

Data-Driven Approach toward Long-Term Equipment Condition Assessment in Sterile Drug Product Manufacturing

Philipp Zürcher, Sara Badr,* Stephanie Knüppel, and Hirokazu Sugiyama



Cite This: *ACS Omega* 2022, 7, 36415–36426



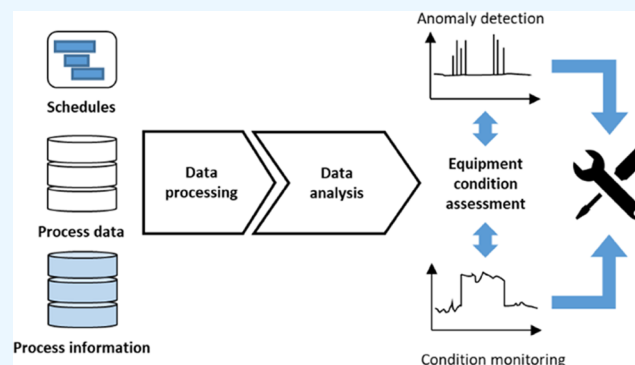
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: A two-stage data-driven methodology for long-term equipment condition assessment in drug product manufacturing is presented with a case study for a commercially operating aseptic filling line. The methodology leverages process monitoring data. Sensor measurements are partitioned using process information and maintenance schedules that are available on different databases. Data is processed to tackle heterogeneity in sources and formats. The data is cleaned to remove the effects of short-term variabilities and to enhance underlying long-term trends. Two approaches are presented for data analysis: first, anomaly detection using independent component analysis (ICA), where clusters of outliers are identified. The frequency and timing of such outliers yield important insights regarding maintenance schedules and actions. The second approach enables condition monitoring using principal component analysis (PCA). Long-term operational baselines are identified and shifts therein are linked with different process and equipment faults. This approach highlights the impact of equipment deterioration on shifting operational data baselines and shows the potential for the combined application of ICA and PCA for equipment condition monitoring. It can be applied within predictive maintenance applications where the installation of new specialized sensors is difficult, like in the pharmaceutical industry.



1. INTRODUCTION

Digitalization is one of the main pillars of Industry 4.0.¹ The emergence of digitized manufacturing facilities recording large amounts of various parameters (i.e., Big Data) in an automated manner allows for the introduction of data-based solutions.^{2,3} Concepts of digitalization are already established in the pharmaceutical industry such as the Pharma 4.0 initiative.^{4,5} Awareness of the importance of such concepts is rising among industry leaders.⁶ A digital transformation would highly benefit the pharmaceutical industry. It is an industry with a high product value, where unexpected production failures lead to expensive product losses. This offers a strong incentive for applications that could reduce production faults and downtime. Furthermore, it is a highly regulated industry with strict procedures regarding keeping records of all processes, and equipment parts directly linked to the product to ensure patient safety as part of good manufacturing practice (GMP).⁷ This ensures a rich environment of historical production data available for analysis.

Quality by design (QbD) and process analytical technology (PAT) are well-established concepts in the pharmaceutical industry.⁸ PAT involves monitoring critical quality attributes. In combination with data analysis techniques, it can ensure product quality, higher production efficiency, and enhanced process understanding.⁹ Applications of PAT in pharma often

include alternative measurements for monitoring process parameters that cannot be measured in real time, for example, concentration measurements via spectroscopy.¹⁰ The same concepts could also apply for equipment condition monitoring, where equipment deterioration is often difficult to measure directly. Innovative data analysis techniques could therefore provide additional insights and enable applications such as predictive maintenance.¹¹

Within the pharmaceutical industry, periodic equipment maintenance is usually performed at predetermined time intervals.¹² The implemented time-based maintenance could cause problems as it could miss equipment deterioration, leading to increased process faults if the intervals are too long. On the other hand, if the intervals are too short, it could lead to unnecessary costs, material consumption, long facility downtime, and organizational effort. Predictive maintenance applications could help optimize this process to avoid

Received: July 4, 2022

Accepted: September 19, 2022

Published: October 10, 2022



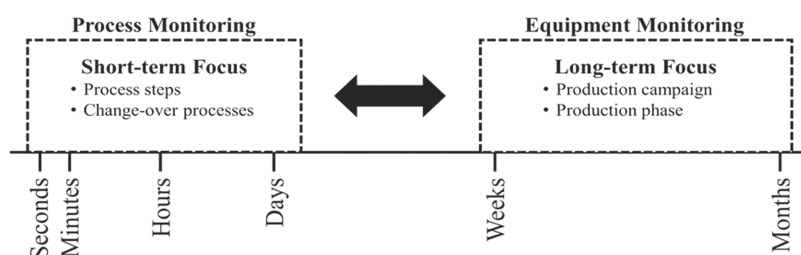


Figure 1. Representation of different timescales for process and equipment monitoring.

production interruptions through either equipment faults or superfluous maintenance activities. Data-driven applications for equipment condition monitoring have been suggested outside pharmaceutical manufacturing.¹³ Tailored sensors for predictive maintenance (e.g., vibrational analysis) are often installed in production facilities.¹⁴ Installing such new sensors in a running pharmaceutical facility, however, could be cumbersome due to the GMP requirements for registration and validation of any changes within the manufacturing facility. The extra costs for the revalidation process are also an additional burden that could deter the application of changes to an operating process.¹⁵ As such, there is a need to explore the existing large pool of historical process and equipment data for potential uses in long-term equipment condition monitoring.

Currently, in addition to backtracking purposes, the recorded process data can be used for process monitoring and fault detection in batch and change-over processes.^{16,17} A monitoring workflow for the formulation, fill, and finish processes was published.¹⁸ Data-driven batch monitoring in drug product manufacturing for change-over processes has been suggested with a focus on cleaning and sterilization in place (C/SIP)¹⁷ and decontamination.¹⁹ The latter work showed that deteriorations could be observed across multiple batches before a failure occurs. This finding is a strong indicator that process monitoring data can be used for equipment condition monitoring over a longer period if adequately addressed. Current research in predictive monitoring and predictive maintenance applications, however, is mainly focused on methodology development using simulated datasets.^{20,21} Case studies using real industrial data are scarce and mostly focused on fields outside pharmaceutical manufacturing.²² In pharmaceutical manufacturing, commercial applications of predictive maintenance are extremely scarce.²³ One of the few published applications of condition-based maintenance for utility systems has been shown through the installation of tailored monitoring sensors such as thermal, vibrational, and ultrasonic analysis which led to an overall reduction in equipment faults.²⁴

The nature of operations in pharmaceutical manufacturing greatly influences the structure of the collected data and the challenges faced in processing it. Drug product manufacturing is typically performed in batch operations.²⁵ Generally, batch process data exhibits nonlinear process behavior as well as intra- and interbatch variability.^{26,27} Information retrieval could be hindered by such variabilities, e.g., variations in batch sizes and processed amounts, or in the lengths of process steps as a result of manual interventions with continuous data recording without separating production and process downtime. Despite the abundance of process and maintenance data, the variations in the formats and locations of data storage could pose additional challenges during data processing and aggregation.

Using collected process data for a secondary application also poses additional challenges by not providing a comprehensive image of the problems at hand, where direct measurements of targeted phenomena are missing. Therefore, a methodology is still needed for properly harnessing information regarding long-term equipment health from the available historical short-term process monitoring data and overcoming the challenges posed by the data.

This work proposes a data-driven approach to leverage collected short-term process monitoring information in operating facilities for long-term equipment condition monitoring in drug product manufacturing. A methodology is presented for data cleaning and analysis using real industrial data. The suggested methodology combines information from online monitoring sensors with data from databases containing equipment information (e.g., downtime, fault types, or maintenance actions). Two approaches for the analysis of long-term equipment conditions are presented in this work. First, anomaly detection is introduced to identify periods with long-term variations across maintenance cycles and the sources of equipment-related disturbances. Second, an approach for condition monitoring is developed, where a higher-resolution analysis is conducted to correlate detected long- and short-term baseline shifts with eventual equipment faults. An industrial case study is presented to demonstrate the application of the methodology, where equipment faults are linked with shifts in operational baselines that were already observed weeks or months earlier in process monitoring data. This methodology gives a long-term comprehensive overview of changes in the equipment and enables the early detection of changes in equipment condition and their sources. It allows for planned early interventions and thus avoiding unexpected product losses and process downtime. In this case study, process data is used from an aseptic filling line collected at a drug product manufacturing facility of F. Hoffmann-La Roche Ltd. in Kaiseraugst, Switzerland.

2. METHODOLOGY

2.1. Overview. A two-stage approach is presented for long-term equipment condition assessment using process monitoring information collected for short-term purposes, e.g., for monitoring individual runs including different process steps, or change-over operations. Sensors are installed in the production facility for univariate process monitoring and include parameters such as pressure and temperature. This work focuses on detecting small variations in equipment condition that last over long periods of operation. Such variations would not cause immediate process failures, but their accumulation could lead to deterred operation over time. Such changes would henceforth be referred to as long-term effects. In contrast, short-term effects are a result of variations within each process or production batch and can be observed in end-to-

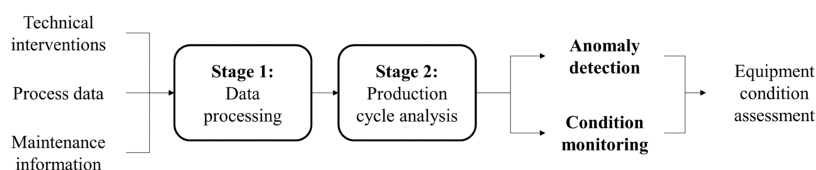


Figure 2. Schematic overview of the developed two-stage methodology.

end process monitoring. Short-term effects may have a larger impact on operation and could lead to process interruptions. In this work, long-term effects were determined as those lasting for more than one operating day (i.e., longer than the span of one batch). Early detection of such long-term effects provides a window for maintenance actions that can be scheduled during regular operations without causing additional downtime. Figure 1 shows a schematic highlighting the main differences in the scope between long- and short-term monitoring applications.

A schematic representation of the developed methodology for equipment condition assessment is shown in Figure 2. The first stage involves data collection from multiple sources in diverse formats followed by data preprocessing. Process information is usually available offline in text form. It includes documentation, e.g., process recipes, information about alarms and failures, as well as production logbooks and diaries. Different databases are accessed to retrieve process data as sensor measurements, which are primarily collected online for process monitoring purposes. Offline maintenance records are also available and include details of actions taken for the various equipment. They are usually available in text form. Tackling the heterogeneity in the data is the goal of this first “data processing” stage, where data from the different sources are linked and cleaned. Data can be connected through process tags or time stamps. Equipment-related failures are analyzed and linked to the maintenance actions in historical records. An example of data cleaning includes filtering out downtime and separating the production data by batch and the corresponding change-over processes. The cleaned data is then used in the second stage “production cycle analysis”. In this stage, anomaly detection is conducted to identify clusters of events representing nonregular operation. Condition monitoring involves identifying baselines for normal operation in addition to baseline shifts indicating ongoing deterioration within the unit and the impact of maintenance actions taken. Findings from both steps are combined for an in-depth equipment condition assessment.

2.2. Case Study. In this work, process data of an aseptic filling line in drug product manufacturing of F. Hoffmann-La Roche Ltd, in Kaiseraugst, Switzerland is used. The aseptic filling line consists of four main units, namely, a vial washing machine, sterilization tunnel, filling machine, and isolator. The developed methodology was applied individually to the four units and can be considered to be independent of each other. Process data is collected for various short-term purposes such as monitoring change-over processes and validating that production steps were within specifications. Two years of production data were considered in this work. In the investigated facility, regular equipment maintenance is performed in a time-based manner at fixed intervals. In this case, maintenance is undertaken following production phases, which typically have a duration of about 4 months. A schematic overview of the annual production and maintenance schedules is given in Figure 3. In the following sections, a

detailed description is given for the application of the developed methodology within the investigated facility.

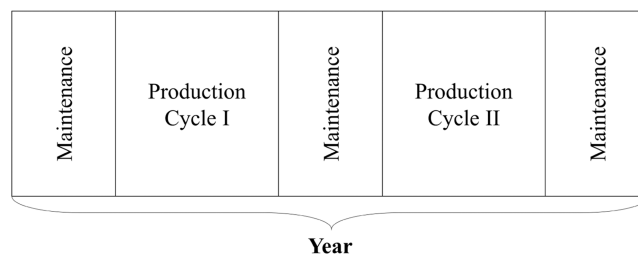


Figure 3. Schematic representation of production and maintenance phases within one year of production.

2.2.1. Stage 1—Data Processing. The aim of this stage is to combine the available information from different sources and in different formats to obtain suitable datasets for long-term equipment condition monitoring. Figure 4 shows a schematic representation of the data processing steps. Individual units were investigated independently of each other. For each unit, process data was retrieved for analysis from online sensor measurements that are primarily recorded for univariate process monitoring purposes. Process data is continuously collected and stored as data historians on a PI OSIsoft database.²⁸ No missing process data was encountered. The data was retrieved in 10 s intervals regardless of equipment status, e.g., in production, change-over, or idle. Table 1 shows examples of recorded process data in each unit. Similar types of sensors are often installed at multiple locations within each unit. To track equipment-related changes, data from batch production processes over a span of 2 years were considered. The comparison of similar production processes in the same equipment over a long time span can yield insights into non-process-related changes, such as equipment deterioration. To achieve this comparison, data from other sources were used for partitioning and labeling the available process data. In the first step, the process data was separated into production and maintenance phases according to the maintenance information, which was available as excel files containing historic maintenance schedules. Recordings during maintenance phases were excluded from the dataset. The remaining dataset still contained all recordings during the production phase, including idle time, change-over operations, and test runs in addition to the targeted production data. In the subsequent process identification step, batch production runs were exclusively extracted from the continuously collected production phase data. The PI OSIsoft database was used in this step. It consists of continuously collected process tags describing the type of ongoing processes. Time points coinciding with production tags were extracted and the process data accordingly refined to retain only production-relevant data. Data recorded during individual batch runs were identified by batch ID and extracted from the production

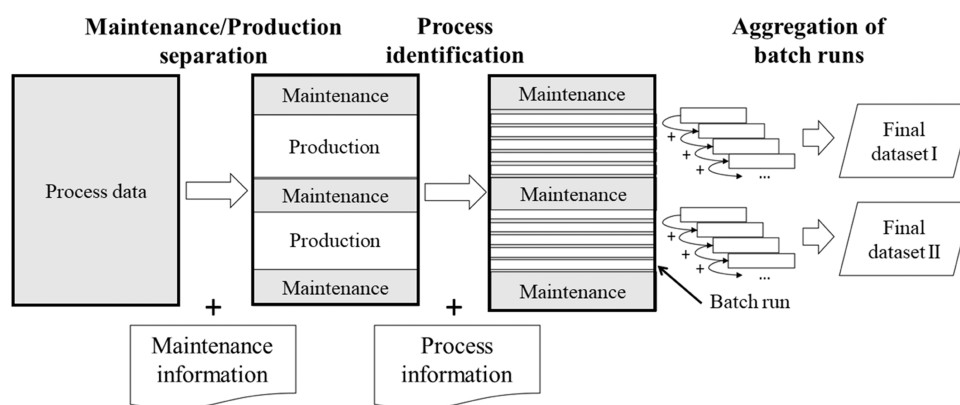


Figure 4. Schematic representation of the data processing methodology to obtain production cycle datasets.

Table 1. Overview of Process Units and Type of Recorded Sensor Data within the Aseptic Filling Line

unit	examples of physical recordings	number of sensors	time interval
vial washing machine	<ul style="list-style-type: none"> spraying time pressure temperature 	14	
sterilization tunnel	<ul style="list-style-type: none"> position equipment pressure transport speed heating power ventilation power temperature 	21	
filling machine	<ul style="list-style-type: none"> pressure flowrate machine power temperature conductivity 	22	10 s
Isolator	<ul style="list-style-type: none"> aeration time pressure H₂O₂ injection rate relative humidity H₂O₂ amount temperature 	47	

dataset. The individual batch datasets within each production cycle were aggregated to the final datasets used for analysis. Data from two production cycles per year were obtained separated by half-yearly maintenance phases. For each dataset over one million data points were included in the two-year period. The length of the final datasets for each unit varied due to differences in individual unit operation time across the production cycles. The described methodology was applied individually for each unit of the aseptic filling process.

Offline databases for process interruptions and technical interventions were used to connect data-driven findings with recorded occurrences in the facility. Process interruptions contain text-based information about events leading to an interruption of the running process. Process interruptions do not only include equipment-related events but can also include process-related problems such as contamination of vials, fallen or displaced vials, or alarms during change-over operations. These logs are recorded by operators on the shop floor on a machine execution system (MES). Technical interventions, where technicians were required on the shop floor, are recorded separately. They are recorded on the SAP plant

maintenance (PM) system by the technicians. Technical interventions include descriptions of maintenance actions taken following equipment-related faults. The text-based data was parsed, and only equipment-relevant data was retained. Excluded entries described general actions not related to manufacturing equipment, e.g., lamp replacements in the room.

The frequency and type of recorded incidents varied depending on the nature of the operation and equipment. The filling machine is the most complex among the investigated units. It featured the highest number of recorded technical interventions and process interruptions. The sterilization tunnel had the smallest number of process interruptions. This unit features a relatively simpler process, where vials are transported during depyrogenation. The number of process interventions and interruptions in this unit were almost the same, indicating that when a problem occurred, it was most likely equipment-related, requiring intervention of a technician. The vial washing machine and the isolator had comparable numbers of recorded process interruptions, which were mostly linked to problems during change-over operations.

After the analysis of the text-based records, extracted faults were then classified into two categories: equipment-related and non-equipment-related (i.e., externally caused) faults as summarized in Table 2. Type A faults are externally driven

Table 2. Summary of the Two Fault Classes

fault class	description
Type A	non-equipment-related incident
Type B	equipment-related fault

incidents that require manual interventions by the operators. Causes for Type A faults necessitating a process reset could be manifold, a prominent example of which is vial glass breakages inside the filling line. Type B faults, on the other hand, result from equipment failure necessitating replacement, maintenance actions, or parameter adjustment.

Information from these databases was used to interpret the data patterns observed after the analysis of the production cycle data. The offline entries were aligned with the refined production cycle data according to the recorded time stamps. The offline entries suffer from delays between the actual event occurrence and the entry into the database, which does not refer to the exact failure time. Therefore, an allowance of +1 day was made regarding the accuracy of matching equipment

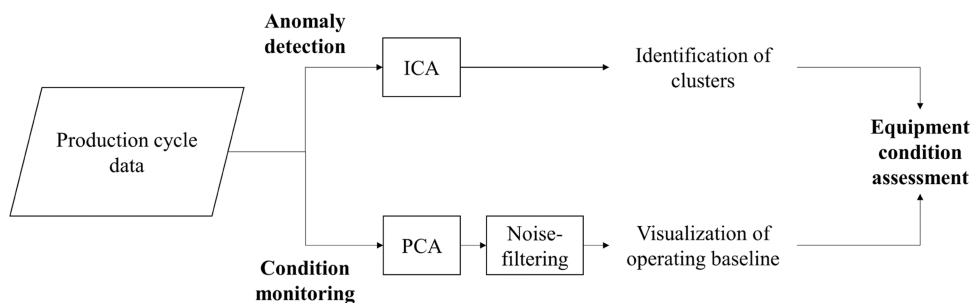


Figure 5. Schematic flowsheet of developed methodology for equipment condition monitoring from short-term process monitoring data.

faults between the offline database and the observed patterns in the data.

2.2.2. Stage 2—Production Cycle Analysis. The production cycle analysis stage aims at identifying two types of patterns in the process data. A schematic representation of the production cycle analysis is shown in Figure 5. First, sources of abnormal process behavior are identified, with a special focus given where repeated instances of outlier events and anomalies are concentrated. An outlier in this work is defined as a source of unusual variance in the production cycle data. Such outliers may not be connected to particular faults and could indicate changes in the system. They are used to track the sources of long-term changes in the system and evaluate actions by the operators (for example the adequacy of maintenance actions in each maintenance period). The frequency and position of anomaly clusters within the production cycle can help decision-makers identify potential equipment-related problems and the needed adjustments to maintenance schedules. Furthermore, detected anomalies indicate the time points of disturbances in the data that could cause equipment deterioration. The second type of data patterns involves establishing baselines for long-term process data for condition monitoring purposes. Observed shifts in the established baselines can indicate changes in equipment condition. Shifts lasting longer than batch durations could indicate long-term changes in the equipment. Such changes highlight ongoing equipment deterioration that could eventually lead to failure or malfunction. The use of different data processing algorithms yields more comprehensive results regarding the sources and impacts of equipment changes in the system. Findings from both, anomaly detection and condition monitoring are used together and linked to offline information about technical interventions to confirm the validity of the methodology for detecting equipment deterioration. This analysis thus links process data with failures, visualizes equipment deterioration, and maps out the impact of maintenance actions on correcting or reestablishing the production baseline.

Dimensionality reduction was crucial in this stage to analyze patterns in the data given the large number of sensors available. Different methods of dimensionality reduction were applied depending on the goal of the analysis. The data was variable-wise unfolded and auto-scaled prior to the multivariate analysis by subtracting the mean of each column and dividing by the standard deviation. Variable-wise unfolding of the data yields an aggregated time series of recorded sensor measurements. Such aggregation of batch data allows for the comparison of long-term changes across the investigated time period.

Anomaly detection was achieved by the application of independent component analysis (ICA)²⁹ to the production cycle dataset. ICA is designed for application to non-Gaussian

datasets. This suits the nature of the data from batch production processes, which are often strongly nonlinear.³⁰ Dimensionality reduction with ICA is based on the idea that measured variables are a mixture of some independent component variables. It is assumed that l measured variables, $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_l(k)]$, at sample k can be expressed as a linear combination of r unknown independent components $[s_1, s_2, \dots, s_r]^T$ (where $r \leq l$); the relationship between them is given by

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S} + \mathbf{E} \quad (1)$$

where n is the number of measurements, $\mathbf{X} = [x_1, x_2, \dots, x_n] \in \mathbf{R}^{l \times n}$ is the data matrix, $\mathbf{A} = [a_1, a_2, \dots, a_r] \in \mathbf{R}^{l \times r}$ is the mixing matrix, $\mathbf{S} = [s_1, s_2, \dots, s_r] \in \mathbf{R}^{r \times n}$ is the independent component matrix, and $\mathbf{E} \in \mathbf{R}^{l \times n}$ is the residual matrix. The ICA problem includes the estimation of the original component \mathbf{S} and the mixing matrix \mathbf{A} from \mathbf{X} . The Fast ICA algorithm was applied in this work.

The data was first whitened using the eigenvalue decomposition method, where considering $\mathbf{x}(k)$ with its covariance $\mathbf{R}_X = E\{\mathbf{x}(k) \cdot \mathbf{x}(k)^T\}$, the eigenvalue decomposition of \mathbf{R}_X was given by

$$\mathbf{R}_X = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (2)$$

where the whitening transformation was expressed as

$$\mathbf{z}(k) = \mathbf{Q} \cdot \mathbf{x}(k) = \mathbf{Q} \cdot \mathbf{A} \cdot \mathbf{s}(k) = \mathbf{B} \cdot \mathbf{s}(k) \quad (3)$$

$$\mathbf{Q} = \mathbf{\Lambda}^{-1/2} \cdot \mathbf{U}^T \quad (4)$$

$$E\{\mathbf{z}(k) \cdot \mathbf{z}(k)^T\} = \mathbf{B} \cdot E\{\mathbf{s}(k) \cdot \mathbf{s}(k)^T\} \cdot \mathbf{B}^T = \mathbf{B} \cdot \mathbf{B}^T = \mathbf{I} \quad (5)$$

where \mathbf{B} is an orthogonal matrix. Then, the following estimate in eq 6 was made to calculate a separating matrix \mathbf{W} to obtain the independent components of the reconstructed data matrix \mathbf{S} as in eq 7

$$\hat{\mathbf{s}}(k) = \mathbf{B}^T \cdot \mathbf{z}(k) = \mathbf{B}^T \cdot \mathbf{Q} \cdot \mathbf{x}(k) \quad (6)$$

$$\hat{\mathbf{S}} = \mathbf{W} \cdot \mathbf{X} \quad (7)$$

resulting in

$$\mathbf{W} = \mathbf{B}^T \cdot \mathbf{Q} \quad (8)$$

The problem of finding an arbitrary full-rank matrix is reduced to the finding of an orthogonal matrix (\mathbf{B}). ICA was performed using the Fast-ICA implementation²⁹ that is part of the scikit-learn package in Python (v3.7). The resulting independent components were studied to identify regions with aggregation of process anomalies.

Condition monitoring was conducted by the application of multiway principal component analysis (MW-PCA)³¹ applied

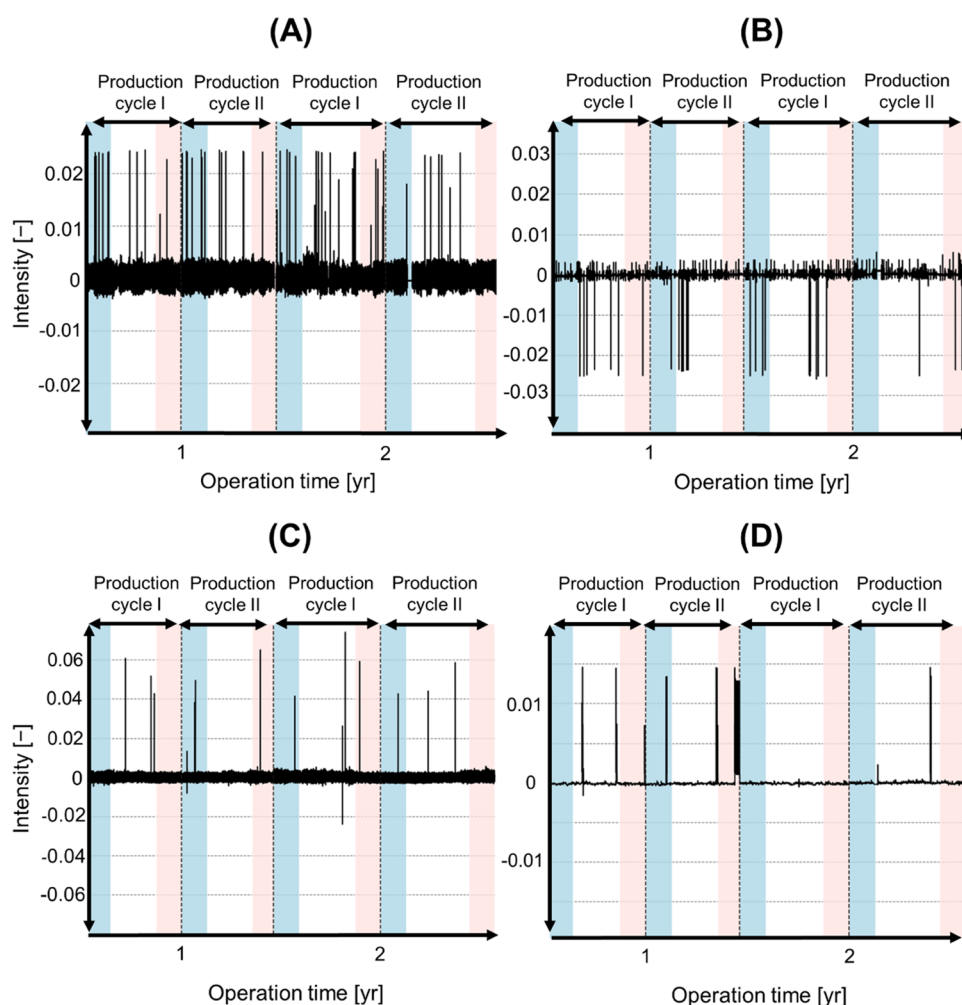


Figure 6. Anomaly detection results using independent components from aggregated batch data over time. Examples of independent components are plotted for two production years for (A) the vial washing machine, (B) the sterilization tunnel, (C) the isolator, and (D) the filling machine. Production cycles are separated by half-yearly maintenance phases. Blue shaded areas indicate starting phases within each production cycle following scheduled maintenance, and red shaded areas indicate the end phase of the production cycle before scheduled maintenance.

on a variable-wise unfolded dataset. In process monitoring applications, batches are compared to each other to detect batch variations and operational faults. In such cases, variable-wise data unfolding can result in difficulties regarding the alignment of batch data when production durations vary. However, in equipment condition monitoring applications, the aggregation of similar operations allows the monitoring of long-term trends over multiple batch runs. Noise created from such batch variations can be filtered by focusing on changes lasting longer than the duration of one batch. Persistent long-term changes in the data can thus be isolated, analyzed, and linked to equipment condition changes.

The decomposition of the matrix X is represented as a sum of the outer product of vectors t_i and p_i . The resulting eqs 9 and 10 are given as follows

$$X = \sum_{i=1}^l t_i p_i^T + \sum_{i=1}^{m-l} \tilde{t}_i \tilde{p}_i^T \quad (9)$$

$$X = \tilde{X} + \hat{X} = TP^T + E = TP^T + \tilde{TP}^T = \overline{TP}^T \quad (10)$$

The dataset undergoing dimensionality reduction did not include time as a factor. Only sensor measurements were considered with a time index. The number of used principal

components was chosen to account for 95% of the explained variance in the data. For all of the investigated units, this resulted in using a number of principal components equal to half the number of available sensors. The same number of parameters was also considered during ICA.

The dependence on operator input in some semiautomated process steps causes additional noise in the datasets. The Savitzky–Golay noise filter was applied during condition monitoring to the data to improve the signal-to-noise ratio and visualize underlying trends in the plotted principal components.³² The noise-filtered components were studied to establish and compare operational baselines across multiple production cycles.

As a follow-up analysis step, the impact of equipment faults and maintenance actions were studied in relation to shifts in the established baselines and the peaks identified in the principal and independent components, respectively. This step aimed to confirm and validate the impact of different actions on the observed trends in the data. Links were established in this step between observed patterns and different types of faults eventually occurring weeks or months later. This approach enabled the identification of characteristic recurring

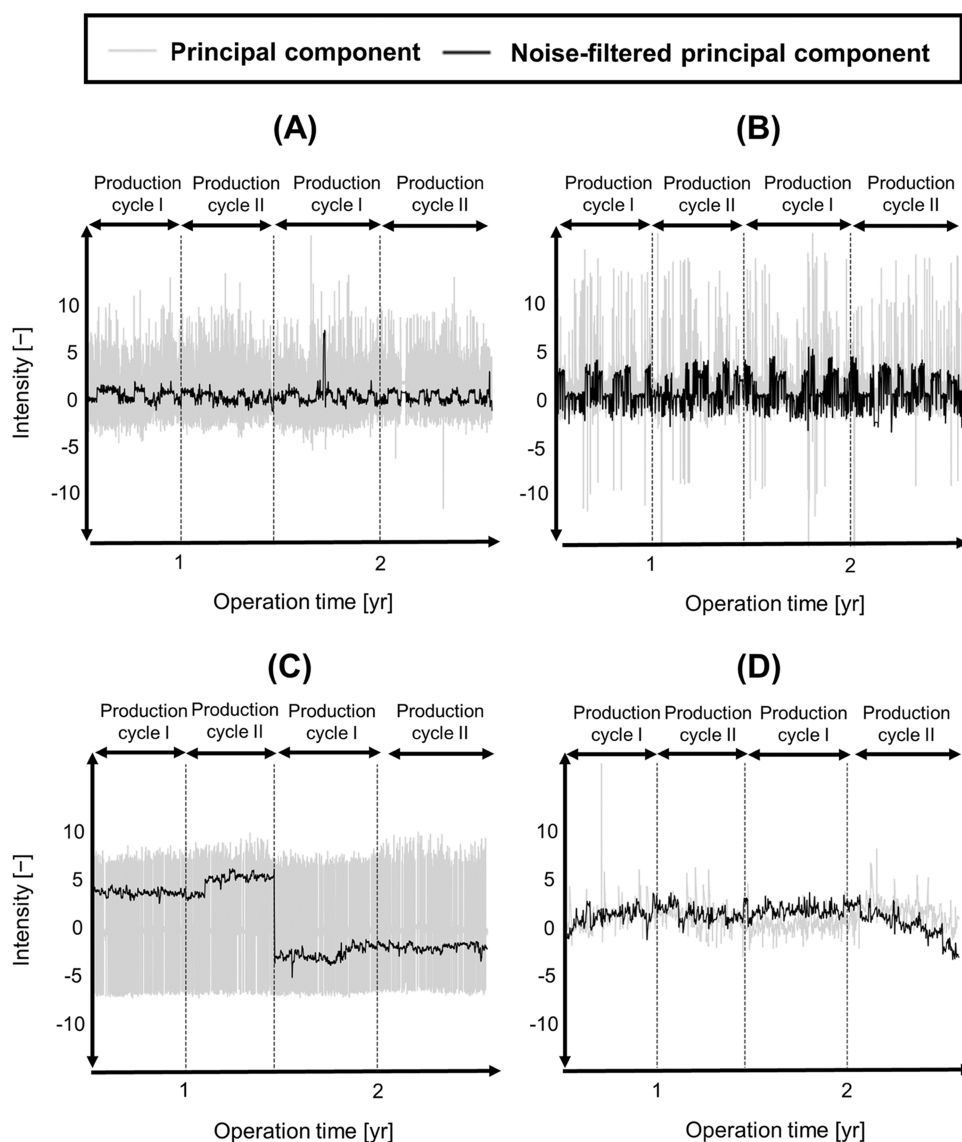


Figure 7. Condition monitoring results using principal components from batch data aggregated over time. Examples of principal components showing long-term variations are plotted for two production years for (A) the vial washing machine (PC6), (B) the sterilization tunnel (PC2), (C) the isolator (PC1), and (D) the filling machine (PC5). Production cycles are separated by half-yearly maintenance phases. The original principal components are given in gray, while the noise-filtered components are shown in black.

patterns in the data and provided insights regarding the relevant equipment condition.

3. RESULTS AND DISCUSSION

3.1. Anomaly Detection. The proposed methodology for anomaly detection was applied to two full years of production data consisting of four production cycles in all four units of the aseptic filling line. Independent components were plotted with time for each investigated unit. Figure 6 shows examples of the plotted independent components for the vial washing unit (A), the sterilization tunnel (B), the isolator (C), and the filling machine (D). Clear outliers were observed for all four process units represented in the case study. The outlier peaks were present in the same locations in all independent components, but with different intensities. The independent components with the best signal-to-noise ratio were shown for each unit in Figure 6.

Clusters of outlier peaks representing process anomalies were observed at different frequencies in each of the units

indicating differences in the sources of variations within the data from each process unit. In Figure 6, start-up phases of production following scheduled maintenance are shaded in blue and late production phases preceding scheduled maintenance are shaded in red for better visualization. Outlier peaks were observed along the course of the investigated production cycles, but clusters of outliers were also observed at the beginning of the production cycles (e.g., the first three production cycles for the vial washing unit (A) and production cycle I of year 2 for the sterilization tunnel (B)). Several production cycles also had a second cluster of anomalies toward the end of the production cycle. This second cluster can be again seen in the first three production cycles of the vial washing machine (A), and production cycle II of year 1 in the filling machine (D). During maintenance phases, large parts of the equipment are dismantled and reconnected before resuming production. A cluster of anomalies at the start of the production phase could be an indication of problems related to improper restoration or connection of the equip-

ment. Clusters of anomaly sources in such phases should be closely monitored as an indicator of the quality of maintenance actions taken. If clusters are repeatedly obtained in the same phases, then the maintenance procedure should be revised, leading to the development of more standardized maintenance procedures and quality checks.

During production, manual interventions may be required to overcome some equipment malfunctions. Anomalies detected during production could thus result from the original malfunction or from problems arising from the subsequent manual intervention. An increasing frequency of observed anomalies during ongoing production, especially toward the end of the cycle, is an indication of equipment deterioration over time necessitating maintenance actions. More outliers are detected in the vial washing machine compared to other units, e.g., the filling unit, which had more process interruptions and technical interventions. This was mainly attributed to the available sensors in each unit. For example, vial breakages frequently occur in both units. However, in the washing machine, the breakages mainly occur by the spray needles and can be more easily detected by the pressure sensors there. In the filling unit, the breakages mainly occur on the conveyor belt system but away from the installed sensors, making them harder to detect by the given sensor infrastructure.

The application of this methodology over extended production periods and the analysis of the clusters of anomalies could provide a useful tool for decision-makers. Analysis of recurring patterns of clusters at the beginning of production cycles could help identify problematic maintenance actions, e.g., difficulties in the proper reconnection of certain equipment parts. This would prompt actions such as adjustment of the checks performed after maintenance or better equipment design in the long run. Limiting manual interventions through better design would thus help lower the occurrences of problems during operations. Recurring patterns of anomalies observed toward the end of the production cycle help decision-makers to identify areas where equipment deterioration is not sufficiently avoided through maintenance schedules at fixed time intervals. The insights gained can be used to adjust maintenance schedules or to implement preemptive targeted interventions when necessary. Outliers only indicate sources of variances, but do not evaluate the quality or impact of that change. Therefore, the second condition monitoring procedure is needed to link equipment faults with the data-driven findings from observed variances in the system.

3.2. Condition Monitoring. Figure 7 shows an example of the principal component scores over time for two production years for each (A) of the vial washing machine, (B) the sterilization tunnel, (C) the isolator, as well as (D) the filling machine. The figure also shows the noise-filtered principal components in black. The application of the noise filtering algorithm has successfully increased the signal-to-noise ratio enabling the identification of underlying long-term shifts and trends. During condition monitoring, all resulting principal components were visually analyzed. Shifts in the underlying baselines in the data could be observed in multiple plotted principal components, but not all of them. Figure 7 only shows selected results of observed shifts for each of the investigated process units.

In Figure 7, long- and medium-term shifts in the baseline were observed for all units at different production cycles. For example, long-term shifts were observed spanning several

months within production cycles for PC1 in the isolator in Figure 7C. Medium-term shifts (on the scale of days to weeks) were observed in the case of the vial washing machine and sterilization tunnel. Figure 7A,B shows a relatively stable baseline across production cycles in the vial washing machine, and the sterilization tunnel, respectively, compared to the other units. Recurring shifts in the respective baselines were observed before being restored to the original position. Figure 7A shows principal component 6 (PC6), while (B) shows the changes in PC2. In the case of the isolator (shown in (C)), sustained shifts in the baseline were observed across different production cycles. During production cycle II of year 1, an upward shift in the baseline was observed, which was not restored until the end of the production cycle. A larger step change was observed in the baseline starting from the following production cycle (cycle I in year 2). This indicates changes applied to the isolator in this maintenance phase between years 1 and 2. During year 2, further fluctuations in the baseline were still observed potentially indicating persisting issues. Finally, for the filling machine (D) showing PC5, no sustained shifts in the baseline were observed. However, a clear downward trend is visible in production cycle II of year 2 as a possible indication of a change in the overall equipment condition.

Equipment condition changes are generally expected to be smaller in magnitude and in their contribution to the overall variance in the data, but persistent over longer spans relative to process-related changes. In this work, the later principal components exhibited more long-term trends (changes lasting more than 1 day), such as in the washing and filling machines, with PCs 6, and 5, respectively. Long-term changes were observed in PC1 in the case of the isolator (C), this indicated a persistent and serious problem, contributing to significant variability in operations. A comparison of the observed long-term shifts with entries in the process logs for the isolator showed that there was indeed an air leak during year 2, which was fixed in the following maintenance cycle. The impacts of this leak were manifested by an increased number of process interruptions and required manual interventions. Only 2 years of operation data were used in this study. The impact of the problem in the isolator on the overall variance in the data could be reduced when a longer time span is considered. Overall, detecting persistent problems is the main goal for the analysis of equipment condition changes, independent of the contribution to the variance. The ability of this analysis to detect the leak in the isolator using historical data proves the usefulness of the data-driven approach to track long-term equipment conditions using process monitoring data. Further analysis of the data was still required to categorize the problems occurring in other units and their sources. One limitation persists, where the underlying sensor architecture would be unable to fully describe all physical equipment failures occurring at the line. For example, in the case of the filling machine, only temperature and pressure measurements are collected, which could be insufficient to describe all recorded equipment failures.

A follow-up analysis of the short-term shifts in the data can be achieved by overlaying detected equipment faults obtained from the maintenance database with the principal components. Technical interventions were not part of the data analysis and are used to connect detected patterns in the data with historical recorded events. This step aims to detect repetitive patterns in the data and provide a better understanding of their impacts and sources by linking them to recorded faults. The

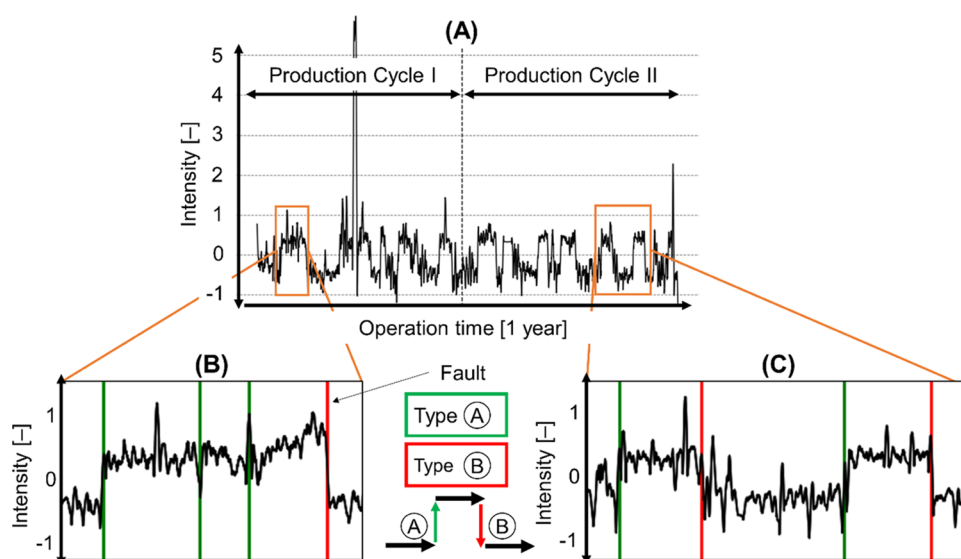


Figure 8. Higher-resolution analysis of principal component 6 from the washing machine during production year 2. (A) PC6 plotted with time over year 2. (B) Zoomed-in view of an example of a long-term baseline shift observed in production cycle I, while (C) gives focus on another example from production cycle II. Equipment faults from historical records are overlaid in (B, C). Externally driven faults (Type A) are shown in green, and equipment-related faults (Type B) are shown in red.

washing machine was used as an example for this analysis. For the washing machine, the first principal components were dominated by short-term variances, and no long-term trends or baseline changes could be observed. Figure 8A shows PC6, zoomed in for year 2 from the data of the washing machine unit. A higher-resolution image is presented in the figure for selected short-term shifts in both production cycles. Equipment faults are overlaid in different colors according to their type in the higher-resolution snapshots. Throughout the analysis of the data, a recurring pattern was observed related to the short-term baseline shifts. (Table 2) The start of the shift generally coincided with a Type A fault (externally driven faults), indicated in green (Figure 8B,C). The shift was maintained as production was resumed for several days up to multiple weeks, sometimes across multiple batches and production campaigns. Eventually, an equipment-related fault (Type B, indicated in red) occurred interrupting production and requiring maintenance actions. Typically, the initial baseline was restored after production was resumed following the Type B maintenance. It should be noted that not all Type A faults resulted in step changes in the baseline as shown in Figure 8B. In that case, an initial shift was observed following a maintenance action linked to a vial counting sensor, this was followed by two glass breakage (Type A) events occurring inside the washing machine, which required debris removal and cleaning. As shown in the figure, in the period following the initial Type A fault, the two others occurred without causing further step changes in the baseline. Instead, more gradual changes in the baseline were observed, especially after the third Type A fault. This behavior indicates changes in the underlying equipment conditions as a result of the initial fault leading to changes in the data patterns. The maintenance action taken after the following Type B fault then restored the original baseline. Other examples are given in the second production cycle in Figure 8C, where two shifts were investigated. Both shifts followed a similar pattern, where a shift caused by a manual intervention following an externally driven fault (Type A) eventually led to an equipment fault

(Type B). In this case, Type A events included vial breakage and maintenance of the conveyor belt system. Type B faults triggered maintenance actions which were then successful at restoring the baseline. Overall, the duration and magnitude of the Type A and Type B faults varied, showing that increased equipment deterioration does not always occur at a comparable level with different categories of external and equipment-related faults.

Figure 9 shows a more in-depth analysis of the same washing machine unit in the same year. In Figure 9A, an example of overlaid principal and independent components for the vial washing machine is shown. Figure 9B shows the same independent component with the recorded technical interventions. It can be seen in Figure 9A that the outlier peaks in the independent component match the baseline shifts in the principal component, which are shown in Figure 8 to coincide with Type A events. The recorded technical interventions in Figure 9B are also shown to be closely linked with the peaks in the independent components and thus the recorded baseline shifts in the plotted principal components. These findings confirmed that the detected anomalies were equipment-related. The persistence of the shift in operational baseline following the technical intervention indicates either that the intervention was not sufficient to fully restore the equipment to its previous state or that the intervention itself had inadvertently caused more changes in the equipment. This analysis can be useful for the detection of the start of the shifts marking equipment deterioration. The early detection of these shifts is important to avoid eventual equipment faults and unexpected downtime.

Overall, 11 distinct baseline shifts in the principal component were identified in year 2 in the washing machine (as shown in Figure 8A). All except one could be related to a recorded technical intervention in the offline database. The additional shift without a record in the offline database indicates that the underlying equipment-related problem was most likely fixed during routine operations with manual adjustments of the line for different product sizes. The remaining 10 baseline shifts were not fixed during normal

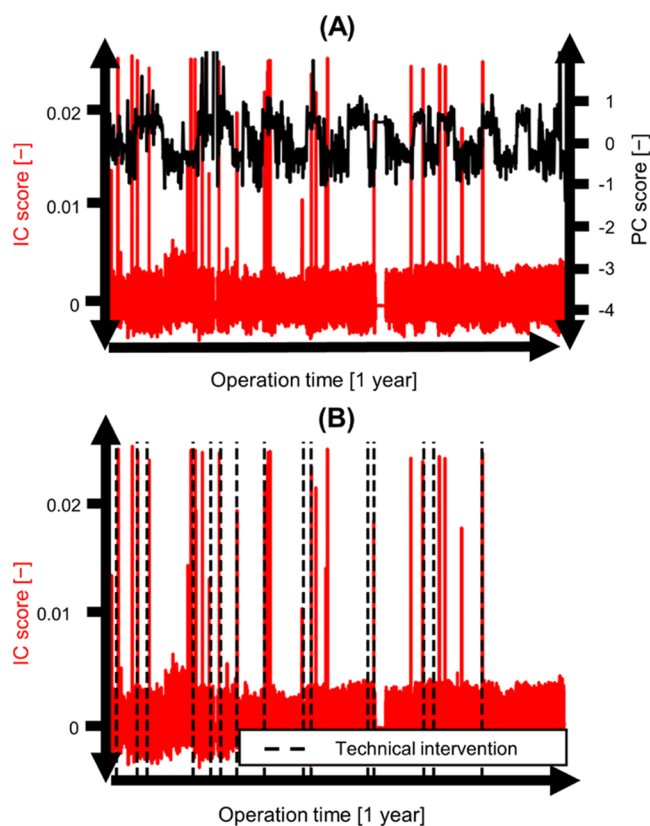


Figure 9. Combined analysis of principal and independent components for the washing machine in relation to recorded equipment-related technical interventions in the unit. (A) Independent component 1 overlaid with principal component 6. (B) Independent component overlaid with recorded technical interventions from the offline database. The figure shows the correspondence of the interventions with the IC peaks, which also coincide with the start of long-term baseline shifts in the PC.

operation and eventually led to type B events requiring additional downtime and technical expertise to fix. This resulted in a total of 24 days of non-value-adding downtime before production was resumed. For example, the type B event shown in red in Figure 8B required 1-day downtime for equipment replacement. The first type B event in Figure 8C was followed by 8 days of downtime until the problem was identified and the necessary equipment was replaced. The second type B event in Figure 8C led to a production stop for 1.5 days. The application of this methodology could therefore be of high use to identify the underlying problem and utilize the planned downtime more effectively to perform the restorative maintenance actions.

This analysis shows that even small equipment-related changes, as evidenced by their small contribution to the variance (PC6), could lead to equipment deterioration and long downtime for the eventual required maintenance. The early detection of such shifts in operational baselines can help avoid long downtime periods by planning maintenance actions in between production runs once a baseline shift is detected. Learning from previous recurring patterns can help shorten the downtime further by linking related events in each PC showing the long-term baseline shifts. Due to the scarcity of some events, while some insights can still be gained regarding the nature and impact of those events, a full characterization of the potential source or effect may not be possible. To address such

cases, all persistent baseline shifts should be investigated in scheduled periods between batch runs. The success of the maintenance actions taken would be indicated by the restoration of the baseline. Repetitive patterns can be more fully investigated. Such recurring patterns are also of highest importance as they are occurring more frequently and thus have the highest potential for downtime reduction due to unexpected failures. Learning from the related historical events can shorten the duration required for investigating the sources and thus help achieve more consistent performance during maintenance.

Overall, this analysis serves as a preliminary effort to infer equipment condition from the available process measurements. Distinct patterns were detected in the data. Historical events were used for the verification of the significance of such patterns. The combination of ICA and PCA allowed for a deeper interpretation of the connections between recorded events. The combined use of the methods allowed for maximizing the insights gained from both. The isolation of equipment-related events from the noisy PCs alone would be more difficult given the small contribution of equipment-related changes to the overall variance. The combination of methods helped identify the sources of the shifts, link them to exact recorded interventions and then map their persisting impact on the operation. Additional insights regarding the causes of the faults, the nature of the deterioration occurring, and its impact could be gained by a deeper analysis of the PCs showing the shifts, whether short- or long-termed ones. Analysis of the loading factors and the sensors contributing to those PCs, as well as the combinations of affected PCs can indicate the type and location of the developing equipment fault. This would reduce the overall time required for maintenance. Early detection and intervention can potentially save valuable production time and avoid unnecessary process downtime. Type A events can be further classified according to the required interventions. Problematic interventions, which result in big shifts in operational baselines can be identified. This supports operators and process designers to propose more convenient courses of action following such events and offer suggestions for better process and machine design. These findings enable the transition from a purely time-based maintenance scheme that is currently in place to a condition-based one.

4. CONCLUSIONS

A two-stage methodology for equipment condition assessment in drug product manufacturing mainly leveraging short-term focused process monitoring data was proposed. The methodology was applied to an aseptic filling line using data from two production years. Underlying long-term equipment-related trends could be extracted from the data.

The anomaly detection showed that while outlier events occurred throughout the production cycles, increased clusters of anomalies could be detected at the beginning and end of the production cycles between maintenance periods in some cases. Increased process understanding was achieved by visualizing outlier events and not just process failures. This finding helps decision-makers adjust the maintenance intervals instead of relying on strict time-based schedules to avoid the observed deterioration at the end of the cycle. The findings could also be used to adjust the maintenance actions taken to achieve smoother operations in the following production cycles.

The condition monitoring identified recurring patterns in the data to show long- and medium-term shifts in the operation baseline data. These shifts were linked to sources obtained from the anomaly detection and to different types of process faults. The role of these faults in shifting operational baselines was investigated. This allows decision-makers to identify deteriorating conditions earlier and apply appropriate preemptive measures, which can result in a significant reduction of non-value-adding downtime. Problematic maintenance actions can also be highlighted, which could eventually lead to improved machine, process, and operation design. Localization of the sources of variance and their impacts would help to further separate the detected condition changes, for example, separating sensor malfunction from equipment deterioration.

This work shows the applicability of predictive maintenance applications given the sensor architecture in the facility. It has been established that equipment-related changes can be inferred from the available data. Future work should further try to identify and precisely localize equipment with deteriorating conditions, address the classification of fault events, and predict failure occurrences in a more automated manner. The findings and insights gained in this work could thus lay the groundwork for predictive maintenance actions in the pharmaceutical industry. This could potentially save valuable production time, reduce unplanned downtime, avoid redundant maintenance actions, and improve process design and operation.

AUTHOR INFORMATION

Corresponding Author

Sara Badr – Department of Chemical System Engineering, The University of Tokyo, 113-8656 Tokyo, Japan; orcid.org/0000-0003-4233-1353; Email: badr@pse.t.u-tokyo.ac.jp

Authors

Philipp Zürcher – Department of Chemical System Engineering, The University of Tokyo, 113-8656 Tokyo, Japan

Stephanie Knüppel – Engineering, Science & Technology, F. Hoffmann-La Roche Ltd., 4303 Kaiseraugst, Switzerland

Hirokazu Sugiyama – Department of Chemical System Engineering, The University of Tokyo, 113-8656 Tokyo, Japan; orcid.org/0000-0002-3067-030X

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c04182>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the many collaborators from F. Hoffmann-La Roche, especially Jens Kühne and Klaus-Dieter Jochmann, for providing resources for this project. This research was funded by a research grant from F. Hoffmann-La Roche. H.S. gratefully acknowledges financial support from a Grant-in-Aid for Young Scientists (17H04964). The Top Global University Project of the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) is also gratefully acknowledged.

REFERENCES

- (1) Diez-Oliván, A.; Del Ser, J.; Galar, D.; Sierra, B. Data Fusion and Machine Learning for Industrial Prognosis: Trends and Perspectives towards Industry 4.0. *Inf. Fusion* **2019**, *50*, 92–111.
- (2) Cohen, Y.; Faccio, M.; Pilati, F.; Yao, X. Design and Management of Digital Manufacturing and Assembly Systems in the Industry 4.0 Era. *Int. J. Adv. Manuf. Technol.* **2019**, *105*, 3565–3577.
- (3) Lin, C. Y.; Hsieh, Y. M.; Cheng, F. T.; Huang, H. C.; Adnan, M. Time Series Prediction Algorithm for Intelligent Predictive Maintenance. *IEEE Rob. Autom. Lett.* **2019**, *4*, 2807–2814.
- (4) Steinwandter, V.; Borchert, D.; Herwig, C. Data Science Tools and Applications on the Way to Pharma 4.0. *Drug Discovery Today* **2019**, *24*, 1795–1805.
- (5) Kumar, S. H.; Talasila, D.; Gowrav, M. P.; Gangadharappa, H. V. Adaptations of Pharma 4.0 from Industry 4.0. *Drug Invent. Today* **2020**, *14*, 405–415.
- (6) Reinhardt, I. C.; Oliveira, D. J. C.; Ring, D. D. T. Current Perspectives on the Development of Industry 4.0 in the Pharmaceutical Sector. *J. Ind. Inf. Integr.* **2020**, *18*, No. 100131.
- (7) Food and Drug Administration FDA. Current Good Manufacturing Practice for Finished Pharmaceuticals. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=211> (accessed November 11, 2021).
- (8) Barenji, R. V.; Akdag, Y.; Yet, B.; Oner, L. Cyber-Physical-Based PAT (CPbPAT) Framework for Pharma 4.0. *Int. J. Pharm.* **2019**, *567*, No. 118445.
- (9) Zobel-Roos, S.; Schmidt, A.; Mestmäcker, F.; Mouellef, M.; Huter, M.; Uhlenbrock, L.; Kornecki, M.; Lohmann, L.; Ditz, R.; Strube, J. Accelerating Biologics Manufacturing by Modeling or: Is Approval under the QbD and PAT Approaches Demanded by Authorities Acceptable without a Digital-Twin? *Processes* **2019**, *7*, No. 94.
- (10) Sommeregger, W.; Sissolak, B.; Kandra, K.; von Stosch, M.; Mayer, M.; Striedner, G. Quality by Control: Towards Model Predictive Control of Mammalian Cell Culture Bioprocesses. *Biotechnol. J.* **2017**, *12*, No. 1600546.
- (11) Bousdekis, A.; Lepenioti, K.; Apostolou, D.; Mentzas, G. Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0. *IFAC-PapersOnLine* **2019**, *52*, 607–612.
- (12) Ding, S. H.; Kamaruddin, S. Maintenance Policy Optimization—Literature Review and Directions. *Int. J. Adv. Manuf. Technol.* **2015**, *76*, 1263–1283.
- (13) Carvalho, T. P.; Soares, F. A. A. M. N.; Vita, R.; da P Francisco, R.; Basto, J. P.; Alcalá, S. G. S. A Systematic Literature Review of Machine Learning Methods Applied to Predictive Maintenance. *Comput. Ind. Eng.* **2019**, *137*, No. 106024.
- (14) Ganesh, S.; Su, Q.; Vo, L. B. D.; Pepka, N.; Rentz, B.; Vann, L.; Yazdanpanah, N.; O'Connor, T.; Nagy, Z. K.; Reklaitis, G. V. Design of Condition-Based Maintenance Framework for Process Operations Management in Pharmaceutical Continuous Manufacturing. *Int. J. Pharm.* **2020**, *587*, No. 119621.
- (15) Zürcher, P.; Shirahata, H.; Badr, S.; Sugiyama, H. Multi-Stage and Multi-Objective Decision-Support Tool for Biopharmaceutical Drug Product Manufacturing: Equipment Technology Evaluation. *Chem. Eng. Res. Des.* **2020**, *161*, 240–252.
- (16) Meneghetti, N.; Facco, P.; Bezzo, F.; Himawan, C.; Zomer, S.; Barolo, M. Knowledge Management in Secondary Pharmaceutical Manufacturing by Mining of Data Historians - A Proof-of-Concept Study. *Int. J. Pharm.* **2016**, *505*, 394–408.
- (17) Casola, G.; Siegmund, C.; Mattern, M.; Sugiyama, H. Data Mining Algorithm for Pre-Processing Biopharmaceutical Drug Product Manufacturing Records. *Comput. Chem. Eng.* **2019**, *124*, 253–269.
- (18) Pretzner, B.; Taylor, C.; Dorozinski, F.; Dekner, M.; Liebinger, A.; Herwig, C. Multivariate Monitoring Workflow for Formulation, Fill and Finish Processes. *Bioengineering* **2020**, *7*, No. 50.
- (19) Zeberli, A.; Badr, S.; Siegmund, C.; Mattern, M.; Sugiyama, H. Data-Driven Anomaly Detection and Diagnostics for Changeover

Processes in Biopharmaceutical Drug Product Manufacturing. *Chem. Eng. Res. Des.* **2021**, *167*, 53–62.

(20) Ahmad, R.; Kamaruddin, S. An Overview of Time-Based and Condition-Based Maintenance in Industrial Application. *Comput. Ind. Eng.* **2012**, *63*, 135–149.

(21) Kumar, A.; Shankar, R.; Thakur, L. S. A Big Data Driven Sustainable Manufacturing Framework for Condition-Based Maintenance Prediction. *J. Comput. Sci.* **2018**, *27*, 428–439.

(22) de Jonge, B.; Scarf, P. A. A Review on Maintenance Optimization. *Eur. J. Oper. Res.* **2020**, *285*, 805–824.

(23) Rantanen, J.; Khinast, J. The Future of Pharmaceutical Manufacturing Sciences. *J. Pharm. Sci.* **2015**, *104*, 3612–3638.

(24) Liggan, P.; Lyons, D. Applying Predictive Maintenance Techniques to Utility Systems. *Pharm. Eng.* **2011**, *31*, 8–16.

(25) Klutz, S.; Magnus, J.; Lobedann, M.; Schwan, P.; Maiser, B.; Niklas, J.; Temming, M.; Schembecker, G. Developing the Biofacility of the Future Based on Continuous Processing and Single-Use Technology. *J. Biotechnol.* **2015**, *213*, 120–130.

(26) Rendall, R.; Chiang, L. H.; Reis, M. S. Data-Driven Methods for Batch Data Analysis – A Critical Overview and Mapping on the Complexity Scale. *Comput. Chem. Eng.* **2019**, *124*, 1–13.

(27) Rendall, R.; Reis, M. S. Which Regression Method to Use? Making Informed Decisions in “Data-Rich/Knowledge Poor” Scenarios – The Predictive Analytics Comparison Framework (PAC). *Chemom. Intell. Lab. Syst.* **2018**, *181*, 52–63.

(28) OSIsoft. PI System. <https://www.osisoft.com/pi-system> (accessed March 16, 2022).

(29) Hyvärinen, A.; Oja, E. Independent Component Analysis: Algorithms and Applications. *Neural Networks* **2000**, *13*, 411–430.

(30) Albazzaz, H.; Wang, X. Z. Statistical Process Control Charts for Batch Operations Based on Independent Component Analysis. *Ind. Eng. Chem. Res.* **2004**, *43*, 6731–6741.

(31) Nomikos, P.; MacGregor, J. F. Monitoring Batch Processes Using Multiway Principal Component Analysis. *AIChE J.* **1994**, *40*, 1361–1375.

(32) Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.