Journal of Pharmaceutical Innovation (2019) 14:359–372
https://doi.org/10.1007/s12247-018-9348-7

ORIGINAL ARTICLE

Advances in Continuous Active Pharmaceutical Ingredient (API) Manufacturing: Real-time Monitoring Using Multivariate Tools

Melanie Dumarey[1] · Martin Hermanto[2] · Christian Airiau[3] · Peter Shapland[1] · Hannah Robinson[1] · Peter Hamilton[1] · Malcolm Berry[1]

Published online: 25 October 2018

Abstract

Purpose The implementation of continuous processing technologies for pharmaceutical manufacturing has increased due to its potential to enhance supply chain flexibility, reduce the footprint of the manufacturing facility, and deliver more consistent quality. Additionally, it facilitates extensive, real-time monitoring by sensors and process analytical technology (PAT) tools without perturbing the process. In the presented case study, the use of multivariate tools for the real-time monitoring and retrospective review of a continuous active pharmaceutical ingredient (API) synthesis was evaluated from process development through to commercialization.

Method A multivariate statistical process monitoring (MSPM) approach summarizing variability in both quality critical (controlled flow rates, temperatures) and non-quality critical parameters (pressures, pump speeds, conductivity) was used to monitor three telescoped chemistry stages of a continuous API synthesis. Four different modeling strategies were presented addressing specific monitoring and analysis requirements during the pharmaceutical development lifecycle.

Results During development (R&D and commercial facility), the implemented multivariate monitoring resulted in the identification of potential failure modes, a deeper understanding of the natural process variability and accelerated root cause analysis for a recurrent reagent blockage. During manufacturing (commercial facility), the multivariate tool confirmed potential for predictive maintenance and early fault detection.

Conclusions While the implemented control strategy based on parametric control and offline analytical testing provided the required quality assurance, the multivariate trends provided additional information on process performance. More specifically, they enabled more detailed process understanding during the development of the continuous API synthesis following quality by design (QbD) principles and demonstrated the potential for enhanced process performance during commercial manufacturing.

Keywords Multivariate statistical process monitoring · MSPM · Continuous manufacturing · PAT · Process monitoring · Multivariate analysis

Introduction

The pursuit of increased efficiency to produce high quality medicines for patients has driven the implementation of continuous processes for active pharmaceutical ingredient (API) and drug product manufacturing [1–3]. The low volume equipment enables high throughput of materials and does not require the storage of intermediate materials between consecutive unit operations. The spatial footprint of continuous manufacturing can therefore be smaller than of traditional batch manufacturing. Continuous processes are also considered more environmentally friendly and economical, because they consume less solvent, consume less energy during extreme temperature processes, and achieve higher yields when an appropriate control system

✉ Melanie Dumarey
melanie.x.dumarey@gsk.com

[1] Product Development and Supply, Research and Development, GlaxoSmithKline, Gunnels Wood Road, Stevenage, SG1 2NY Hertfordshire, UK

[2] Technical Development, Global Manufacturing Site, GlaxoSmithKline, Jurong, Singapore

[3] Product Development and Supply, Research and Development, GlaxoSmithKline, Upper Merion, PA, USA

is in place [4]. These benefits are supplemented by a more consistent quality achieved by efficient mass and heat transfer, which reduces process heterogeneity. Finally, continuous processes enable manufacturing tailored to demand as batch size is no longer constrained to a given vessel size. All these major advantages make continuous processing a fast progressing innovation within the pharmaceutical world, despite some initial challenges (e.g., little regulatory precedents) [5, 6]. The International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) recently announced to support the introduction of continuous manufacturing into the pharmaceutical industry with a new ICH guideline: continuous manufacturing (Q13) [7].

In pharmaceutical industry, product quality and process performance are ensured by the implementation of a suitable control strategy as defined by the ICH Q8 and Q11 guidelines [8, 9]. Parametric, attribute, and/or procedural controls are designed based on product and process understanding and aim to control critical quality attributes within their specification limits [8, 9]. For continuous processes, this is typically achieved by the integration of feedback or feedforward loops, which control selected parameters and attributes by tuning manipulated variables (actuators) in real time. Additionally, the control strategy applied for continuous processes may prevent the collection of material generated outside the accepted ranges by the integration of divert to waste valve(s) at relevant positions within the equipment [10]. Typically, the last step of the control strategy for pharmaceutical products consists of release testing using offline analytical methods [8, 9].

A continuous process under a state-of-control exhibits constant key characteristics over time at every position within the equipment (static profile). Process consistency can therefore easily be monitored in real time [6] by a variety of tools [10–12]: from simple process sensors (e.g., flow rate, temperature, and pressure, [10]) to process analytical technology (PAT) such as online near-infrared (NIR) or infrared (IR) spectroscopy, focused beam reflectance measurement (FBRM) [11, 12], or online HPLC [13]. The resulting trends can then constitute an essential component of a control strategy in the manufacturing environment [6–9]. Additionally, the wealth of data generated is commonly used to increase process understanding supporting quality by design (QbD) [14].

A multivariate monitoring approach enables monitoring continuous process performance in a holistic way by summarizing the variability of multiple process parameters and/or attributes in a few trends [15] based on the collinear features of the investigated process data. This method is very sensitive to subtle changes in correlation patterns between the sensor readings, which are difficult or even impossible to detect when studying univariate trends only. The use of MSPM for real-time monitoring, diagnosis, or control of process performance is well established across a range of industries, e.g., in biotechnology for the supervision of a fermentation process [16], in the paper industry for fault detection for a paper board manufacturing process [17], in the petrochemical industry for fault detection and diagnostics [18], in the soft drink industry for process control [19], or in the polymer industry [20]. Similar applications within the pharmaceutical industry, however, are rarely described in literature. Machin M et al. discussed the use of multivariate statistical process control technology to monitor and control a granulation process [21], while an unsupervised monitoring application (i.e., the multivariate model does not use process output information) for drug product manufacturing was investigated by Zomer S et al. [22].

The slower uptake, mainly of unsupervised MSPM, in pharmaceutical manufacturing might be caused by the added complexity of rigid quality management systems and high regulatory scrutiny: (1) no precedent for the use of unsupervised models, (2) unclear position in the control strategy as multivariate trends implemented for process performance monitoring also highlight process changes not affecting product quality, (3) requires additional resource at the manufacturing site for model maintenance, and (4) provides little added benefit for quality assurance within the current ways of working heavily reliant on analytical testing. Additionally, API process data is often scarce at the start of commercial manufacture and insufficient to build a robust multivariate model for monitoring.

The case study presented in this paper, evaluates the use of multivariate monitoring as a complementary tool to the formal control strategy (i.e., a combination of parametric process control and offline analytical testing) applied during continuous API manufacturing. While the traditional control strategy focusses on quality assurance, the multivariate monitoring aims to increase process understanding during development and process performance during manufacturing. For this purpose, a generic MSPM methodology based on principal component analysis (PCA [23]) was evaluated from early development, when little process experience was available, until the start of commercial manufacture. Four implementation modes were discussed targeting different requirements during the pharmaceutical development lifecycle: (1) understand natural process variability at state-of-control (i.e., common cause variability) with respect to the implemented control limits, (2) identify failure modes and increase process understanding, (3) ad hoc root cause analysis, and (4) real-time monitoring of process performance and fault diagnosis during flexible commercial manufacturing.

Methods

Process Set up

The continuous API synthesis consisted of five chemical transformations. Two consecutive transformations in the stage

1 reactor (reactor 1) were followed by the removal of excess gas (reagent 2) in the stripping column. The resulting material was collected in a buffer pot and then pumped through reactors 2–3 for the next three chemical transformations. Note that the selectivity of the last reaction was controlled by the low jacket temperature (approximately − 5 °C) in reactor 3, while the increased jacket temperature (approximately 30 °C) in reactor 4 ensured completion of the reaction. The crude API was collected in a holding tank before it was crystallized as a batch process. The complete API synthesis process scheme is presented in Fig. 1. Note that two separate valves enabled the diversion to waste of non-conforming material during the process: one in stage 1 before the stripping column and one after reactor 4 before the material was collected in the holding tank (Fig. 1).

Process Sensors

In the R&D facility, the process as presented in Fig. 1 was monitored by a large variety of sensors (29 sensors for stage 1 and 40 sensors for stages 2 and 3, see Table 1 in Appendix):

- Temperature was monitored by multiple sensors in the reactors (Fig. 1). Note that sensor redundancy was built in to enable the differentiation between sensor failures and real process disruptions [24].
- The heat transfer fluid (HTF) temperature (Fig. 2) was measured at the reactor supply and return. The supply temperature was used to maintain the temperature at set point and as an action limit (univariate) for divert to waste events as defined in the control strategy ("Process Control" section).
- Sensors measuring flow rate and pump speed were used to maintain a constant flow rate (Fig. 2). The flow rate measurements were also used to trigger diversion to waste in case of excursions exceeding the action limit as defined in the control strategy ("Process Control" section).
- Pressure sensors in the pumps protected equipment from pressure damage via feedback loops and univariate alarms (Fig. 2). Additionally, the pressure in each reactor was monitored.
- Pressure in the gas stripping column was measured at the inlet and outlet (Fig. 1).
- Conductivity sensors after each chemical transformation and at the inlet of reactor 1 provided a highly sensitive, but not selective indication of process consistency (Fig. 1).

The sensor set up in the commercial equipment followed the same principles, but small modifications were made reducing redundancy and optimizing information captured by the sensor readings (e.g., target biggest potential temperature and conductivity changes):

- For all pumps, the pressure and filter inlet, outlet and differential pressures were monitored instead of suction and discharge pressures
- In stage 1, conductivity was only measured at the outlet (no inlet and content conductivity), while in stage 2, an additional conductivity probe was placed at the reactor 2 inlet
- The number of temperature sensors was reduced but still provided redundancy (2 fewer in reactor 1, 1 fewer in reactor 2)
- Sensors monitoring reagent 6 flow control output, reactor 2 outlet pressure and reactor 4 back pressure were added

Process Control

For the discussed case study, a state-of-control was declared when all controlled process parameters had remained within the predefined ranges for a predefined time. These ranges and timings were based on process understanding (e.g., kinetics and residence time distribution) and risk assessment activities linking process variability to quality as part of the control strategy definition. Similarly, action limits to divert non-confirming material to waste after chemistry stages 1 and 3 were defined considering both the magnitude and the duration of a process parameter disturbance (Fig. 1).

The process was maintained at state-of-control by feedback loops manipulating the appropriate actuators (e.g., pump speed) to control selected process parameters (i.e., flow rate, pump pressure, and temperature of the HTF at supply) as visualized in Fig. 2. At the R&D site, this was achieved with single input—single output proportional-integral-derivative (PID) controllers, while at the manufacturing site, a ratio control for flow rate was implemented within a cascade set up (i.e., the reagent flows were controlled relative to the main flow).

At both facilities, the output material was monitored by an online HPLC method, which could separate all process related impurities. Eventually, the API was released downstream, i.e., after the crystallization step (batch), against specifications as described in the control strategy (e.g., assay, impurity level, ID test, and solvent content)

Multivariate Monitoring Approach

Selection of Representative Process Data

Model sensitivity is critical for a multivariate monitoring approach to ensure that all relevant process drift or changes (i.e. special cause variability) are detected. Therefore, the

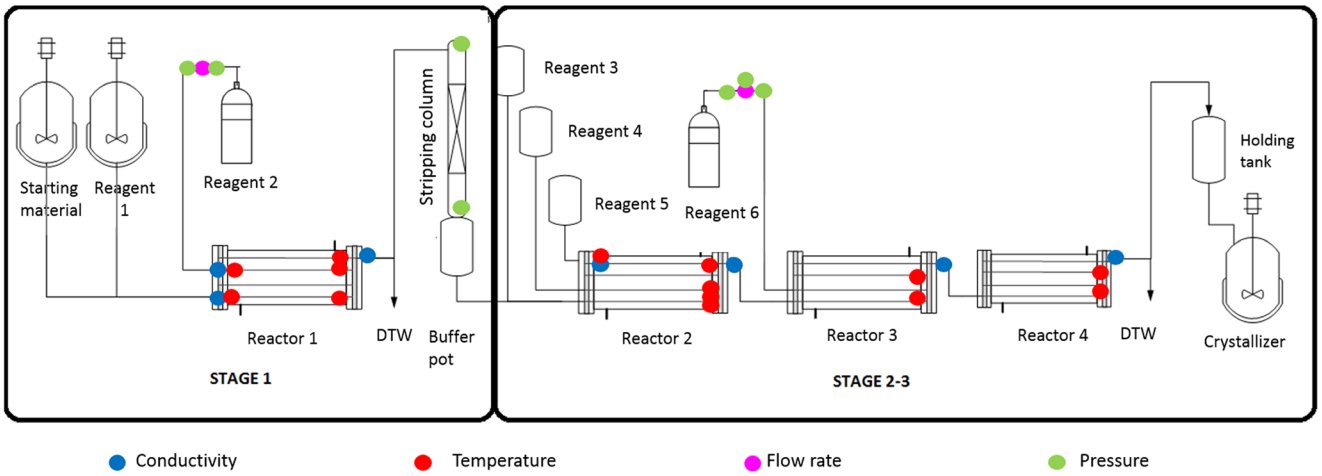

Fig. 1 Overview of continuous API synthesis with marked sensors and divert to waste (DTW) valves (excluding feedback loops and pump sensors)

calibration data used to build the PCA model excluded all process sensor readings obtained at a time when the process was not under a state-of-control, i.e., during start up, during shut down and prior to or during a disturbance (as defined in the "Process Control" section). Such observations would introduce additional variability widening the model limits and reduce sensitivity to process changes.

A multivariate monitoring model also needs to be robust to small, natural process variability (common cause variability) in order to minimize false alarms as these would require unnecessary interventions by operators. Therefore, the calibration model should include all expected common cause variability. However, during some of the development campaigns presented in this paper, the processes were run only for a short duration at each experimental condition and therefore the calibration sets were based on the largest representative data set available at the time (i.e., from 2 until 192 h).

Preprocessing

Preprocessing is critical to ensure that a multivariate model captures the relevant trends and correlation patterns in the data, while filtering out noise. At the start of process development, little information was available on failure modes and process variability. Therefore, all process sensor readings were given equal weight by scaling them to unit variance (i.e., division by the observed standard deviation for the process sensor). As required for PCA modeling, all process sensor readings were also mean centered.

Model Development

A standard MSPM model, which summarizes multiple, collinear process sensor readings in fewer dimensions by principal component analysis (PCA), was implemented for real-time monitoring [23, 25]. A PCA model projects the calibration data in a lower dimensional space maximizing the variance captured in the newly calculated latent variables, i.e., principal components (Fig. 3a). Mathematically, this is achieved by a singular value decomposition producing one matrix of scores T (i.e., the projections of the observations in the space defined by the principal components) and one matrix of loadings P (i.e., the weight of the process sensors for all selected principal components) as shown in Eq. 1.

$$X = T\,P^T + E, \tag{1}$$

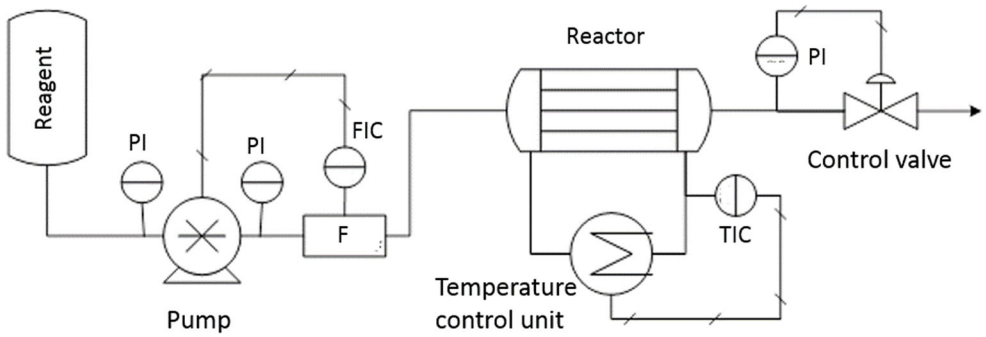

Fig. 2 Overview of temperature, pressure, and flow rate control for one reactor. PI, pressure indicator; FIC, flow indicator controller; F, flow meter; and TIC, temperature indicator controller. Note that HTF supplied to the reactor was running through a recycling loop managed by the temperature control unit

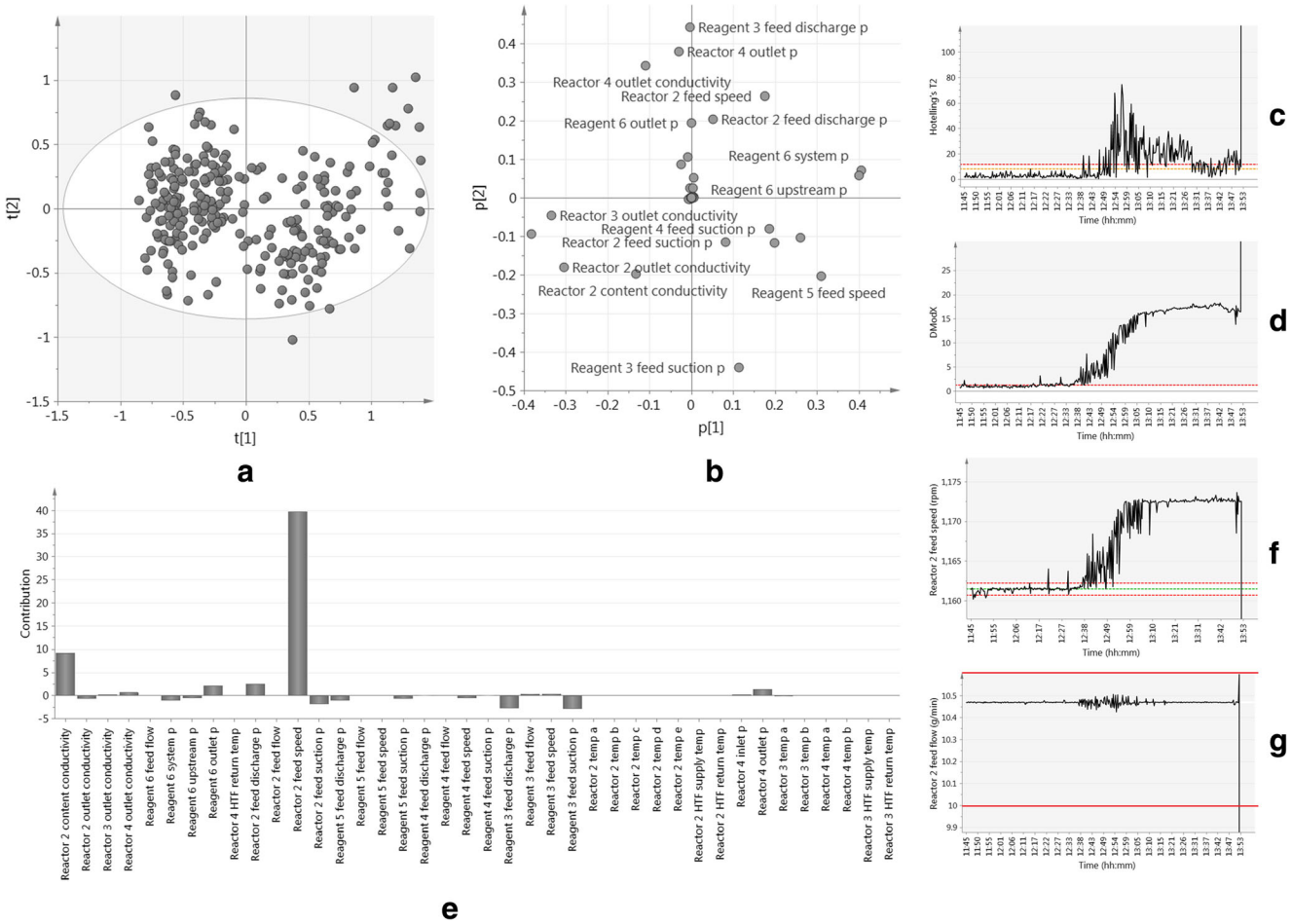

Fig. 3 Multivariate monitoring process for a continuous API synthesis displaying symptoms of a pump blockage which resulted in a flow rate disturbance and a pump switch at 13:52: the operational space projected on principal components 1 and 2 defined by the stages 2–3 sensor data from 02FEB2016 10:45 till 02FEB2016 12:16 (**a**); the corresponding loading plot (**b**); Hotelling's T^2 trend based on three principal components with the 99 and 95% statistical limit marked by the red and orange dotted line (**c**); the DModX trend based on three principal components with the critical limit marked by the red dotted line (**d**); contribution plot based on the increase of Hotelling's T^2 trend at 12:50 (**e**); univariate trend of reactor 2 feed pump speed with three times standard deviation marked in red and average trend in green (**f**); and univariate trend of reactor 2 feed flow rate with proposed accepted ranges marked in red (**g**)

where X is the ($m \times n$) data matrix containing all sensor readings (n) over several time points (m), T is the (m x k) score matrix, P is the ($n \times k$) loading matrix, and E is the model residual. ^{T} denotes transposal of the matrix and k is the number of selected principal components in the model.

A separate PCA model was built for stage 1 and stages 2–3 based on the selected, pre-processed data as described above using Simca v13.0 (Umetrics, Umeå, Sweden), because the physical separation introduced by the buffer pot (Fig. 1) broke the correlation patterns between the sensor readings. The number of model components was selected based on cross validation. However, due to the correlated nature of the time series data describing continuous processes, it was easy to predict scores for sensor readings close in time, which resulted in over fitting models. Therefore, a heuristic approach was taken to exclude the last component(s) selected by cross validation if the explained variability was below 1%. This fit for purpose approach enabled the fast processing of large data sets within a validated environment.

Monitoring

The calibration models as constructed in the "Model Development" section were used to calculate two multivariate monitoring diagnostics in real time: Hotelling's T^2 and DModX. Hotelling's T^2 (Fig. 3c) is the generalized distance of the predicted score from the mean score of the historical calibration set [26, 27]. It detects variability greater than historically observed common cause variability. The statistical control limit for Hotelling's T^2 was calculated from the calibration data ("Model Development" section) at a 95% significance level assuming a beta distribution [26, 27]. The DModX (aka the Q residual, Fig. 3d) is the distance of the real-time sensor measurements from the PCA model space [26] and detects new variability patterns, which were not captured by the PCA model.

When an unexpected process trend or excursion outside the control limit was observed for one of the multivariate diagnostics, a contribution plot (Fig. 3e) was generated in real time to investigate the root cause. The contribution of one process sensor reflects the difference between its value at the time of excursion and its historical average as calculated in Eq. 2a and 2b [26].

$$\text{Contribution}\left(\text{Hotelling's } T^2\right) = \Delta X \times \text{weight} \quad (2a)$$

$$\text{Contribution}\,(\text{DModX}) = e_k \times \text{weight} \quad (2b)$$

where the weight is the square root of the explained sum of squares for each variable, ΔX is the difference between the sensor reading at excursion and its historical average, and e_k represents the residual of the variable. As a consequence, the magnitude of the bars in the plot is proportional to the contribution of the process sensor to the changing process trend, while the direction of the bar indicates whether the contribution is positive or negative. The process sensors with the largest bars were then investigated in more detail by visualizing their individual trends (Fig. 3f). It should be noted that flexible, real-time data visualization is critical to ensure fast information review during manufacturing by non-experts.

Informatics Set up

Process sensor readings were recorded by a process control system, i.e., Labview (National Instruments Corporation, Austin, USA) in R&D or Delta V (Emerson, St. Louis, USA) at the manufacturing site, and collected by a SIPAT collector station (SIMATIC SIPAT v4.1, Munich, Germany), which sent the data to the data base server. The calibration data were extracted from this server via the data miner of the SIPAT client and then imported to SIMCA for offline PCA model construction. Once the model was finalized, it was uploaded to SIMCA-online v13.3 (Umetrics), where the Hotelling's T^2 and DModX values were calculated from sensor readings in real time. SIMCA-online directly receives the sensor readings from the SIPAT collector station at the programmed frequency, which was 20 s for the discussed application. This setting was supported by statistical analysis, which confirmed that no information was lost reducing the frequency from every second (equipment capability) to every 20 s (minimizing the required data base space on the server): the mean, standard deviation and area under the curve for each process sensor was not significantly different between 1 and 20 s frequency data at and outside state-of-control. Additionally, a review of multivariate models acquired at different frequency demonstrated that the trends were similar and detected the same process changes, because the process changes leading to process disturbances (as defined in the "Process Control" section) were not instantaneous and took minutes rather than seconds.

Results and Discussion

Explore Common Cause Variability at State-of-Control

The MSPM methodology as described in the "Methods" section was initially implemented during early process development to gain more insight in the common cause variability inherent to the continuous process.

From multivariate trends (not shown), it was observed that not all common cause variability was random for the studied continuous API synthesis: the pressure sensors monitoring the feed pumps exhibited continuously decreasing readings as the fill level of the feed vessel dropped and a step change was observed when the feed vessel was replaced (Fig. 4). This non-random variability was very small (approximately 0.02 barg for the sensor shown in Fig. 4) and did not affect process performance, but did raise an excursion from the statistical control limit.

Secondly, it was observed that small, irrelevant temperature changes (0.01 °C) caused false alarms in Hotelling's T^2 plot (Fig. 5a). This is caused by the fact that the implemented feedback loops controlled temperature and flow rate very tightly, and consequently the observed variability at state- of-control was much smaller than the ranges conferring acceptable product quality. This resulted in an overly tight control limit calculated from the multivariate model, when based on a limited amount of process data.

Lastly, it was observed that most of the common cause variability was random within the well-controlled manufacturing process, which was confirmed by the low amount of variance captured by the supporting PCA models, e.g., the model presented in Fig. 3a contained three principal components explaining 10, 4, and 3% of the total process variability. These low values are indicative of the lack of structured changes (special cause variability) in process parameters compared to batch processes, where process parameters change over time following certain correlation patterns [26]. However, the correlation structures identified in the loading plot of the model (Fig. 3b), were considered scientifically relevant: high correlation between system and upstream pressure for reagent 6, correlation of all stage 2 and 3 conductivity measurements (all located in the same quadrant), etc. Moreover, during extensive monitoring activities over extended time frames (i.e., external validation), it was observed that models capturing only a low amount of variability were still able to detect relevant process disruptions during process development. It should also be noted that new variability patterns not captured by the PCA model are captured by the DModX diagnostic.

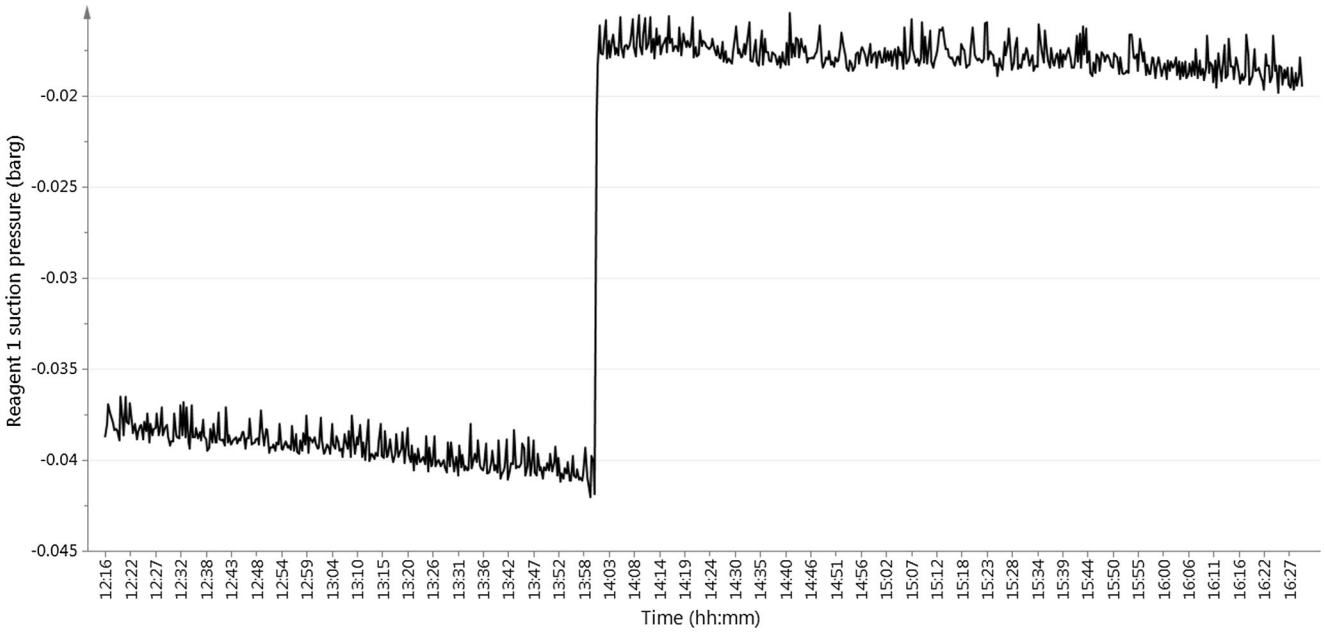

Fig. 4 Natural drift of reagent 1 pump suction pressure observed as the feed vessel was emptied followed by a step change around 14.00 when the feed vessel was replaced

Increase Process Understanding and Identify Failure Modes

During an extensive development campaign at the R&D facility spanning a 4-month period, the process settings were stretched in a multivariate way to assess their impact on API quality. The MSPM models based on common cause variability only were rebuilt for each process setting to target simplicity and maximal sensitivity. The modeling strategy was adjusted based on the knowledge gained during the exploration activities as described in the "Explore Common Cause Variability at State-of-Control" section. Ideally, a full cycle for each feed vessel (i.e., approximately 24 h) was included in the calibration model to avoid false alarms by natural pressure drift or feed vessel changes. However, this was practically impossible as most settings were only run for one day. Secondly, the temperature and flow rate sensor readings were custom scaled to address the limited variability experienced during early development. Therefore, all flow rate and temperature sensors (italicized in Table 1) were down-weighted in the model by dividing the readings by the accepted standard deviation (SD), i.e., one fourth of the one-sided accepted range assuming normal distribution (Eq. 3), instead of the observed standard deviation.

$$\text{Acceptable SD} = (\text{defined upper limit–set point})/4 \qquad (3)$$

The custom scaling eliminated the excursion from the Hotelling's T^2 trend (Fig. 5b) and increased the sensitivity to pressure changes as reflected by the appearance of short spikes caused by online HPLC sampling and short pressure disruptions. These short spikes would be easier to manage by fine tuning the alarm settings (e.g., only trigger an alarm after three consecutive trend excursions) than extended multivariate trend excursions caused by small temperature or flow rates shifts. The custom scaling thus enabled monitoring failure modes related to flow (e.g., pump blockage) and temperature (e.g., leak in the HTF pipes) control, while avoiding false alarms caused by small shifts in these process parameters.

Although the models included limited processing time frames (typically 2–4 h), they still provided detailed process insights for the development team and detected meaningful process changes in real time. More specifically, several failure modes were identified during early development at a time of limited operation experience: two types of pump blockages and one recurrent blockage of a reagent supply.

Note that during this development campaign, no actions were taken based on the multivariate trends as the methodology was still under evaluation. This approach delivered the advantage that the effect of a detected process change could be investigated over a longer time to confirm the initial interpretation of the multivariate trend change. As such, these data are critical for model validation or verification purposes, when implementing multivariate monitoring for predictive maintenance in a good manufacturing practice (GMP) environment at a later stage.

As expected, the models based on limited time frames did result in multiple false alarms, i.e., the multivariate trend (Hotelling's T^2 or DModX) exceeding the statistical test limits without any process implication (e.g., Fig. 3d). In a development environment, this was not considered as an issue as no decisions were based on the multivariate trends. Moreover,

high model sensitivity for maximal information extraction is considered more important than model robustness at this stage of the pharmaceutical development life cycle.

One example of a failure mode detected during the R&D campaign was the blockage of the pump feeding the material from the buffer pot to reactor 2. A PCA model (built as described above) summarizing natural variability in process sensor readings in stages 2–3 for 1.5 h was used to monitor the continuous synthesis in real time from 12:16 onwards. Although the system was not pronounced as under a state-of-control at this point following the definition in the "Process Control" section, the process sensor readings were stable within the operational space defined by three principal components of the model (Fig. 3a). The DModX trend soon exceeded the statistical limit (Fig. 3d), but the contribution plot (not shown) did not reveal any meaningful changes in sensor trends. Therefore, it was decided that these small, initial excursions were false alarms caused by a lack of model robustness as explained earlier. From 12:40 onwards, a larger trend change was observed for DModX and also a departure from the normal operating space was observed in Hotelling's T^2 trend from 12:50 onwards (Fig. 3c). The contribution plot (Fig. 3e) trend revealed that an increase in stage 2 feed pump speed (Fig. 3f) caused the drift of the Hotelling's T^2 trend. This root cause was confirmed by the contribution plot for the DModX trend (not shown). A higher pump speed was thus needed to maintain the flow rate at the set point by an integrated feedback loop, which indicates a blockage. In Fig. 3g, the flow rate was confirmed to be still within the accepted ranges (red lines) at the time of the multivariate alarm, but the noisier signal did indicate some disturbance of the reagent flow. At 13:52, a large flow rate excursion was observed, which triggered the pump alarm and confirmed the pump blockage hypothesis. The engineers then switched from the duty to the standby pump to maintain appropriate delivery of feed solution to the stage 2 reaction and diverted the product to waste for a predefined time (as defined in the "Process Control" section) ensuring no compromised material ended up in the API.

In this example, the multivariate monitoring approach detected the upcoming pump failure approximately 1 h before

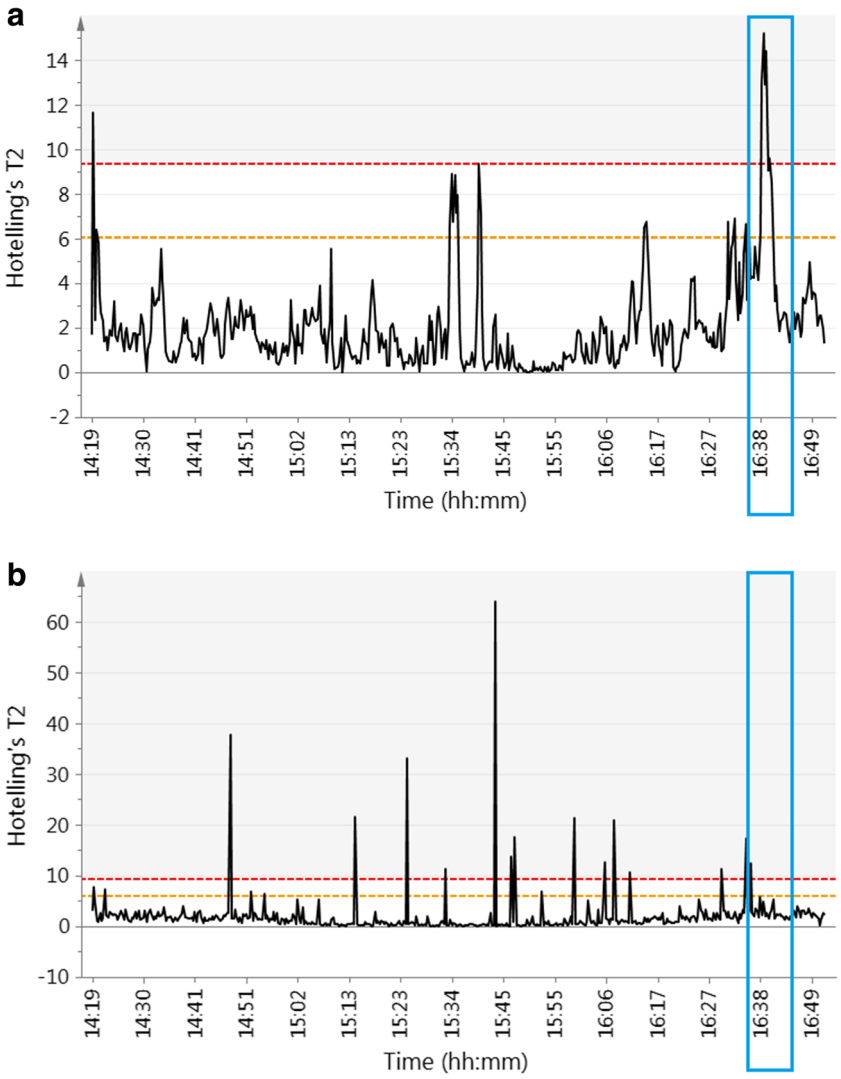

Fig. 5 Hotelling's T^2 trend summarizing stage 2–3 process sensor variability over approximately 2.5 h based on data scaled to unit variance (**a**) and custom scaled data (**b**). The blue box indicates the time frame of the small, insignificant temperature change

the univariate pump flow alarm, which would have provided sufficient time to perform corrective actions avoiding the controlled process parameter to exceed the accepted ranges and triggering divert to waste.

The described MSPM approach was later applied during continuous synthesis of the same API as part of a development campaign at the manufacturing plant, which exhibited a slightly different process sensor set up (see the "Process Set up" section). Therefore, a new model was constructed at the start of the campaign and when process settings were adjusted. This resulted in similar observations as before: (1) small continuous drift in pump sensors (pressure and speed) over time, (2) a full feed vessel cycle needed to be included in the model to ensure robustness (i.e., avoid false alarms), and (3) the multivariate trends successfully detected process changes. Note that again the multivariate monitoring was under evaluation and no actions were taken based on the multivariate trend observations. This enabled a fair comparison with the current manufacturing process monitored by univariate alarms for flow rate and temperature only and provided data for further model verification activities.

In three instances, an upcoming blockage of reagent 2 supply (gas) was detected 1 to 23 h before the flow rate alarm was activated. Another two upcoming process failures, i.e., pump blockages were detected using multivariate monitoring approximately 20 min before the flow rate exceeded the accepted ranges. The root causes for the detected process changes were fluctuations in the flow rate and increased filter pressure in one of the reagent pumps. Finally, a process disruption caused by reagent depletion was also detected using multivariate monitoring 13 min before the flow rate exceeded accepted ranges, and a decrease in reagent supply pressure was correctly identified as the root cause.

All these observations confirmed the value of multivariate tools for the monitoring of process performance in real time. Firstly, the increased process understanding could be used to optimize the control strategy definition and inform the selection of relevant statistical process control charts (SPC). Secondly, the early fault detection highlighted their potential for predictive maintenance during routine manufacturing. These benefits could be exploited to achieve increased process performance for API manufacture.

Event Root Cause Analysis

The implemented informatics set up also proved useful as a data source for a detailed root cause analysis of a recurrent reagent blockage, which was flagged by multivariate monitoring. Therefore, the regions before seven different spikes observed in Hotelling's T^2 (Fig. 6) were selected from a time point where the trend was stable to the time point before the actual spike (2–16 min depending on the spike), and a new PCA model was built to identify trends and correlation patterns for all process sensors. Note that for this application, the set up differed from the earlier approach: the PCA model is used to identify structured variability (i.e., trends or outliers) within the selected data set in a retrospective manner instead of comparing it to historical data in real time. This approach ensured the model detects new correlation patterns, which were not captured by the historical model based on common cause variability, in the most sensitive way.

The resulting score plot (Fig. 7a) revealed that for most instances, little process variability was observed prior to the spike as most observations are clustered around the center, except for the first spike appearing at 01:57. The corresponding loading (p) plot (Fig. 7b) revealed that the largest variability before the spike at 01:57 was observed for the pressure sensors located in the reagent supply (reagent 2 downstream pressure) and the degassing column (change of 0.02 atmospheric bar considered irrelevant) as they are located at the outskirts of the plot. This implied that the blockage was not originating from the main flow path of the reactors: low variability in reactor pressures confirmed by their central location in the loading plot. The score plots based on principal components 3–5 were investigated in an analogous way (not shown), but did not reveal additional trends. This multivariate analysis thus reduced the number of investigations needed to drill down to the actual root cause, which was finally identified as a local mixing effect at the point of reagent 2 addition. This was permanently addressed by a small equipment modification at the manufacturing site, which successfully prevented the blockage of reagent 2 supply in all further manufacturing runs to date.

Real-Time Fault Diagnosis and Monitoring of Process Performance

During the first commercial campaign, the process was run at different productivity rates enabling maximum production flexibility. This was achieved by independently varying the main feed flow rates of stage 1 and stages 2–3 maintaining a constant fill level in the buffer pot. Note that the ratio between the main feed and reagent flow was kept constant at all times.

The described multivariate monitoring methodology was not practical to monitor this flexible processing mode as a model update would be required at every productivity change adjusting for the changed flow rates and correlated parameters. While flow rates could be monitored as a deviation from their set point to avoid model excursions when changing the process productivity, it was impossible to correct other sensor readings, because there is no quantitative model describing the effect of flow rate change on pump speed, pressure, and heat exchange. The option of progressively building an all-encompassing MSPM model was deemed impractical, as no

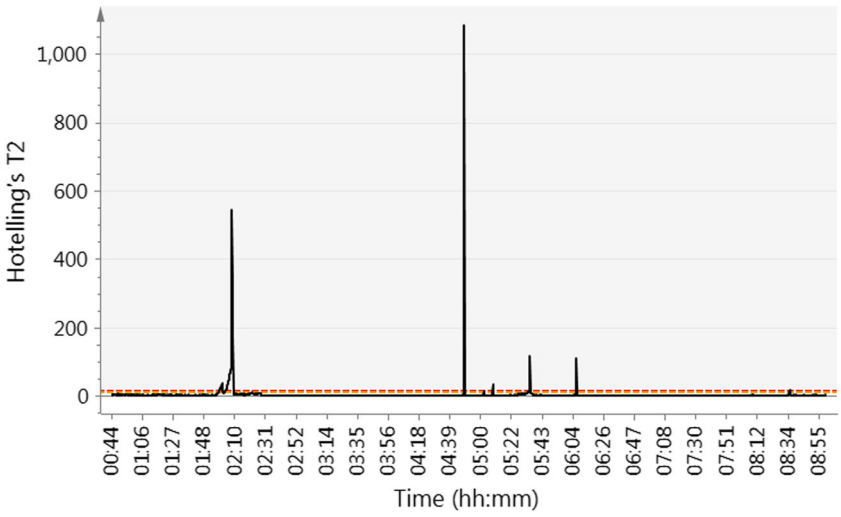

Fig. 6 Spikes in Hotelling's T^2 trend caused by blockage of the reagent 2 supply in stage 1

monitoring would be feasible until sufficient information is gathered at each newly tested productivity rate. Therefore, it was decided to model the difference between the current time point and the previous time point for each sensor instead of the actual sensor readings, i.e., focusing on process variability instead of absolute process sensor readings. The supporting models explained only a small fraction of common cause variability as was observed for the multivariate monitoring during development ("Explore Common Cause Variability at State-of-Control" section). The adjusted monitoring approach successfully delivered conforming multivariate trends while at state-of-control regardless of the applied productivity rate (Fig. 8). Additionally, the observed drift in pump pressures as feed bottles were emptying and small temperature or flow rate changes ("Explore Common Cause Variability at State-of-Control" section) did not cause continued excursions in the multivariate trends. This removed the need for custom scaling and the inclusion of a whole feed bottle cycle as applied during the process development stage of the product lifecycle. A feed or productivity change did cause a temporary spike (Fig. 8) in the multivariate trend and therefore monitoring was paused during these specific events.

While this methodology was fit for purpose during routine manufacturing, it would have been less informative for real-time process insights as targeted during the development campaigns. This is caused by the fact that the model captures relations between sensor variabilities instead of actual sensor values, which complicates the interpretation of the sensor contributions and reduces the sensitivity to small, continued process drift.

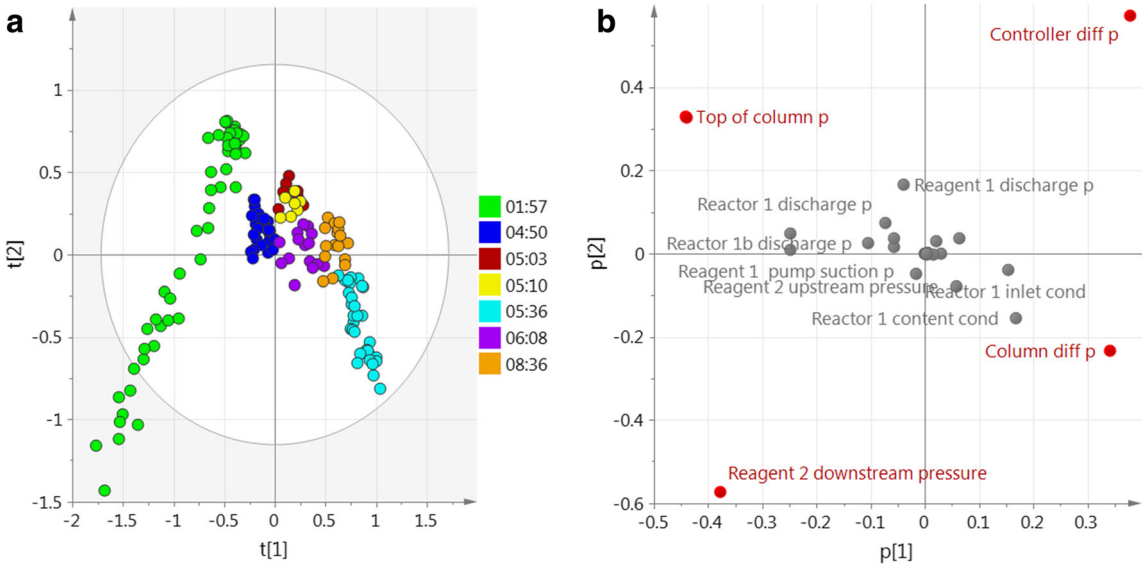

Fig. 7 Score (**a**) and loading (**b**) plot of the PCA summarizing stage 1 sensor variability during selected time frames before observed spikes (Fig. 6) in reagent flow. In the score plot, observations are colored by the time when the spike in reagent flow occurred, while in the loading plot, the process sensors exhibiting highest variability prior to reagent 2 blockage (most extreme locations in the plot) are marked in red.

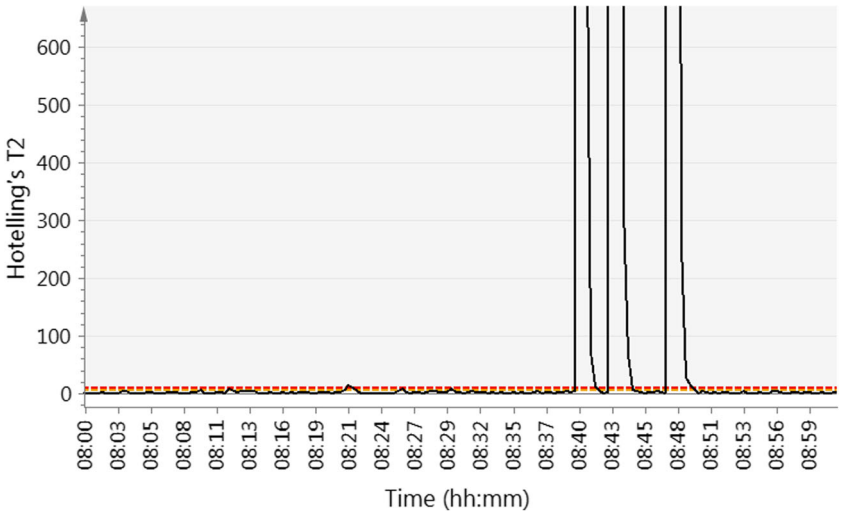

Fig. 8 Hotelling's T^2 trend of one commercial stage 1 process with three peaks occurring during the performed productivity decrease (i.e., decrease in flow rate set points in three steps) from 08:41 am till 08:51 am on 29th of November 2017

Other Advantages of Multivariate Monitoring

During all described development and commercial campaigns, multivariate monitoring was implemented by specialized chemometricans as part of an evaluation exercise. Chemists and engineers were closely involved in the selection of relevant process sensors and data, and linked detected changes in multivariate trends to observed changes in the process. The advanced visualization interface available to the whole matrix team enabled both a process overview based on multivariate summary trends and an interactive drill down to evaluate individual sensor readings. This encouraged chemists, engineers, and operators to monitor the process more closely and resulted in more real-time, data-driven process understanding activities during process development. More specifically, recurrent process sensor patterns or disruptions detected by multivariate trends were investigated that may have previously been ignored, e.g., pump flow rate was more variable as the feed vessel emptied or a reagent 2 cylinder change introduced bubbles to the system affecting both pressure and conductivity sensors.

Discussion

It was shown that multivariate monitoring adds value from early development of continuous API synthesis processes until commercial manufacturing and provides the flexibility to create fit for purpose methodologies for the various stages of the development life cycle. Although the potential for predictive maintenance preventing divert to waste or forced shut down during commercial manufacturing was demonstrated by many observations (i.e., external test sets), more work is needed to enable routine implementation by operators.

Firstly, alarm settings and corresponding decision trees must be developed such that the operator can make an unambiguous decision to act upon a changing multivariate trend. Hereby, it is important to minimize false alarms and unnecessary actions by the operator, as these would complicate operations and increase the risk of human error.

Secondly, routine implementation within the operational control room requires model verification and maintenance following GMP standards. The supporting procedures will differ from the approaches for established PAT predictive models, e.g., models to predict tablet assay from NIR spectra. Firstly, a quantitative estimation of accuracy and precision is not possible as there are no reference measurements available. Secondly, PAT models are typically built around an intentional stretch of process conditions across a predefined range, whereas the models described in this paper are based on variability around the target conditions. The main implication of the resulting higher model sensitivity is the increased likelihood of frequent model updates for multivariate monitoring to include changed variability patterns within the design space, e.g., after a pump change, change of input materials, change of weather conditions, and therefore requiring a novel model maintenance approach compared to established PAT applications. It is expected that the frequency of required updates will reduce over time as the model captures more and more common cause variability.

All the models described in the paper would be classified as "low impact" models as per ICH Q8/Q9/Q10 definition [28], because they were not used as an indicator of quality. As the supporting models were based on common cause variability observed at fixed set points rather than on accepted variability within the defined design space, they typically flag different types of process changes.

Conclusions

Multivariate trends summarizing many collinear process sensor readings by PCA enabled real-time monitoring of the

continuous synthesis of API during development and commercial manufacturing activities. The flexibility of well-established MSPM tools enabled designing a low resource, fit for purpose implementation mode for each stage of the product development life cycle despite a limited amount of process experience.

While the implemented control strategy based on parametric control and offline analytical testing provided the required quality assurance, the multivariate trends provided additional information on process performance. During development, this resulted in the identification of potential failure modes, a deeper understanding of the natural process variability and accelerated root cause analysis for a recurrent reagent blockage. During manufacturing, the multivariate tool demonstrated a potential for predictive maintenance as it flagged upcoming process disturbances 10 min to few hours before the controlled parameters (i.e., flow rate or temperature) exceeded their accepted ranges. This time frame would enable corrective interventions (e.g., pump switch) to avoid a divert to waste event and/or forced shut down of the process. More representative data is required to develop a robust implementation strategy for routine monitoring focusing on meaningful alarm limits and unambiguous decision trees for the supporting operators.

Finally, it was also demonstrated that multivariate models based on real-time process analytics provided an increased scrutiny on the quality of the manufactured products by maximizing the use of all relevant data generated on the line. This drove the reduction of process variability and the increase of process robustness ensuring continued delivery of product quality for the patients. Although the current applications were focused on process understanding, predictive maintenance, and continuous process improvement, the long-term objective should be real time, automated process optimization, and control. This could further enhance the quality of material generated by continuous processes, while enabling the highest level of manufacturing efficiency and flexibility.

Acknowledgements We would like to thank Erik Johansson for the support on custom scaling of the process sensors. We would like to thank Ian Barylski for the informatics platform support.

Appendix

Table 1 Overview of sensor set up for the monitoring of a continuous API synthesis at the R&D pilot plant listing the applied scaling factor. Controlled parameters are marked in italics. The accepted SD was calculated using Eq. 3

Stage 1		Stage 2–3	
Sensor description	Scaling	Sensor description	Scaling
Reactor 1 inlet conductivity	SD	Reactor 2 feed discharge p	SD
Starting material discharge pressure	SD	*Reactor 2 feed flow*	Accepted SD
Starting material feed flow	Accepted SD	Reactor 2 feed speed	SD
Starting material pump speed	SD	Reactor 2 feed suction p	SD
Starting material suction pressure	SD	Reagent 3 feed discharge p	SD
Reagent 1 discharge pressure	SD	*Reagent 3 feed flow*	Accepted SD
Reagent 1 feed flow	Accepted SD	Reagent 3 feed speed	SD
Reagent 1 pump speed	SD	Reagent 3 feed suction p	SD
Reagent 1 suction pressure	SD	Reagent 4 feed discharge p	SD
Reactor 1 content temperature a	Accepted SD	*Reagent 4 feed flow*	Accepted SD
Reactor 1 content temperature b	Accepted SD	Reagent 4 feed speed	SD
Reactor 1 content conductivity	SD	Reagent 4 feed suction p	SD
Reagent 2 feed flow	Accepted SD	Reactor 2 content conductivity	SD
Reagent 2 upstream pressure	SD	Reagent 5 feed discharge p	SD
Reagent 2 downstream pressure	SD	*Reagent 5 feed flow*	Accepted SD
Reactor 1 content temperature c	Accepted SD	Reagent 5 feed speed	SD
Reactor 1 content temperature d	Accepted SD	Reagent 5 feed suction p	SD
Reactor 1 content temperature e	Accepted SD	*Reactor 2 temperature a*	Accepted SD
Reactor 1 discharge pressure	SD	*Reactor 2 temperature b*	Accepted SD
Reactor 1 outlet pressure a	SD	*Reactor 2 temperature c*	Accepted SD
Reactor 1 outlet pressure b	SD	*Reactor 2 temperature d*	Accepted SD

Table 1 (continued)

Stage 1		Stage 2–3	
Sensor description	Scaling	Sensor description	Scaling
Reactor 1 supply temperature	Accepted SD	*Reactor 2 temperature e*	Accepted SD
Reactor 1 return temperature	Accepted SD	*Reactor 2 HTF supply temperature*	Accepted SD
Reactor 1 outlet conductivity	SD	*Reactor 2 HTF return temperature*	Accepted SD
Top of column pressure	SD	Reactor 2 outlet conductivity	SD
Bottom of column pressure	SD	*Reagent 6 feed flow*	Accepted SD
Column differential pressure	SD	Reagent 6 system p	SD
Controller differential pressure	SD	Reagent 6 upstream p	SD
N2 flow to stripper	SD	Reagent 6 outlet p	SD
		Reactor 3 temperature a	Accepted SD
		Reactor 3 temperature b	Accepted SD
		Reactor 3 HTF supply temperature	Accepted SD
		Reactor 3 HTF return temperature	Accepted SD
		Reactor 3 outlet conductivity	SD
		Reactor 4 inlet p	SD
		Reactor 4 temperature a	Accepted SD
		Reactor 4 temperature b	Accepted SD
		Reactor 4 HTF return temperature	Accepted SD
		Reactor 4 outlet p	SD
		Reactor 4 outlet conductivity	SD

References

1. Poechlauer P, Manley J, Broxterman R, Gregertsen B, Ridemark M. Continuous processing in the manufacture of active pharmaceutical ingredients and finished dosage forms: an industry perspective. Org Process Res Dev. 2012;16:1586–90.
2. Anderson NG. Using continuous processes to increase production. Org Process Res Dev. 2012;16:852–69.
3. Nasr MM, Krumme M, Matsuda Y, Trout BL, Badman C, Mascia S, et al. Regulatory perspectives on continuous pharmaceutical manufacturing: moving from theory to practice September 26–27, 2016, International symposium on the continuous manufacturing of pharmaceuticals. J Pharm Sci. 2017;106(11):3199–206.
4. Schaber SD, Gerogiorgis DI, Ramachandran R, Evans JM, Barton PI, Trout BL. Economic analysis of integrated continuous and batch pharmaceutical manufacturing: a case study. Ind Eng Chem Res. 2011;50:10083–92.
5. Allison G, Cain YT, Cooney C, Garcia T, Bizjak TG, Holte O, et al. Regulatory and quality considerations for continuous manufacturing. J Pharm Sci. 2014;104:803–12.
6. Lee SL, O'Connor TF, Yang X, Cruz CN, Chatterjee S, Madurawe RD, et al. Modernizing pharmaceutical manufacturing: from batch to continuous production. J Pharm Innov. 2015;10:191–9.
7. International Council for Harmonization of Technical Requirements for Human Use (ICH): Press Releases. ICH Assembly, Kobe, Japan, June 2018 (http://www.ich.org/ichnews/press-releases.html).
8. ICH, Pharmaceutical development Q8. 2009.
9. ICH, Development and manufacture of drug substances (chemical entities and biotechnological/biological entities) Q11. 2012.
10. Myerson AS, Krumme M, Nasr M, Thomas H, Braatz RD. Control systems engineering in continuous pharmaceutical manufacturing. J Pharm Sci. 2015;104:832–9.
11. Gouveia FF, Rahbek JP, Mortensen AR, Pedersen MT, Felizardo PM, Bro R, et al. Using PAT to accelerate the transition to continuous API manufacturing. Anal Bioanal Chem. 2017;409:821–32.
12. Chandi A, Daly AM, Foley DA, LaPack MA, Mukherjee S, Orr JD, et al. Industry perspectives on process analytical technology: tools and applications in API development. Org Process Res Dev. 2015;19:63–83.
13. Braden TM, Johnson MD, Kopach ME, McClary Groh J, Spencer RD, Lewis J, et al. Development of a commercial flow Barbier process for a pharmaceutical intermediate. Org Process Res Dev. 2017;21:317–26.
14. Ferreira AP, Tobyn M. Multivariate analysis in the pharmaceutical industry: enabling process understanding and improvement in the PAT and QbD era. Pharm Dev Technol. 2015;20(5):513–27.
15. Kourti T, MacGregor JF. Process analysis, monitoring and diagnosis using multivariate projection methods. Chemom Intell Lab Syst. 1995;28:3–21.
16. Albert S, Kinley RD. Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. Trends Biotechnol. 2001;19:53–62.
17. Miletic I, Quinn S, Dudzic M, Vaculik V, Champagne M. An industrial perspective on implementing on-line applications of multivariate statistics. J Process Control. 2004;14:821–36.

18. Kiran KL, Selvaraj S, Hua JLC. Application of fault monitoring and diagnostics techniques and their challenges in petrochemical industries, IFPAC Proceedings Volumes (IFPAC-Papers Online) 2012; 8: 702–707.
19. Saavedra J, Cordova A. Multivariate process control by transition scheme in soft-drink process using 3-way PLS approach. Procedia Food Sci. 2011;1:1181–7.
20. Kourti T, Lee J, MacGregor JF. Experiences with industrial applications of projection methods for MSPS. Comput Chem Eng. 1996;20:S745–50.
21. Machin M, Liesum L, Peinado A. Implementation of modelling approaches in the QbD framework: examples from the Novartis experience. Eur Pharm Rev. 2011;16:5–8.
22. Zomer S, Zhang J, Talwar S, Chattoraj S, Hewitt C. Multivariate monitoring for the industrialization of continuous wet granulation tableting process. Int J Pharm. 2018;547(1–2):506–19.
23. Esbensen KH, Geladi P. Principal component analysis: concept, geometrical interpretation, mathematical background, algorithms, history, practice. In: Brown SD, Tauler R, Walczak B, editors. Comprehensive chemometrics. Amsterdam: Elsevier; 2009. p. 211–26.
24. Stork C, Kowalski B. Distinguishing between process upsets and sensor malfunction using sensor redundancy. Chemom Intell Lab Syst. 1999;46:117–31.
25. Ferrer A. Multivariate statistical process control based on principal component analysis (MSPC-PCA): some reflections and a case study in an autobody assembly process. Qual Eng. 2007;19(4): 311–25.
26. Eriksson L, Byrne T, Johansson E, Trygg J, Wikström C, editors. Multi- and megavariate data analysis part I basic principles and applications, third revision. Umeå: Umetrics; 2013.
27. Hotelling H. Multivariate quality control – illustrated by the air tasting of sample bomb sights. In: Eisenhart C, Hastay MW, Wallis WA, editors. Techniques of statistical analysis. New York: Mc Graw Hill; 1947. p. 111–48.
28. ICH Quality IWG. Points to consider for ICH Q8/Q9/Q10 implementation. 2011.