
57
58
59
60

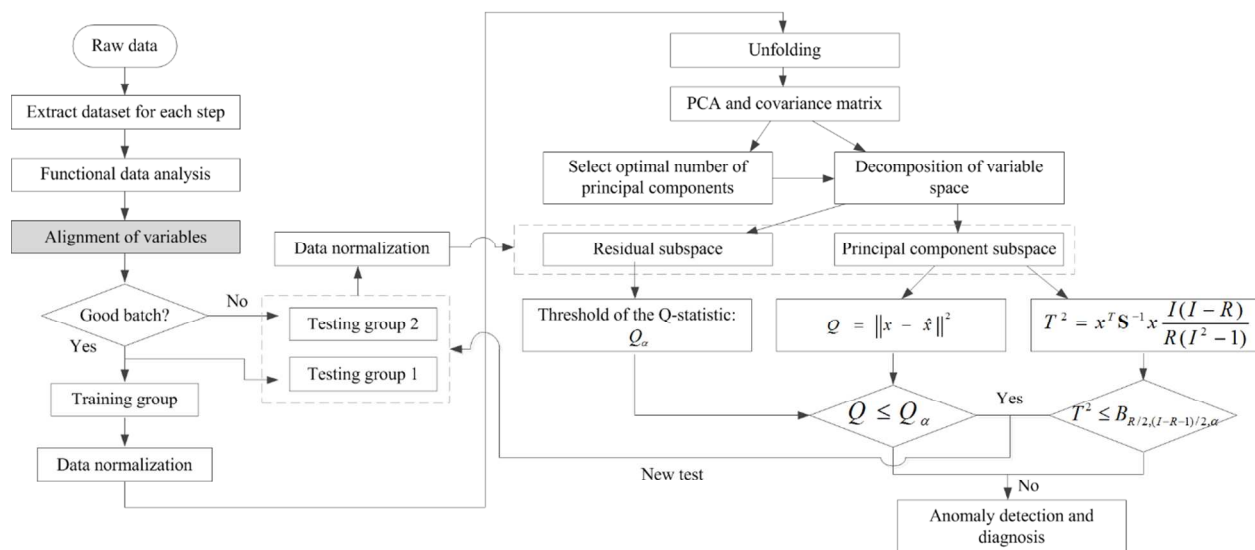


Figure 1. Proposed methodology for multivariate analysis in the level 2 analysis. The step “Alignment of variables” is eliminated when performing non-synchronization analysis.

4.2.1 Functional data analysis

Each cleaning step was analyzed separately. The start and end points of each step were manually determined based on the qualitative pump or valve status data. The starting time of a cleaning step was defined as time zero. In this level, five quantitative variables (Table 2) were analyzed. Due to the uneven length of the available time series for each measurement, the first treatment of the data was to project all measurement variables into uniform time indices and intervals, as illustrated in Figure 2. Functional data analysis (FDA) was applied as a method to produce a smooth functional dataset from the discrete dataset by reducing random errors^{13,14}.

The standard form of FDA is the “signal plus noise” model as:

$$\mathbf{y} = \mathbf{x}(\mathbf{t}) + \boldsymbol{\varepsilon} \quad (1)$$

where the observed dataset \mathbf{y} is expressed by the function $\mathbf{x}(\mathbf{t})$ and noise $\boldsymbol{\varepsilon}$. $\mathbf{x}(\mathbf{t})$ represents a linear expansion of a series of basic functions. The best known basic expansions are the Fourier

series for periodic data and the spline basic system for open-ended data¹³. Only the spline basis was used in this work, as shown:

$$x(t) = \sum_i^M c_i \Phi_i(t) \quad (2)$$

$\Phi_i(t)$ are fourth order polynomials with the i^{th} term coefficient c_i . The number of basic functions, was resolved as follows¹⁵:

$$\text{Number of functions} = \text{Knots} + \text{Order} - 2 \quad (3)$$

The knots equaled time trajectories of the measured variable. The coefficients, c_i , for the spline functions were fitted to the data by minimizing the penalized sum of squares (*PENSSE*) with an additional penalty term, $J[x]$, to impose smoothness in the function output. The function's roughness was the integrated squared second derivative $[D^2x(t)]^2$, as shown in equation 5. The smoothing parameter, λ , was defined to be 1, which was close to the optimal value determined by means of the generalized cross-validation (GCV) criterion^{13,15}.

$$\text{PENSSE} = \sum_i^n (y_i - x(t_i))^2 + \lambda \cdot J[x] \quad (4)$$

$$J[x] = \int [D^2x(t)]^2 dt \quad (5)$$

In order to unify the time trajectories for all variables, a new time vector with one second interval was defined. The values of different variables in the new time series were evaluated from the smoothed function.

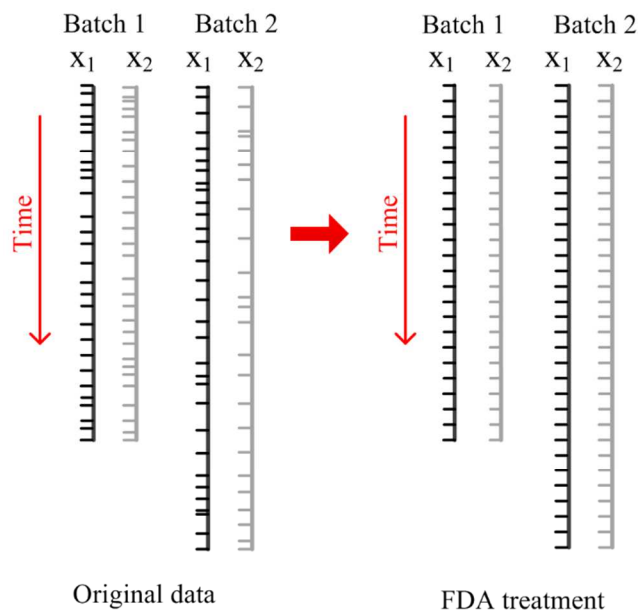


Figure 2. Illustration of unifying the time stamps of different variables by the means of FDA.

The new dataset after FDA treatment has a one second time interval

4.2.2 Alignment of variable trajectories

FDA can be used to align different variables within a single batch. However, following FDA, the unequal and unaligned data across batches still exist because the original data were collected in independent length. Three approaches were tested to solve the problem, and the results are compared below (see Figure 5).

The first and the simplest method was to cut all batches to the minimum length or a specific length. This approach kept the nature of batches, at the expense of losing all information available in the data after the cutting point for the time series that had to be cut. However, the shifts of variable profiles caused by the operation of the dynamic pump could still lead to the challenges of data analysis.

1
2
3 The second method was the indicator variable technique, the concept of which was to replace
4 time by another process or derived variable to indicate the progress of the batch ¹⁶. In this study,
5 liquid volumetric consumption, calculated from time and flow rate data, was chosen as the
6 indicator variable to quantify the percent completion of each batch. Linear interpolation was used
7 to transform the time dimension into the indicator variable dimension.
8
9
10
11
12
13
14

15 Another technique to solve this problem was to locally translate, compress and expand the
16 pattern of one sequence to match another sequence of similar features, called dynamic time
17 warping (DTW) ¹⁷. DTW nonlinearly warps two data series in such a way that similar events are
18 aligned and the distance between them is minimized ¹². In this study, pressure was an active
19 variable, whose variance influenced the patterns of other variables. Therefore, the warping path
20 was obtained by aligning the pressure measurements of all batches to a selected reference batch
21 demonstrating good performance, and other variables were aligned in the same warping path.
22
23
24
25
26
27
28
29
30
31

32 4.2.3 Training and testing datasets

33
34
35 The CIP batches for which data were available were labeled as “good” or “bad” operations. A
36 CIP batch was regarded as “bad” when possessing at least one of the following characteristics:
37
38
39
40

- 41 • The pump stopped for a long time during a step, which was beyond the normal stop
42 period;
- 43 • There were more or less pump stops than foreseen in the normal operation procedure;
- 44 • The cleaning step was missing, or the step time was too long or too short compared
45 with other batches;
- 46 • The ionic properties of the chemical solutions were abnormal;
- 47
48
49
50
51
52
53
54
55
56
57
58
59
60

- Unreasonable observations of flow rates, i.e. zero flow rate due to centrifugal pump idling.

Ca. 15 good batches, the number of which varied when analyzing different cleaning steps, were used as training dataset for constructing the reference models. The rest, consisting of both good and bad batches, were used as testing dataset for validating the model.

4.2.4 Data unfolding and normalization

The functional data were provided as a three-dimensional matrix with I batches, J time indices and K variables. The MPCA approach requires that the 3-D matrix is unfolded to a 2-D matrix. There are normally two ways to unfold the batch data: a batch-wise unfolding results in an $(I \times JK)$ matrix, and a variable-wise unfolding yields an $(IJ \times K)$ matrix. Most of the statistical process monitoring techniques apply the batch-wise unfolding approach, allowing one to analyze the variability among batches with respect to variables and their time variation. However, when being used on-line, it is required to estimate the future portions of variable trajectories. On the contrary, the variable-wise unfolding does not require such estimation, making it applicable for on-line monitoring^{9,18,19}. Wold *et al.*²⁰ concluded that the former was oriented to evaluating the completed batch as normal or abnormal. While the latter focused more on the monitoring of individual time points²⁰. The primary objective of this study was to identify the occurrence of an anomalous event on-line. Therefore, only the variable-wise unfolding method was applied to the data.

The new unfolded matrix was then mean-centered and variance scaled. Mean centering was applied to the raw data by subtracting the average variable values for each time point. Then the variance scaling was carried out to the centered data by a slab-scaling method, by dividing with a

1
2
3 unit root-mean-square of each variable. We noted that other scaling methods, like column-
4 scaling by dividing the root-mean-square of each time point, resulted in different results. In the
5 pre-study, lower variability was explained when scaling with column-scaling, because this
6 method causes significant errors when data were unsynchronized or inadequately synchronized.
7
8 Therefore, only the robust slab-scaling approach was applied in the present study, which was
9
10 also recommended when dealing with an industrial dataset ²¹.
11
12
13
14
15
16

17 4.2.5 Multivariate analysis

18
19
20
21 PCA is a commonly used multivariate statistical analysis method for monitoring purposes,
22 which defines the principal components through a linear transformation of the original variables
23 and explains the maximum variance. Instead of analyzing all the variables, the PCA method
24 focuses only on analyzing the principal components when monitoring the system and thereby
25 reduces the dimensionality of the data ^{18,22}.
26
27
28
29
30
31
32

33 In this work, PCA was applied to the normalized data matrix X , which was decomposed into a
34 score matrix T and a principal loading matrix P .
35
36
37

$$38 X = TP^T \quad (6)$$

39
40
41 Only the eigenvectors, $Q(Q \in \mathfrak{R}^{M \times K}, M < K)$, and P were retained in the PCA model. The
42 number of selected principal components in this model was $M = 3$, which explained
43 approximately 90% of the cumulative percent variance (CPV). The new validation data were
44 normalized in the same way as the training dataset. The estimate of the testing groups of data
45 was done according to:
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

$$\hat{X}_{new} = X_{new}Q^TQ \quad (7)$$

4.2.6 Anomaly detection

Two commonly used indexes to detect anomalies or faults are the Q-statistic¹⁸ and the Hotelling's T^2 statistic^{9,23}. The Q-statistic is computed as the squared sum of residuals⁸:

$$Q - statistic = \|X_{new} - \hat{X}_{new}\|^2 \leq \delta^2 \quad (8)$$

where δ^2 denotes the threshold for the Q-statistic, which is determined by:

$$\delta^2 = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \quad (9)$$

where $\theta_i = \sum_{j=M+1}^K \sigma_j^{2i}$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$ and c_α is the normal deviate corresponding to the $(1 - \alpha)$ percentile. σ_j is the j^{th} eigenvalue. When there is no anomaly, the Q-statistic is less than δ^2 , representing the normal dynamics and noise which cannot be avoided. On the other hand, if an anomaly or fault exists, the Q-statistic is higher than the value of δ^2 .

The Hotelling T^2 statistic is computed as follows¹²:

$$T^2 = \hat{X}_{new} S^{-1} \hat{X}_{new}^T \frac{R(R - K)}{K(R^2 - 1)} \quad (10)$$

where S is the $(K \times K)$ covariance matrix, R is the number of batches in the reference set. The threshold for the Hotelling T^2 statistic is given by:

$$B_{R/2, (I-R-1)/2, \alpha} \sim \frac{(K/(R - K - 1))F_{K, R-K-1, \alpha}}{1 + (K/(R - K - 1))F_{K, R-K-1, \alpha}} \quad (11)$$

where F is the F-distribution value with R and $I - R - 1$ degrees of freedom at the α significance level. The interpretation of the T^2 statistic is similar as the Q-statistic, i.e. the out-of-threshold signal denotes an abnormal event.

4.2.7 Anomaly diagnosis

The diagnosis of a detected anomaly can be performed by a univariate method to analyze the monitoring values one by one, or by a multivariate method, i.e. the contribution plot. Contribution plots indicate the contributions of each variable to the deviations from normal operations. For the Q-statistic, the contribution of variable k at time j is calculated from the residual, e , as follows^{12,24}:

$$C_{Q,jk} = e_{jk}^2 \quad (12)$$

For the Hotelling T^2 , the contribution of each variable to the T^2 statistic is given by *Miller et al.*²⁴:

$$C_{T^2,jk} = \sum_{m=1}^M x_{jk} p_{km} \quad (13)$$

where p_{km} is the loading for variable k in dimension m .

4.3 End-of-batch performance evaluation

For level 3 of the **approach**, a standard range was defined for the cleaning time of each step, liquid consumption or on-off ratio (the period when the pump is “on” divided by the whole step time) on the basis of some good batches. The standard was made for each cleaning step, including the means, standard deviations (STD) and a parameter γ , to allow a judgement range as described in equation 14.

$$Reference = Mean \pm \gamma \times STD \quad (14)$$

The testing groups of data were used for validation, which included both good and bad batches (defined in section 4.2.3). During the test, the cleaning was regarded as “normal” if the indexes were within the reference range. On the contrary, the cleaning was “abnormal” if the indexes were out of the range. However, wrong judgement might occur, i.e. that a good operation was denoted as bad or a bad operation was denoted as good. Therefore, there are four possible outcomes²⁵:

- True positive (TP) correctly recognizes the good operation as “good”;
- True negative (TN) correctly recognizes the bad operation as “bad”;
- False positive (FP) wrongly recognizes the bad operation as “good”;
- False negative (FN) wrongly recognizes the good operation as “bad”.

The judgement decision depends on the value of γ . A new index, η , was defined to quantify the ratio of the correct judgements over the total number of cleaning batches. A high value of η represents a high probability to make the right decision.

$$\eta = 1 - \frac{FP + FN}{All} \quad (15)$$

5. Results and discussions

5.1 Pre-cleaning: Performance of the feed pump

Local wall shear stress is an important flow parameter for the removal of single species bacterial biofilm and spores from surfaces. The value of wall shear stress can be determined by the liquid properties, velocity gradient or flow rate^{26,27}. A critical wall shear stress value was

found when cleaning a radial flow cell according to the EHEDG-based test method²⁸. The value of the critical wall shear stress varies when investigating pipes or tanks. At least, the cleaning may not be achieved or be costly once the criterion is not achieved, especially in the dead areas where the flow velocity is relatively lower than the bulk flow. It is estimated that reducing the flow rate by 5% leads to nearly a 10% decrease of local wall shear stress on a flat plane when flows are turbulent²⁸. As illustrated in Figure 3, the startup time for the feed pump to reach the working flow rate has not changed significantly ($p = 0.23$) in the past two years, but the static flow rate reveals a clearly decreasing trend ($p < 0.01$). This decrease can result from the mechanical variance of the pump, leaking of liquid, fouling of the transfer pipe, low level of liquid in the chemical tank, calibration of sensors, etc. Since the variance of flow rate is a long-term performance metric, the occasional faults can be ignored, like the leaking problem and the liquid level. However, it is still difficult to find an exact reason for the anomaly due to the limited number of measurements.

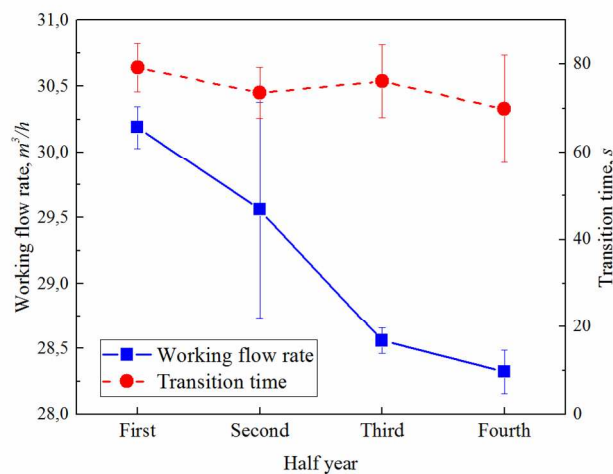


Figure 3. Working flow rate and startup transition time of the feed pump over the past two years.

1
2
3 For the pressure-controlled pump, as in this case, the differential pressure (the difference
4 between the suction and discharge heads) is critical. The outlet pressure is controlled to be
5 constant. However, if there is fouling in the suction line, the inlet pressure drops and the
6 differential pressure increases, and consequently the flow rate drops. The fouling in the
7 discharging line (which is to be cleaned) can increase the system friction and reduce the flow rate
8 as well ²⁹. Besides, the cavitation of the pump and the change of impeller clearance can also
9 influence the pump performance. Therefore, the observation in Figure 3 reminds the operators to
10 perform a necessary inspection of the pump or of the whole system. There is a risk of failure
11 once the flow rate is less than the required value.
12
13
14
15
16
17
18
19
20
21
22
23

24 The above-mentioned analysis can be accomplished automatically if the system can extract the
25 flow rate information of previous cleanings by coupling the pump status and output, thereby
26 displaying the performance of the feed pump before a new operation starts. A similar analysis
27 can be applied to other systems besides the feed pump. For example, checking heat exchangers'
28 working temperature and heating time to reach the temperature can characterize (though
29 irrelevant in this study) the performance of heat exchangers. Keeping the whole system in high
30 performance is a prerequisite to avoid a potential cleaning failure.
31
32
33
34
35
36
37
38
39
40

41 5.2 Cleaning or post cleaning: Data pretreatment and MPCA model

42
43
44

45 The multivariate analysis was performed for each cleaning step separately. An example of the
46 data pre-processing results for the alkaline treatment without circulation step is shown in Figure
47 4.
48
49
50
51
52
53
54
55
56
57
58
59
60

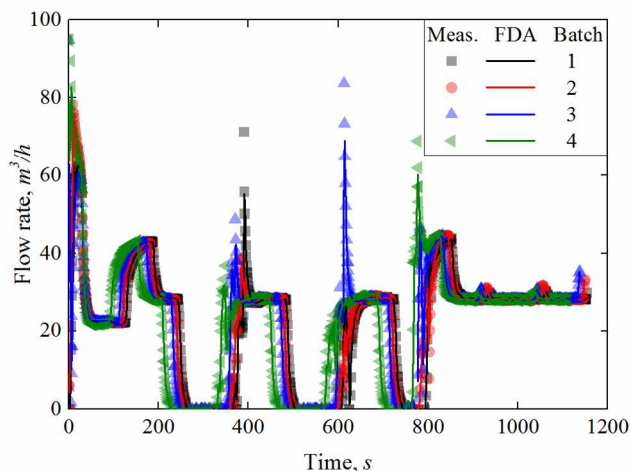


Figure 4. The raw discrete data (symbols) and the FDA results (curves) of the alkaline treatment step without circulation. The roughness penalty coefficient in FDA, λ , is 1.

The smoothness is controlled by the penalty coefficient, λ , in equation 4. For the alkaline step, the optimal λ determined by the GCV criterion varies from batch to batch and variable to variable, most of which lie in the range of $10^{-1} \sim 10^1$. Choosing a small value of λ , the function fits the raw observations better but the smoothness becomes less. On the contrary, as λ becomes larger, the function places more emphasis on the smoothness and less on fitting the observations. However, in this study, the influence of changing λ between 0.1 and 10 is visually small, because the measurement noise is not significant. Therefore, a fixed value of 1 was selected for all computations.

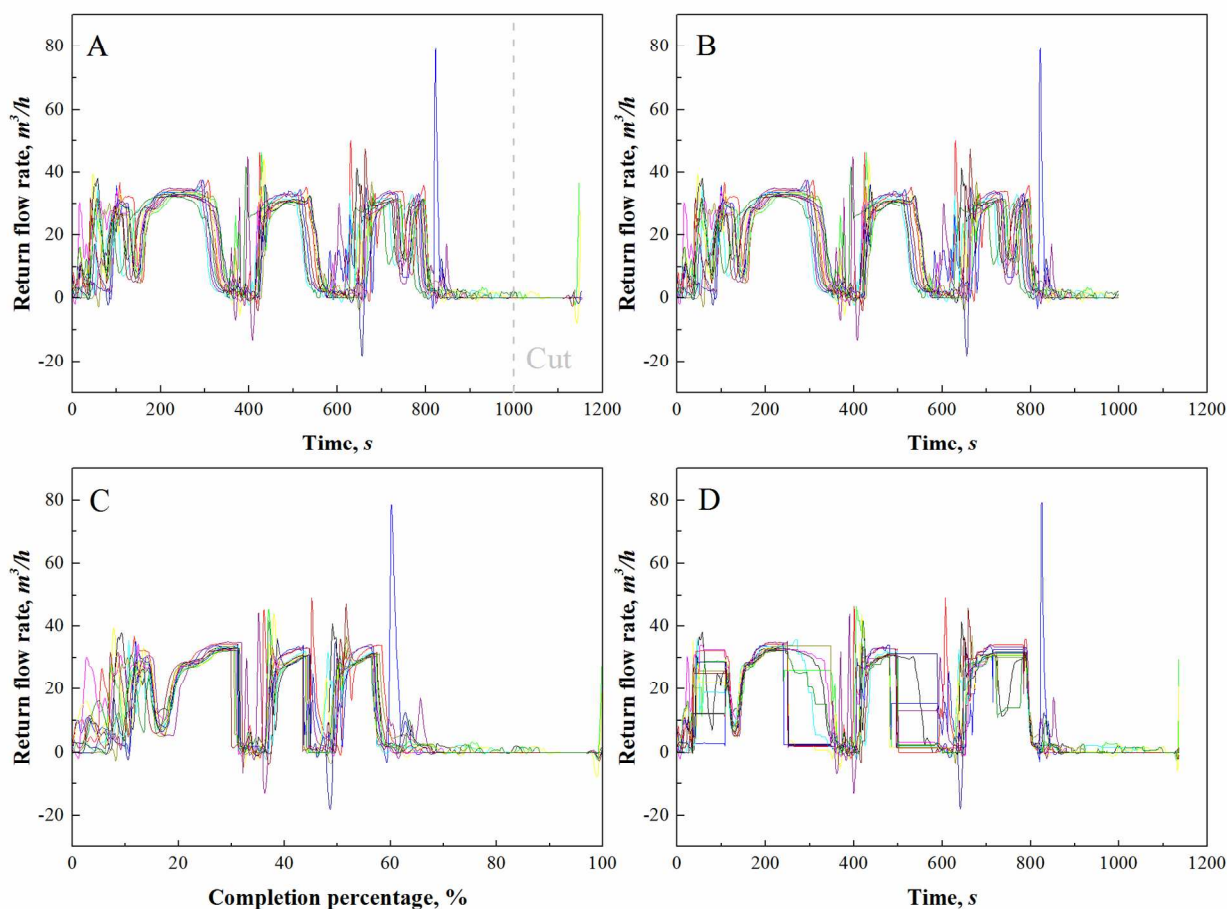


Figure 5. An example of the alignment of variable trajectories for the alkaline treatment step without circulation. (A) is the original measurement. (B) is obtained by cutting the data after 1000 s. (C) is based on the indicator variable technique. (D) is based on DTW.

After FDA treatment, similar curves are obtained for different batches. However, there is still a shift or delay of the shapes, as displayed in Figure 5A. Without trajectory alignment, it is necessary to keep different batches of the same length, for example by cutting the remaining measurements after a specific time point (Figure 5B). However, this kind of treatment eliminates information after the cutting point, meaning the on-line monitoring loses references when a new batch has not completed. On the other hand, the shape differences can be reduced by applying an

1
2
3 indicator variable or time warping the data, as displayed in Figure 5C and D. Both techniques
4 retain the whole dynamic trajectory of the data. Nevertheless, variances and fluctuations still
5 exist, especially when the pump starts. This transition startup problem will further lead to large
6 variances at these time points.
7
8
9
10
11
12

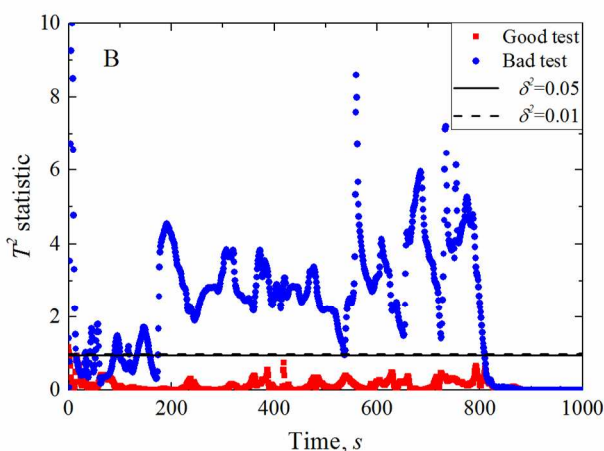
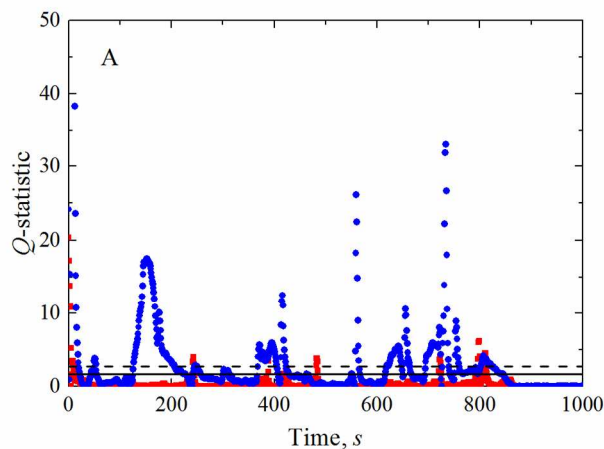
13 PCA was performed to the normalized and unfolded data matrix. According to *Lehman et al.*
14³⁰, the selection of the number of PCs should explain at least 80% of CPV. The final model uses
15 three PCs to meet this criterion.
16
17
18
19
20

21 5.3 Post cleaning: End-of-batch anomaly detection and diagnosis based on the multivariate 22 analysis 23 24 25

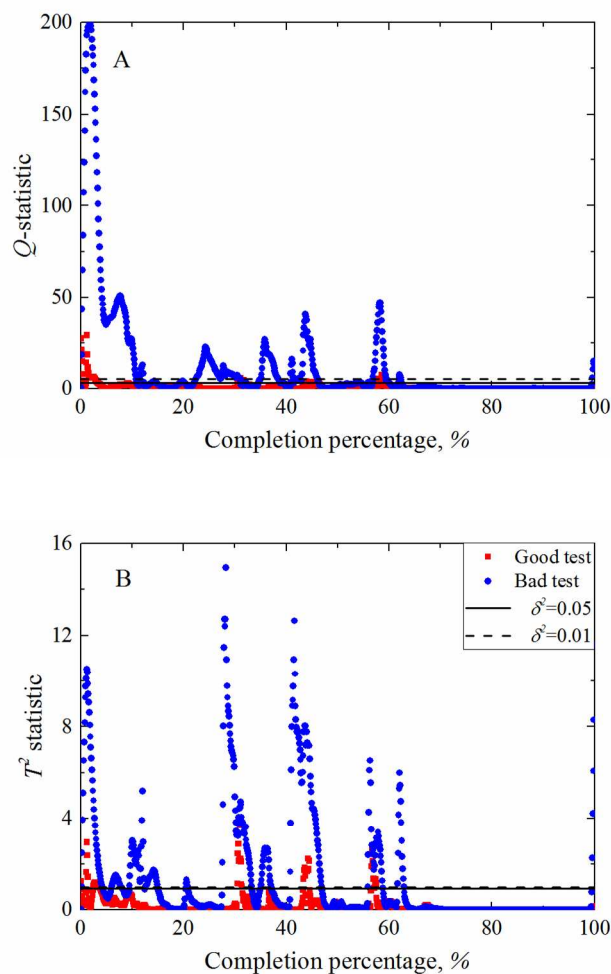
26 A good and a bad batch are introduced respectively for showing the validation of the MPCA
27 model. Figure 6 - 8 display the off-line anomaly detection with the trajectories either unaligned
28 or aligned by two methods. Since there might be spike disturbances from measurements, it
29 makes no sense to recognize an anomalous event when only one or a few samples jump above
30 the threshold²³. So it is more reasonable to determine an anomaly when a continuous series of
31 points jump outside of the control region. Increasing the significance level, α , lifts the detection
32 threshold¹².
33
34
35
36
37
38
39
40
41
42

43 For both Q and T^2 statistics, there are none or only a few anomaly alarms triggered for the
44 good test batch. Some out-of-threshold points at the beginning of cleaning can be mostly caused
45 by the various initial states in different batches, like product or liquid filling in pipes. If
46 neglecting some short-term out-of-threshold observations, all models successfully avoid wrong
47 detection by recognizing a normal situation as abnormal. On the contrary, the detection of the
48 bad test batch is very sensitive to the selected statistical method and alignment approach. The
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Hotelling T^2 statistic triggers more alarms than the Q-statistic, meaning the former is more
4 sensitive to the anomalous event. The large peaks of values are the results of dynamic pump
5 profiles. Applying trajectory alignment can increase the robustness of the model, and that is
6 because after alignment, all batches are adjusted to be as close as possible to the reference.
7 Therefore, the time shift between batches, which causes the firm alarms in Figure 6B, becomes
8 no longer the source of anomalies. Aligning the data by indicator variable is more applicable for
9 detecting a sudden variation. For the alignment by DTW, however, a long-term anomaly occurs
10 as a gradual change.
11
12
13
14
15
16
17
18
19
20
21
22



1
2
3 **Figure 6.** The anomaly detection plots by (A) the Q -statistic and (B) the Hotelling's T^2 statistic
4 with the unaligned trajectory discarding the part of the dataset that was cut off. The example is
5
6 based on the alkaline treatment step without circulation.
7
8
9



42 **Figure 7.** The anomaly detection plots by (A) the Q -statistic and (B) the Hotelling's T^2 statistic
43 with the trajectory aligned by indicator variable. The example is based on the alkaline treatment
44 step without circulation.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

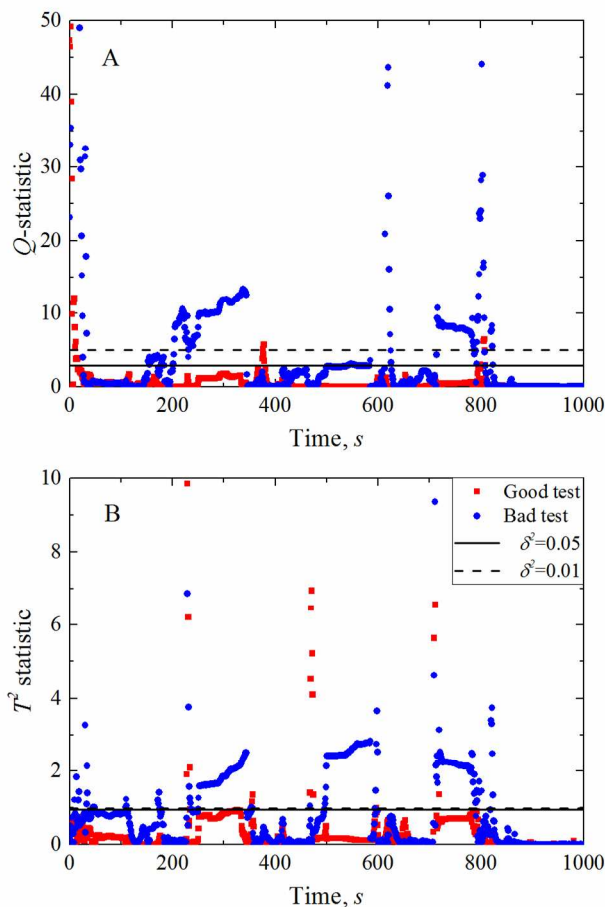


Figure 8. The anomaly detection plots by (A) the Q -statistic and (B) the Hotelling's T^2 statistic with the trajectory aligned by DTW. The example is based on the alkaline treatment step without circulation.

Once an anomaly is found, it is necessary to understand which measurement deviates significantly and why the anomalous event occurs. This can be done by analyzing the measurement values one by one or by using the contribution plot²⁴. However, when there are a large number of measurements or more than one measurement is deviating from the reference, the troubleshooting approach using contribution plots becomes more effective.

Figure 9 and Figure 10 display the contributions of each variable to the Q -statistic and the T^2 statistic for the dataset with trajectories aligned by the indicator variable technique and DTW,

1
2
3 respectively. Both represent the same events at 300 *s* and 400 *s* reference times in Figure 8 when
4 aligning the trajectory using DTW. The two alignment methods are almost identical with regard
5 to diagnosing the observations. At 300 *s*, the anomaly occurs to the bad batch. While at 400 *s*,
6 both batches are under control. The axis scaling of the bad test at 28.3% completion or 300 *s* is
7 greater than the other normal tests in all plots, indicating that the absolute difference of each
8 measurement value in the abnormal state is larger than in the normal states. This anomalous
9 event is mostly caused by the significant deviations of the conductivity measurements in either
10 feed flow or return flow. However, the current analysis cannot figure out the exact causes of the
11 anomaly. But the operator is led to checking the chemical solution to diagnose the problem. A
12 detailed off-line diagnosis of the system confirms that the real cause is due to the extreme dosage
13 of concentrated caustic. Therefore, the troubleshooting approach presented in Figure 9 and
14 Figure 10 indicates the right tracing of the fault. Note that contribution plots suffer from a
15 smearing effect, under which the contribution of a variable is affected by interferences with other
16 variables³¹. As a result, the diagnosis accuracy can sometimes be reduced. Despite some
17 available modified or alternative methods, it is advisable to interpret the diagnosis task based on
18 both contribution plots and univariate isolation, i.e. analyzing the deviation of each variable from
19 its expected value³².

20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

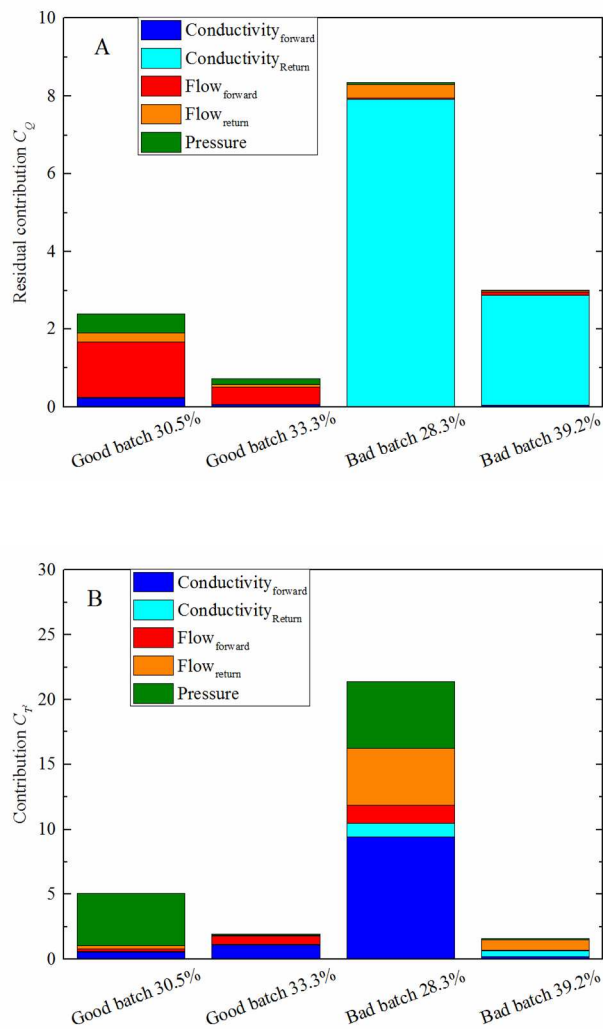


Figure 9. Anomaly diagnosis of aligned dataset by the indicator variable technique. (A) represents the residual contributions to the Q -statistic. (B) represents the absolute contributions to the T^2 statistic. The selected percentages of completions correspond to 300 s and 400 s reference times respectively when aligning the trajectory by DTW. Only “bad batch 28.3 %” is a detected anomaly state. The others are still under the control limit.

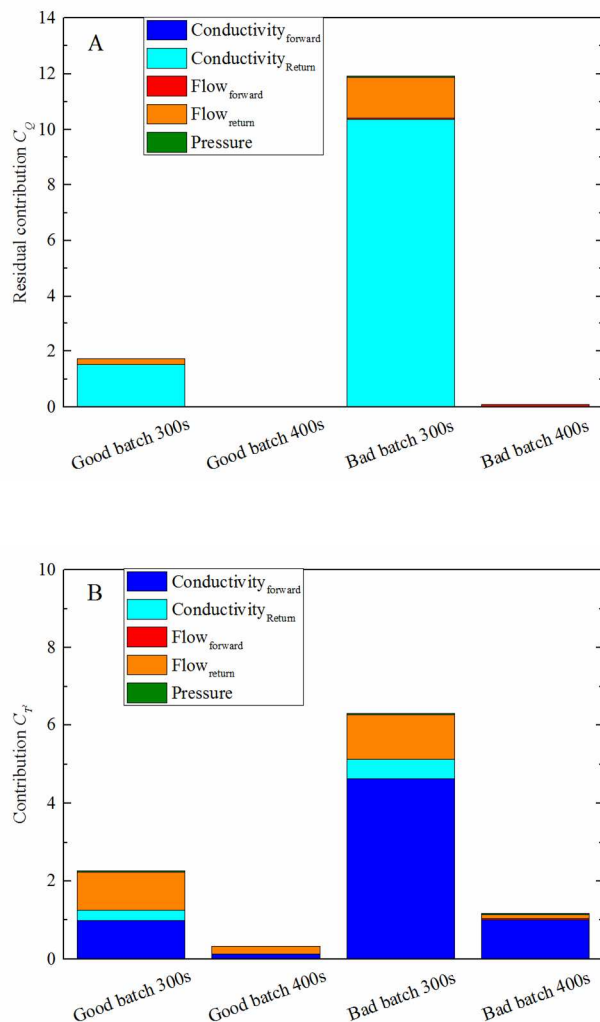


Figure 10. Anomaly diagnosis of aligned dataset by DTW. (A) represents the residual contributions to the Q -statistic. (B) represents the absolute contributions to the T^2 statistic. The selected reference times correspond to the completion percentages in Figure 9 when aligning the trajectory using the indicator variable technique. Only “bad batch 300 s” is a detected anomaly state. The others are still under the control limit.

5.4 Cleaning: On-line anomaly detection and diagnosis based on multivariate analysis

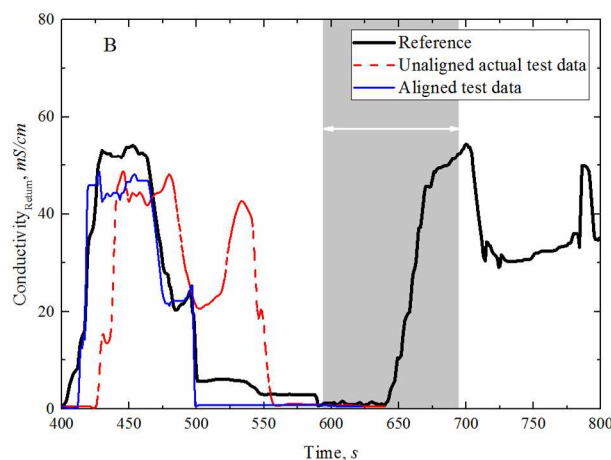
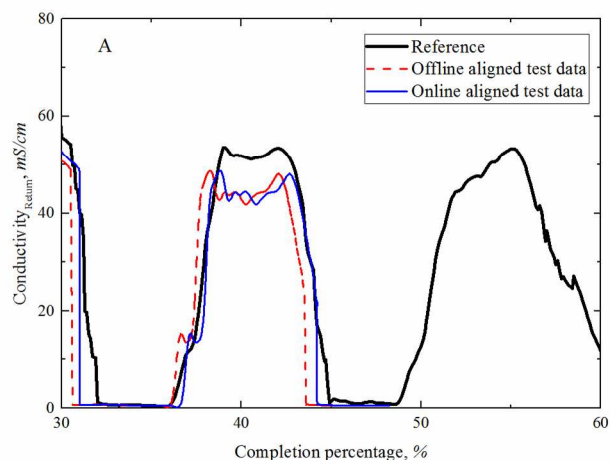
The advantage of variable-wise unfolding is to treat the new monitoring data directly on-line without estimating the future variable measures that are required when unfolding in batch-wise

1
2
3 direction. In order to implement on-line analysis, a coupled measurement system is necessary,
4
5 where all measurements are made in the same time series or can be functionalized to estimate the
6
7 variable values at consistent time points. Without trajectory alignment, the new monitoring
8
9 variables can be directly normalized, unfolded, PCA transformed and analyzed by statistics as
10
11 illustrated in Figure 1. Similar results can be obtained as shown in Figure 6. However, the
12
13 unaligned approach results in high uncertainty once time shifting is significant in the new batch.
14
15 In this regard, trajectory alignment becomes more credible, plus more challenging.
16
17
18
19

20 For the alignment based on the indicator variable, the completion percentage is calculated from
21
22 the consumption of liquid volume divided by the total consumption. However, the total
23
24 consumption of liquid is unknown until the end of the batch. Therefore, a constant value of the
25
26 average consumption in reference batches is used for the new batch. This value can be altered to
27
28 the real consumption when performing end-of-batch analysis. For the alignment using DTW, the
29
30 challenge in on-line monitoring is to pair the current observation time to an endpoint in the
31
32 reference in order to synchronize the new trajectory with the reference timeline. *Kassidas et al.*
33
34 introduced the lower and upper bounds for searching the optimal path that provides the best
35
36 matching between the reference and the uncompleted batch³³. The point resulting in the minimal
37
38 accumulated distance is determined as the endpoint corresponding to the current observation. In
39
40 this study, the alignment path can be obtained by aligning the pressure intensity along the
41
42 timeline. The procedure is one-way incremental and repeated when every new measurement
43
44 comes in.
45
46
47
48
49

50 To explain on-line monitoring, an example is shown in Figure 11 based on functionalized data.
51
52 In Figure 11A, the difference between the on-line and off-line aligned data is due to the fact that
53
54 the real overall consumption may differ from the expected value. However, the deviation only
55
56
57
58
59
60

1
2
3 causes a slight shift of the profile. The alignment based on DTW can also effectively remove the
4 time difference (Figure 11B). However, caution should be taken when defining the searching
5 time difference (Figure 11B). However, caution should be taken when defining the searching
6 bounds of the endpoint, which should lie internally between the lower and upper bound. To sum
7 up, the indicator variable based approach surpasses the DTW method in terms of computation
8 efficiency, because the latter requires accumulative calculations of all possible points in the
9 range in order to obtain the minimal distance. After alignment, the new monitoring variables can
10 be further transformed for anomaly detection, as described in Figure 1.
11
12
13
14
15
16
17
18
19



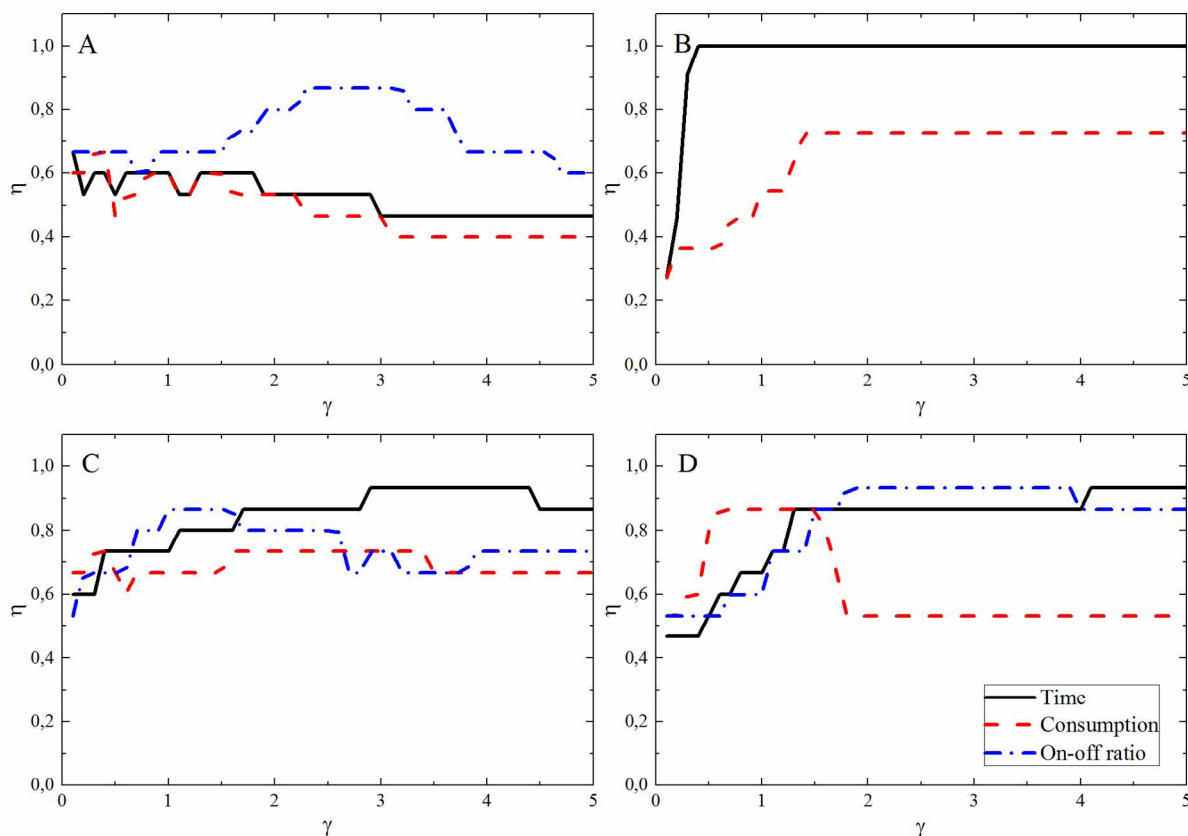
1
2
3 **Figure 11.** An example demonstrating on-line alignment of the variable trajectory based on (A)
4 indicator variable technique and (B) DTW. The current observation time is 640 s, corresponding
5 to 47.6% completion of the batch. The gray region in figure B refers to the searching range of the
6 endpoint with ± 50 s bounds.
7
8
9
10
11

12 13 5.5 Post cleaning: Validation of SOP 14 15

16 Operators would like to evaluate a CIP operation using a simple method, using indicators
17 which can be easily measured or computed, i.e. the step time, the liquid consumption and the on-
18 off ratio of the pump. Therefore, the operators only need to compare these values with the
19 recommended range, and get an idea about whether the cleaning is performed according to the
20 standard range or performance. The standard range is related to the safety margin parameter, γ ,
21 which further influences the judgement of the cleaning performance.
22
23
24
25
26
27
28
29
30

31 As shown in Figure 12, the increase in γ generally improves the correctness of the judgement,
32 as more good cleaning operations are covered in the range. However, if the value of γ is beyond
33 a specific limit, the correctness of the judgement is reduced, because more and more bad
34 cleanings are incorreced considered as good. Therefore, it is recommended to select the value of
35 γ between 2 and 3. Using the step cleaning time or the on-off ratio is more effective than using
36 the liquid consumption to make a decision, where the accuracy is up to 80%. This can be
37 understood by the fact that the current recipe is based on time, and the main anomalies in the CIP
38 system result from the feed pump or accessory valves. Once a low flow rate occurs, the system
39 automatically takes an action by controlling the flow rate at the pre-defined value and adjusting
40 the cleaning time to be long enough. As a result, the cleaning time deviates from the reference
41 range, but the total consumption does not vary significantly. Therefore, the operator is advised to
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 use the cleaning time or on-off ratio of the pump to evaluate the cleaning results. Once there is a
4 rising trend of the number of abnormal events, it is time to perform an overall inspection of the
5
6
7
8 CIP system.
9



10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40 **Figure 12.** The ratio of the correct judgement of SOP against the margin parameter: (A) alkaline
41 step without circulation, (B) alkaline circulation, (C) acid circulation, and (D) disinfectant
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
circulation. There are no pump stops during the alkaline circulation step.

5.6 Practical applications and future perspectives

The three-level approach provides a benchmark in analysis of historical data and evaluation of the CIP system. The above analysis was based on the cleaning of a brewery fermenter. However, the application is not limited to the brewery case. The potential applications could be:

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- *Training of operators*: The operator is the most important person to implement a CIP operation. If the operator gets some simple indexes to assess a CIP cycle (i.e. the step time), a potential failure can be evaluated immediately after cleaning, thereby avoiding potential delays in the production process.
 - *Maintenance of the equipment*: It mainly refers to the pumps and valves (as well as heat exchangers, though not involved in this study). A systematic shift of the observations (i.e. flow rate and transition time) may indicate an accumulated error or damage of the equipment.
 - *Optimization of the system*: A reference has been established based on historical data of good cleaning operations. Optimization could therefore be achieved by altering one parameter at a time and comparing the new results with the reference.
 - *Flexible selection of cleaning recipes*: In some processes where the equipment is used to produce different products, there are normally multiple recipes to remove different soiling materials when altering products. A reference model helps to examine if a complex recipe is selected to clean a loose soil or a simple recipe is selected to clean a harsh soil. Sometimes, a short cleaning run should be selected before production if the equipment is held for long time. Choosing a full recipe in this case will waste considerable money and resources. A reference model helps to reduce waste, because a wrong selection of recipe results in unexpected monitoring outputs when comparing with the reference.
 - *Alarm of system anomaly*: Conventionally, the alarm is triggered when one or more monitoring variables are out of range, meaning that the failure or error already happens.

1
2
3 A multivariate analysis approach can not only capture such an existing problem, but
4 also identify abnormal trends that may suggest a potential failure. Early detection of
5 anomalies reduces the damage of equipment and saves cleaning cost.
6
7
8

- 9
10
11 • *Counting of anomalous events*: End-of-batch analysis can be performed periodically by
12 counting anomalies over a specific period, in order to gain a better understanding of the
13 system.
14
15
16

17
18
19 There are also some considerations associated with the existing study to support an effective
20 application in the future. First, only the two-year dataset of one brewery fermenter is
21 investigated. But the model is implemented in a software. As a benchmark case, the model is
22 open to involve other systems and more datasets. Second, the current system does not consist of
23 any on-line monitoring of the cleaning quality, which is a common problem in most CIP
24 practices. New sensor technology (i.e. turbidity) enables to detect the residues in the liquid
25 phase. Involving the quality indicators for the cleaning operation into the second level analysis
26 will deliver more information and potentially allow to predict the end of the cleaning operation ⁹.
27
28 Finally, an advanced monitoring and control system is required to implement the analysis
29 automatically. Especially, monitoring all variables of the same trajectory and equal time space
30 should allow simplifying the model construction and reducing the uncertainty. An ideal
31 operation of CIP is expected to be implemented in such a way:
32
33
34
35
36
37
38
39
40
41
42
43
44
45

- 46
47 • when clicking the recipe button, the control interface displays the system performance
48 based on previous cleaning operations and off-line measurements, and reports if the
49 selected recipe is able to complete the cleaning task in an efficient and economic
50 manner;
51
52
53
54
55
56
57
58
59
60

- when performing the cleaning, the abnormal detection triggers the alarm immediately when it happens, and the potential causes of anomalies and proposed actions are described;
- after cleaning, a report is automatically generated including the cleaning duration, the chemical consumption and the number of anomalies, as well as a self-evaluation of the result.

6. Conclusions

A three-level analysis **approach** has been described to provide a benchmark to analyze historical data in CIP cycles. The case study analysed here is the cleaning of a fermenter in an industrial brewery. Reference models are established in order to evaluate if the future cleanings are performed normally or abnormally. The analysis **approach** is expected to be carried out prior to cleaning, during cleaning and at the end of the cleaning process.

In the demonstrated case of the cleaning of a brewery fermenter, a decreasing trend of the feed pump performance is observed, indicating the necessity of an overall inspection of the system. The process monitoring using multivariate analysis methods is able to detect the localized abnormal events in the process. Trajectory alignment is necessary to overcome time shifting between batches and hence improve the robustness of the model. Using liquid volume consumption as indicator variable to synchronize the dataset is more efficient than DTW in terms of computation speed. Both the Q-statistic and Hotelling's T^2 statistic can indicate the occurrence of an anomalous event. The contribution plots can help to diagnose the cause of the anomaly. When the cleaning is finished, the operators are advised to check the cleaning time or on-off ratio of the pump to evaluate if the cleaning has been performed normally.

1
2
3 The application of the proposed analysis **approach** is not limited to the CIP operations of the
4 brewery fermenter. The possible applications are to utilize historical knowledge to benefit the
5 daily operations, the maintenance and optimization of the system, and the monitoring of new
6 cycles, in order to obtain a deeper understanding of the process and reduce cleaning costs. The
7 model is flexible to involve more data to make it more informative. An ideal situation is to
8 automatically detect and diagnose the anomalies in the system with help of advanced monitoring
9 and control systems.
10
11
12
13
14
15
16
17
18
19
20
21
22

23 ASSOCIATED CONTENT

24
25
26 The following file is available free of charge, and helpful for reading and reviewing:

27
28
29 Supplemented_glossary.docx
30
31
32
33

34 AUTHOR INFORMATION

35 36 37 **Corresponding Author**

38
39
40 * Ulrich Krühne, E-mail: ulkr@kt.dtu.dk
41
42
43
44

45 ACKNOWLEDGMENT

46
47 This paper results from the DRIP (Danish partnership for Resource and water efficient Industrial
48 food Production) project. We acknowledge that this work is partly funded by the Innovation
49 Fund Denmark (IFD) under contract No. 5107-00003B, and by the Technical University of
50 Denmark (DTU).
51
52
53
54
55
56
57
58
59
60

NOMENCLATURE

B	The threshold for the Hotelling's T^2 statistic
c	The coefficient of the basic function
c_α	The normal deviation corresponding to the $(1 - \alpha)$ percentile
C_Q	Contribution to the Q-statistic
C_{T^2}	Contribution to the Hotelling's T^2 statistic
e	Residual
F	The F-distribution value
h_0	A coefficient for computing the threshold of the Q-statistic
I	Number of batches
J	Number of time indices
$J[x]$	Roughness term
K	The number of variables
M	The number of principal components
p	Probability value
p_{km}	Loading for variable k in dimension m
P	Loading matrix
Q	Principal loading matrix
R	The number of batches in the reference set
S	Covariance matrix
t	Time, [s]
T	PCA scores
T^2	Hotelling T^2 statistic
x	The predicted function in FDA

X	The matrix of variables
y	The observed dataset in FDA
α	Significance level
σ_j	The j^{th} eigenvalue
ϵ	The noise term in FDA
Φ	Basic function
γ	Safety margin
η	The ratio of the correct judgement
λ	Roughness penalty coefficient
δ^2	The threshold for the Q-statistic
θ_i	The i^{th} coefficient for computing the threshold of Q-statistic

ABBREVIATIONS

CCP, Critical control point; CIP, Cleaning-in-Place; CPV, Cumulative percent variance; DTW, Dynamic time warping; EHEDG, European Hygienic Engineering & Design Group; FDA, Functional data analysis; FN, False negative; FP, False positive; GCV, Generalized cross-validation; MPCA, Multiway principal component analysis; PC/PCs, Principal component(s); PCA, Principal component analysis; PENSSE, Penalized sum of squares; SCADA, Supervisory control and data acquisition; SOP, Standard operation procedure; STD, Standard deviation; TN, True negative; TP, True positive

REFERENCES

- (1) Lelieveld, H. L. M.; Mostert, M. A.; Holah, J. *Handbook of Hygiene Control in the Food*

- 1
2
3 *Industry*, 1st ed.; Woodhead Publishing: Cambridge, UK, 2005.
- 4
5
6 (2) Yang, J.; Jensen, B. B. B.; Nordkvist, M.; Rasmussen, P.; Gernaey, K. V.; Krühne, U.
7
8 CFD Modelling of Axial Mixing in the Intermediate and Final Rinses of Cleaning-in-
9
10 Place Procedures of Straight Pipes. *J. Food Eng.* **2018**, *221*, 95–105.
- 11
12
13 (3) EHEDG Secretariat. *Cleaning Validation in the Food Industry - General Principles (Doc.*
14
15 *45, Part 1)*; EHEDG: Frankfurt, Germany, 2016.
- 16
17
18 (4) Goode, K. R. *Characterising the Cleaning Behaviour of Brewery Foulants - To Minimise*
19
20 *the Cost of Cleaning in Place Operations*, University of Birmingham, 2012.
- 21
22
23 (5) Davey, K. R.; Chandrakash, S.; O'Neill, B. K. A New Risk Analysis of Clean-In-Place
24
25 Milk Processing. *Food Control* **2013**, *29* (1), 248–253.
- 26
27
28 (6) Trinh, L.; Willey, A. R.; Martin, P. J.; Ashley, J.; Tothill, I. E.; Rodgers, T. L. Rate-Based
29
30 Approach to Cleaning-in-Place. *Ind. Eng. Chem. Res.* **2017**, *56*, 6695–6702.
- 31
32
33 (7) Martin, E.; Montague, G.; Robbins, P. A Quality by Design Approach to Process Plant
34
35 Cleaning. *Chem. Eng. Res. Des.* **2013**, *91* (6), 1095–1105.
- 36
37
38 (8) Leo H. Chiang; Russell, E. L.; Braatz, R. D. *Fault Detection and Diagnosis in Industrial*
39
40 *Systems*, 1st ed.; Springer-Verlag London, 2001.
- 41
42
43 (9) Roy, K.; Undey, C.; Mistretta, T.; Naugle, G.; Sodhi, M. Multivariate Statistical
44
45 Monitoring as Applied to Clean-in-Place (CIP) and Steam-in-Place (SIP) Operations in
46
47 Biopharmaceutical Manufacturing. *Biotechnol. Prog.* **2014**, *30* (2), 505–515.
- 48
49
50 (10) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining*,
51
52
53
54
55
56
57
58
59
60

- 1
2
3 *Inference, and Prediction*, 2nd ed.; Springer-Verlag New York, 2009.
- 4
5
6
7 (11) Kjellberg, K. Rotary Jet Head ‘Burst’ Cleaning Technology Delivers Significant Savings
8 in Cleaning Costs. *EHEDG Yearbook 2015/2016*. 2016, 96–98.
- 9
10
11 (12) Ündey, C.; Cinar, A. Statistical Monitoring of Multistage, Multiphase Batch Processes.
12 *IEEE Control Systems*. October 2002, 40–52.
- 13
14
15 (13) Ramsay, J. O.; Silverman, B. W. *Functional Data Analysis*, 2nd ed.; Springer-Verlag New
16 York, 2005.
- 17
18 (14) Mears, L.; Nørregård, R.; Sin, G.; Gernaey, K. V.; Stocks, S. M.; Albaek, M. O.; Villez,
19 K. Functional Unfold Principal Component Regression Methodology for Analysis of
20 Industrial Batch Process Data. *AIChE J.* **2016**, 62 (6), 1986–1994.
- 21
22 (15) Ramsay, J. O.; Hooker, G.; Graves, S. *Functional Data Analysis with R and MATLAB*, 1st
23 ed.; Springer-Verlag New York, 2009.
- 24
25 (16) Nomikos, P.; MacGregor, J. F. Multivariate PSC Charts for Monitoring Batch Processes.
26 *Technometrics* **1995**, 37 (1), 41–59.
- 27
28 (17) J. Keogh, E.; Pazzani, M. J. Derivative Dynamic Time Warping. In *Proceedings of the*
29 *2001 SIAM International Conference on Data Mining*; Society for Industrial and Applied
30 Mathematics: Chicago, USA, 2001; 1–11.
- 31
32 (18) Wang, S.; Xiao, F. AHU Sensor Fault Diagnosis Using Principal Component Analysis
33 Method. *Energy Build.* **2004**, 36 (2), 147–160.
- 34
35 (19) Ündey, C.; Ertunç, S.; Çınar, A. Online Batch/Fed-Batch Process Performance

- 1
2
3 Monitoring, Quality Prediction, and Variable-Contribution Analysis for Diagnosis. *Ind.*
4
5 *Eng. Chem. Res.* **2003**, *42* (20), 4645–4658.
6
7
8
9 (20) Wold, S.; Kettaneh, N.; Fridén, H.; Holmberg, A. Modelling and Diagnostics of Batch
10
11 Processes and Analogous Kinetic Experiments. *Chemom. Intell. Lab. Syst.* **1998**, *44* (1–2),
12
13 331–340.
14
15
16 (21) Gurden, S. P.; Westerhuis, J. A.; Bro, R.; Smilde, A. K. A Comparison of Multiway
17
18 Regression and Scaling Methods. *Chemom. Intell. Lab. Syst.* **2001**, *59* (1), 121–136.
19
20
21 (22) Kadlec, P.; Gabrys, B.; Strandt, S. Data-Driven Soft Sensors in the Process Industry.
22
23 *Comput. Chem. Eng.* **2009**, *33* (4), 795–814.
24
25
26
27 (23) Zhao, C.; Gao, F. Fault-Relevant Principal Component Analysis (FPCA) Method for
28
29 Multivariate Statistical Modeling and Process Monitoring. *Chemom. Intell. Lab. Syst.*
30
31 **2014**, *133*, 1–16.
32
33
34
35 (24) Miller, P.; Swanson, R. E.; Heckler, C. E. Contribution Plots: A Missing Link in
36
37 Multivariate Quality Control. *Appl. Math. Comput. Sci.* **1998**, *8* (4), 775–792.
38
39
40 (25) Celebi, M. E.; Aydin, K. *Unsupervised Learning Algorithms*, 1st ed.; Springer
41
42 International Publishing, 2016.
43
44
45 (26) Lelièvre, C.; Legentilhomme, P.; Gaucher, C.; Legrand, J.; Faille, C.; Bénézech, T.
46
47 Cleaning in Place: Effect of Local Wall Shear Stress Variation on Bacterial Removal from
48
49 Stainless Steel Equipment. *Chem. Eng. Sci.* **2002**, *57* (8), 1287–1297.
50
51
52
53 (27) Jensen, B. B. B. Numerical Study of Influence of Inlet Turbulence Parameters on
54
55 Turbulence Intensity in the Flow Domain: Incompressible Flow in Pipe System. In
56
57
58
59
60

- 1
2
3 *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process*
4 *Mechanical Engineering*; 2007; Vol. 221, 177–186.
5
6
7
8
9 (28) Jensen, B. B. B.; Friis, A. Critical Wall Shear Stress for the EHEDG Test Method. *Chem.*
10 *Eng. Process.* **2004**, *43* (7), 831–840.
11
12
13
14 (29) Singh, R. P.; Heldman, D. R. *Introduction to Food Engineering*, 5th ed.; Academic Press,
15 2014.
16
17
18
19 (30) Lehman, A.; O'Rourke, N.; Hatcher, L.; Stepanski, E. J. *JMP® for Basic Univariate and*
20 *Multivariate Statistics: Methods for Researchers and Social Scientists*, 2nd ed.; SAS
21 Institute: Cary, NC, 2013.
22
23
24
25
26
27 (31) Westerhuis, J. A.; Gurden, S. P.; Smilde, A. K. Generalized Contribution Plots in
28 Statistical Process Monitoring. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 95–114.
29
30
31
32 (32) Van Den Kerkhof, P.; Vanlaer, J.; Gins, G.; Van Impe, J. F. M. Analysis of Smearing-out
33 in Contribution Plot Based Fault Isolation for Statistical Process Control. *Chem. Eng. Sci.*
34 **2013**, *104*, 285–293.
35
36
37
38
39
40 (33) Kassidas, A.; MacGregor, J. F.; Taylor, P. A. Synchronization of Batch Trajectories Using
41 Dynamic Time Warping. *AIChE J.* **1998**, *44* (4), 864–875.
42
43
44
45
46
47
48

49 ABSTRACT GRAPHIC
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

