ENDBOSS: Industrial endpoint detection using batch-specific control spaces of spectroscopic data

Tim Offermans a, Terry-Hale Wijkera, Rita Folcarellib, Rudi Heemskerk b, Packo P. Lamersb, Maria Proençab, Thanh N. Tran a,b, Lutgarde M.C. Buydensa, Jeroen J. Jansen a,*

a Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ, Nijmegen, the Netherlands
b Corbion, Arkelsedijk 46, 4206 AC, Gorinchem, the Netherlands

ARTICLE INFO

Keywords:
Endpoint Detection Data Industrial ENDBOSS

ABSTRACT

(Bio)chemical industrial batch reactions have to be terminated timely, to prevent waste of resources and decreased production quality due to prolonging the production when the (primary) reaction has already finished. Approaches for detecting endpoints on off-line and/or on-line analysis exist, but may be inaccurate for productions with high batch-to-batch variations. In this study, we present a novel multi-step strategy for endpoint detection named ENDBOSS (ENdpoint Detection using Batch-specific cOntrol Spaces of Spectroscopic data). This strategy is designed to have higher robustness against batch-to-batch variation than endpoint detection methods reported in literature and to be implemented for on-line monitoring. We demonstrate ENDBOSS on three industrially relevant reactions with high batch variations. A method for optimizing the settings of ENDBOSS for a given production process is proposed and demonstrated. The correlations between detected and reference endpoint were validated to be 0.96, 0.80 and 0.18 for the three demonstrator reactions. ENDBOSS has high performance for two reactions. For the third reaction, the size of the dataset was too limited, indicating that ENDBOSS does benefit from quantitative integration with strategic data collection. ENDBOSS is furthermore shown to outperform endpoint detection methods currently reported in literature for all demonstrator reactions.

1. Introduction

Many intermediate and consumer products in the food and pharmaceutical industries are manufactured in batch processes [1]. Such productions have to be carefully monitored and controlled to ensure both production efficiency and quality [2,3]. Much of today’s research focuses on developing methods for automation of process controls, and integrating physical measurements at the industrial plant with several monitoring, control and management tasks to facilitate an industrial cyber-physical system [4]. Although many of the automation methods that are currently used within such a system focus on keeping one or more process variables (pressure, temperature, viscosity or chemical concentration) within a predefined limit, another important feature of batch productions to control accurately is the point at which the production should be terminated. This feature is also referred to as the endpoint, and is highly variable for many production processes [5].

Terminating a batch production too early, before the reaction has completely finished, can lead to a non-optimal yield and an increased demand for purification steps. In addition, the reactor has to be cleaned and (re)prepared more often, which is time and cost intensive. On the other hand, continuing a batch reaction for too long may lead to a waste of energy, material, time and manpower. In some cases, it also leads to the formation of byproducts, which effectively reduces both production yield and quality [6].

Traditionally, reaction endpoints are detected by quantitative wet chemical analysis of samples taken out of the reactor, for instance using chromatography [6,7]. These analyses are relatively cumbersome and slow, and can therefore not be done frequently. Additionally, they do not allow for mitigating control actions based on the observation. Even when the quantitative determination itself is accurate, these practical shortcomings can result in a late and inaccurate endpoint detection.

In-line spectroscopic analysis is commonly used in multivariate statistical process control, as it provides both quantitative and qualitative chemical information, and can be installed to facilitate automatic in-line sampling at high frequency. In quantitative analysis, regression models can be calibrated to predict the chemical composition from the
spectroscopic data. Accurately monitoring this predicted composition in real-time allows for (automated) production endpoint detection. This strategy is for instance used in the pharmaceutical industry for powder blending and crystal polymorph conversion [8,9]. Prediction of physical properties like moisture from spectroscopic data can also be used for endpoint detection, as illustrated for fluidized batch granulation by Frake et al. and Findlay et al. [2,10].

Obtaining an accurate prediction model may however be challenged by reaction mixtures containing multiple compounds with large spectral overlap. This is especially common for bioproductions, due to the nature of working with large molecule structures [11,12]. In such cases it is impossible to fully distinguish the different compounds in the spectral data. An alternative solution is to monitor the spectra without using a predictive model. Principal Component Analysis (PCA) can be used to extract the major sources of variation in the data relevant to the end-point detection. These sources can be monitored over time to track changes in the chemical state of the reaction mixture. A lack of such changes indicates that the reaction has ended and that the production process should be terminated.

This approach is used by Svensson et al. to determine the endpoint for an industrial pharmaceutical synthesis [13]. In this study, a PCA model was calculated on the spectroscopic data from a single golden batch. For new batches, the incoming spectra are projected in that model and the distances between the projection of the subsequent spectra are calculated. Monitoring these distances, and thus the amount of chemical change, allowed for the accurate determination of the new batches’ endpoints. The threshold for the chemical change allowed without ending the reaction was determined arbitrary, and have to be re-considered for different production facilities.

In their work, Svensson et al. used the data from one completed golden batch to calculate a PCA model and define a control space. The use of one or more completed batches as control space for PCA modelling is common in MSPC [14–16]. Although this approach is accurate in those studies, it may be of limited use for processes that suffer from high batch-to-batch variations that still retain in the PCA model (which we will demonstrate in this work for three demonstrator reactions). Such variations can be the result of changes in, for example, raw material, weather, operators and consumer wishes, and are common in the pharmaceutical and food industries [17–19].

In this paper, we introduce ENDBOSS: ENdpoint Detection using Batch-specific cOntrol Spaces of Spectroscopic data. ENDBOSS is a validated multi-step strategy for the endpoint detection of industrial batch reactions. It is designed to have higher robustness against batch-to-batch variation than the aforementioned methods. As it is specifically developed for on-line process monitoring in real-time, it can potentially be integrated in an overarching industrial cyber-physical system [20]. Furthermore, in contrast to the endpoint detection methods mentioned in the literature referred to earlier, an automated routine for finding the optimal settings and thresholds of ENDBOSS for a given production facility is given.

ENDBOSS works by using spectral data measured at the start of each individual production batch as a control space, rather than data from one

Fig. 1. Schematic representation of ENDBOSS.
or more completed batches. Data from after the start of a batch is then compared to the control space of that specific batch to detect the end in chemical changes, and thus the end of production. Further robustness against batch-to-batch variation is obtained using spectral data preprocessing.

2. Methods and data

ENDBOSS is schematically shown in Fig. 1, and consists of two major steps: modelling and monitoring. The modelling step is performed after the first (estimated) 10–20% of the batch have passed, and preferably when the actual reaction is started (in cases where the reaction does not start immediately at the start of the batch). Using pre-processed spectroscopic data from these first few hours, a batch-specific control space is created using Principal Component Analysis (PCA) [21]. The monitoring step starts right after the modelling steps and lasts until the batch is ended. Incoming data is pre-processed and projected into the PCA-model in real-time. This projection is continuously monitored and compared with the control space in terms of the Hotelling T2- or Q-statistic. When the statistic reaches a threshold defined on the control space, the reaction is detected as ended. Each of the (sub)steps of ENDBOSS will be explained in more detail in the following subsections. Source code of the implementation of ENDBOSS in Matlab used to generate the results presented in this manuscript are accessible via https://www.ru.nl/science/analyticalchemistry/research/software/.

2.1. Modelling step

The modelling step is executed after the first (estimated) 10–20% of the batch have passed and the first spectroscopic data is collected. For productions where the actual reaction is initialized a certain period after the measurements have started, for instance by starting the feed of a substrate, the end of that period should be used to execute the modelling step of ENDBOSS. In this step, the batch-specific control space is defined by modelling all spectroscopic data collected up until that point. Prior to modelling, spectroscopic data needs to be pre-processed as they often suffer from scattering and/or fluorescence artefacts. To reduce batch-to-batch variation caused by irrelevant variation such as fluorescence effects rather than actual chemical differences, the spectra are typically pre-processed using a baseline and scatter correction method. Asymmetric Least Squares smoothing (AsLS) is first used to remove any differences in baseline offsets of the individual spectra by combining a smoother with an asymmetric weighting of deviations from the smoothed signal [22, 23]. To correct for scattering artefacts, a Standard Normal Variate transformation (SNV) is applied to the data. This transformation processes each individual spectrum by first subtracting the mean of that specific spectrum and then dividing it by the standard deviation of the spectrum [24].

The preprocessed spectroscopic data from the preparation phase is, for each batch, modelled using Principal Component Analysis (PCA). PCA is a multivariate statistical modelling technique that rotates a data matrix of many variables into a set of orthogonal variables (principal components, PCs) with maximized variance. An algebraic expression of PCA is given in Equation (1). \( X \) denotes the original data matrix with \( M \) rows (representing samples) and \( N \) columns (representing variables); \( T \) is the score matrix of the original samples for the principal components with \( M \) rows and \( A \) columns (representing components); \( P \) is the loading matrix of the original variables for the components with \( N \) rows and \( A \) columns; \( E \) is the model residual matrix with \( M \) rows and \( N \) columns [21,25]. For this study, all spectra are always mean-centered prior to modelling.

\[
X = TP^T + E = \sum_{a=1}^{A} t^a P^a + E
\]  

Equation 1

For each spectrum in the control space, the Hotelling T2- and Q-statistic are calculated. The Hotelling T2-statistic expresses the variation of a spectrum captured and described by a PCA model, and can be calculated from the PCA scores (\( T \) in Equation (1)). The formula for this summary statistic is given in Equation (2) for spectrum \( i \). \( A \) represents the number of components projected to; \( v_i \) represents the variance in the original data matrix explained by component \( a \) [26]. In ENDBOSS, the Hotelling T2-statistic is used to quantify how similar the chemistry measured in one spectrum is to the overall chemistry measured in the spectra used to calculate a PCA model.

\[
T^2_i = \sum_{a=1}^{A} \left( \frac{T^a_i}{v_i} \right)^2
\]  

Equation 2

The Q-statistic expresses the remaining variation of a spectrum that is not captured by a PCA model, and can be calculated from the PCA residual matrix (\( E \) in Equation (1)). The formula for this summary statistic is given in Equation (3) for spectrum \( i \) and spectral variables \( J \) [26]. In ENDBOSS, the Q-statistic is used to quantify how dissimilar the chemistry in one spectrum is to the overall absorption measured in the spectra used to calculate a PCA model.

\[
Q_i = \sum_{j=1}^{J} E_{ij}^2
\]  

Equation 3

2.2. Monitoring step

The monitoring step is initiated directly after the modelling step, and continues until the production is stopped. During this step, incoming spectra are projected in the existing PCA model according to Equation (4) for spectrum \( i \). The data is preprocessed using the same steps as the ones used for the modelling step. Next, either the Hotelling T2-statistic or the Q-statistic is calculated for each of the incoming spectra. This choice depends on whether the actual chemical reaction has already started during the production period used in the modelling step or not, as will be explained in more detail further on. The values for the statistic for the incoming spectra are in real-time calculated and compared to the values for the spectra in the model’s control space to facilitate the endpoint detection.

\[
\hat{T}_i = X_i P
\]  

Equation 4

The Hotelling T2-statistic is monitored for endpoint detection when the chemical production has already started in the production period that is used as a control space. The chemistry of the product is then captured by the PCA model, and changes in its concentration are thus represented in the Hotelling T2-statistic. This scenario is most applicable for productions where the chemical reaction is initiated directly at the start of production. As the product is formed directly at the start of the batch, it will be captured by a PCA model calculated on the first few hours of production data. To monitor the chemical productivity, the Hotelling T2-statistic should thus be used to compare the incoming spectra with the PCA model.

The Q-statistic is monitored for endpoint detection when the chemical productivity has not yet started in the period used to define the batch-specific control space. This is the case when the actual reaction is started only a few hours after the measurements have started, and the measurements from before the reaction start are used as control space. As the product represents chemical variation that is not in the control space and thus described by the PCA model, monitoring its formation in real-time should be done by calculating the Q-statistic for incoming spectra and comparing its value to the values in the control space.

The endpoint of the production is defined by ENDBOSS as the point when the Hotelling T2- or Q-statistic of the incoming spectra no longer increases. This would signify that the chemistry in the reaction vessel is no longer changing, and thus that the reaction has ended. This stop in increase is found by checking whether or not the first derivative of the statistic drops below a certain noise level. In practice, operators who
might use ENDBOSS for their plant can use a visualization of this first derivative over production time as a graph on a screen to keep track on whether the endpoint has been reached or not, and even to estimate how long the production will still take.

The first derivative of the Hotelling $T^2$-or Q-statistic is calculated as the difference between the statistic for the current spectral measurement and the one measurement before that, as exemplified in Equation (5) for the Hotelling $T^2$-statistic for measurement $i$, measurement times $t$, window width $w$ and $N$ samples that were measured within the window limits. Secondly, the smoothened derivative is not compared to exactly zero, but to a noise-threshold that is the window size used for the moving average filter, which replaces each first derivative value by a mean value calculated over a centered window moving over production time [27]. This is shown in Equation (6) for the Hotelling $T^2$-statistic of measurement $i$, measurement times $t$, window width $w$ and $N$ samples that were measured within the window limits. Secondly, the smoothened first derivative is not compared to exactly zero, but to a noise-threshold that is slightly above zero. This noise threshold is calculated as the standard deviation of the smoothened first derivative of the Hotelling $T^2$-or Q-statistic, multiplied with a noise factor. More details on the selection of the window size used for the moving average filter and of the noise factor are given in the following subsection.

Firstly, the derivative is smoothened using a moving average (Boxcar) filter, which replaces each first derivative value by a mean value calculated over a centered window moving over production time [27]. This is shown in Equation (5) for the Hotelling $T^2$-statistic for measurement $i$, measurement times $t$, window width $w$ and $N$ samples that were measured within the window limits. Secondly, the smoothened first derivative is not compared to exactly zero, but to a noise-threshold that is slightly above zero. This noise threshold is calculated as the standard deviation of the smoothened first derivative of the Hotelling $T^2$-or Q-statistic, multiplied with a noise factor. More details on the selection of the window size used for the moving average filter and of the noise factor are given in the following subsection.

$$
T^2_{i,t} = \frac{T^2_{i,t} - T^2_{i,t-1}}{t_i - t_{i-1}} \quad \text{Equation 5}
$$

$$
T^2_{\text{smooth},i} = \frac{1}{N} \sum_{j=0}^{j=N} T^2_{j,i} \quad \text{Equation 6}
$$

### 2.3. Parameter optimization

An overview of the different steps in ENDBOSS for which different methods or settings are considered, is given in Table 1. Both baseline and scatter correction are optional preprocessing steps, as they might not be necessary for each production process. The duration of the control space is a parameter only when the Hotelling $T^2$-statistic is monitored for endpoint detection, and data recorded after the start of the reaction is used as a control space. When the Q-statistic is monitored, simply all data collected before the start of the reaction is used as control space. Finally, the number of principal components to include in the model, the width of the moving average filter used to smoothen the PCA-statistic and the factor used to calculate the noise threshold for the PCA-statistic can all be varied, and their settings can highly influence the accuracy of endpoint detection. Note that the settings for the smoothing width and the control space duration that are considered are defined not absolute but relative to the average batch length for a certain production process. The values given in Table 1 are factors by which this average batch length is multiplied.

The optimal settings for the steps of ENDBOSS are different per production process, depending on the chemical nature of the process and the spectroscopic method used for monitoring. Although background knowledge of the process and the measurement can and should be employed to set (part of) these parameters, such knowledge might not always be available. The settings for these parameters as given in Table 1 were therefore optimized per demonstrator process using a full-factorial experimental design. This approach would correspond to the worst-case scenario that no background knowledge is available and ENDBOSS has to be optimized data-driven. The parameter settings leading to the highest performance of ENDBOSS are on the presence or absence of a (potentially outlying) batch.

To further estimate the significance of the validated performance, a permutation test was applied [29]. The endpoints and spectral data of the production batches were re-assigned to each other at random, after which the entire validation routine was performed to obtain an $r_{\text{detected, reference}}$ for the permuted data. This permutation test was repeated 100 times and a 95%-confidence interval was calculated for the detection performance. Furthermore, the main and interaction effects of each parameter on the detected performance were studied by analyzing the correlation coefficients per setting and using three-way ANOVA [30].

### 2.4. Demonstrator process I: simulated penicillin production

The first demonstrator process used to test ENDBOSS is an advanced simulation of an industrial penicillin production named IndPenSim. This simulator is developed by Goldrick et al. for the development of data analysis tools applicable to the biopharmaceutical industry, and features simulation of on-line Raman spectroscopic data [31]. A set of 100 pre-simulated batches was retrieved from Mendeley [32], of which 42 were used to test ENDBOSS. These 42 batches showed a clear maximum in penicillin concentration, which is used as reference endpoint, and last for 204–278 h. A complete overview of all 42 batches is given in Table S2 in the supplemental material. The penicillin and substrate concentration of the first of the 42 batches is shown in Fig. 2a as an example. The Raman data for this batch is shown in Fig. 2b. As the reactions in all of these batches were initialized immediately by adding substrate, only the use of the Hotelling $T^2$-statistic for endpoint detection is considered.

### 2.5. Demonstrator process II: real-world industrial biochemical production

The second demonstrator process on which ENDBOSS was tested is an industrial biochemical production facility. Data was collected for 19 production batches, of which a complete overview is given in Table S1 in the supplemental material. The time scales of all batches have been normalized to the duration of the longest batch (batch 15). The batches represent 11 different recipes, and all featured a startup phase during which no substrate is added to the reactor. The production batches are monitored using off-line High-Performance Liquid Chromatography (HPLC) and on-line Raman spectroscopy [33,34]. Examples of both data sources are given in Fig. 3a–b for batch 13. Raman signals below 250 cm$^{-1}$ and above 2750 cm$^{-1}$ were always discarded as these regions showed no signal for any of the production batches.

| Level | Baseline correction | Scatter correction | Control space duration (T2 only) | PCs | Smoothing width | Noise factor |
|-------|----------------------|---------------------|----------------------------------|-----|----------------|-------------|
| 1     | No                   | No                  | 0.05                             | 1   | 0.05           | 0.05        |
| 2     | AnaLS                | SNV                 | 0.10                             | 2   | 0.10           | 0.10        |
| 3     | AnaLS                | SNV                 | 0.20                             | 3   | 0.20           | 0.20        |
| 4     | AnaLS                | SNV                 | 0.50                             | 4   | 0.50           | 0.50        |
| 5     | AnaLS                | SNV                 | 1.00                             | 5   | 1.00           | 1.00        |
All production batches featured a primary reaction, but 8 batches also featured a secondary reaction that starts after the endpoint of the primary reaction is reached. The batch for which the data is exemplified in Fig. 3a–b also features this secondary reaction. For all batches, undesired byproducts start to accumulate at the end of the process. A strategy for detecting the endpoints of both reactions is therefore desired. Although these reactions are comparable and the same spectroscopic instrument is used to monitor them, the optimal ENDBOSS settings cannot be assumed to be the same for both reaction types. These parameter settings were therefore optimized for the primary reactions alone, the secondary reactions alone, and for both reaction experiments combined.

Furthermore, both the uses of the Hotelling T²-statistic and the Q-statistic for endpoint detection were tested for this demonstrator process. When using the Hotelling T²-statistic, the data collected in the first few hours after a reaction was started (primary or secondary) was used as control space. When using the Q-statistic, the preparation phase was used as control space for the detection of the primary endpoint and the primary reaction stage was used as control space for the detection of the secondary endpoint.

### 2.6. Comparing ENDBOSS to an alternative method

To place the performance of ENDBOSS for the demonstrator cases better in the context, the endpoint detection method proposed in the work of Svensson et al. was also applied to the demonstrator cases [13]. From the endpoint detection methods reported in literature, this method is fundamentally most similar to ENDBOSS.

For this method, a PCA model is trained on one golden batch of a certain reaction. The data for newly measured batches are projected in this model, after which the Euclidian distances between the PCA scores of subsequent spectra are calculated. These Euclidian distances are smoothed over time using a moving average. The derivative of these smoothed Euclidian distances is calculated, which is smoothed a second time with the same method as before. The point in time where this signal drops below a certain threshold is marked as the endpoint for the new batch.

For their case study, Svensson et al. preprocessed the spectra only using SNV, modelled the golden batch using two principal components, smoothed the Euclidian distances using a window width of five samples and used an endpoint detection threshold of 0.005. These settings for these parameters were set arbitrary, but will be different for the demonstrator processes reported in this manuscript. The method settings were therefore optimized for each demonstrator process using the same approach as are used for ENDBOSS (a full-factorial design validated with double leave-one-out cross-validation).

To ensure that the comparison with ENDBOSS is as close as possible, the same parameter settings were considered. Both AsLS and SNV were considered for baseline and scatter correction and a maximum of three principal components were considered for modelling. The window widths considered for the two moving average-smoothings are 0.05, 0.1 and 0.2 times the average batch length of the respective demonstrator process. Finally, as detection thresholds the standard deviation of the smoothened first derivative of the Euclidian distances of the PCA model of the golden batch times the noise factors given in Table 1 are considered.
For demonstrator data simulated using IndPenSim, no golden batch is indicated. Batch 3 (in Table S1) was therefore used as golden batch, as it has the median total penicillin yield of all batches at 2.675.300 kg. For demonstrator process II, batch 15 (in Table S2) was used as golden batch for both reaction types, as it was marked as such by plant experts.

3. Results and discussion

3.1. Demonstrator process I: simulated penicillin production

For demonstrator process I, the parameters for ENDBOSS were only optimized for the scenario in which the Hotelling \( T^2 \)-statistic is used during the monitoring step, as there is no preparation stage available for this production. The parameter settings found optimal using double cross-validation are to perform no baseline or scatter correction, use a period of 0.2 times the average production time as control space, model using only one PC, use a smoothing window width of 0.2 times the average production time and use a noise factor of 0.05. Fig. 4 shows the trajectory of the differentiated Hotelling \( T^2 \)-statistic and the noise-threshold for the batch of which the raw data is shown in Fig. 2a–b when using these optimal settings. Such a visualization could be used by process operators to interpret the results of ENDBOSS in real-time and to keep track on how far the endpoint of the batch is due. Remarkable is that the signal is still quite noisy despite the smoothing step, but that the detection is still accurate.

The detection versus reference endpoints found after double cross-validation of the optimization are plotted in Fig. 5. The validated correlation between detected and reference endpoint is 0.964, which illustrates that ENDBOSS has high accuracy for this production process. For the permutation tests, the mean and 95% confidence interval of the \( r_{\text{detec}} \) was found to be 0.054 ± 0.050. This illustrates that the performance of ENDBOSS found using double cross-validation is statistically significantly higher than random. ENDBOSS does however perform particularly poor for one batch (batch 91 in the overview in Table S2), which is also the shortest batch in the dataset.

The main effects of the parameters on the detection accuracy are visualized in Fig. 6. This figure shows that the duration of the control space is the parameter with the largest effect on the detection performance. For each parameter, the setting that according to the results shown in Fig. 6 lead to the highest correlation between detected and reference endpoint match the ones found optimal using double cross-validation. The results furthermore suggest that increasing the smoothing parameter beyond 0.2 would increase the detection accuracy even more. However, using such wide windows is undesirable as the detection will be late by half of the window width by design.

The main results of the ANOVA for the first demonstrator process are shown in Fig. 7. This figure shows the standardized main and interaction effect sizes, calculated as the point-biserial correlation \( r_{\text{pb}} \). The corresponding full ANOVA-table is given in Table S3 in the supplemental material. For all parameters, except for the noise factor, a \( p \)-value below 0.05 was obtained, indicating that these parameters have a significant effect on the detection accuracy. The results furthermore show significant two- and three-level interactions of quite some parameters, for instance between the baseline and scatter correction.

The significant interaction between baseline and scatter correction can be explained by the fact that they both correct for fluorescence artefacts and are therefore related operations. Both corrections also interact with the smoothing parameter, as for this process they correct for the same type of variation. Baseline and scatter correction respectively remove an additive and multiplicative factor from the different spectra, which for this process gradually changes over production time. As smoothing attempts to reduce variation over time, the choice of performing a baseline and/or scatter correction or not affects the optimal window width setting for smoothing. Interactions between the control space duration and the baseline correction, scatter correction and smoothing are caused by the fact that variances due to fluorescence can be better described by PCA itself when more data is included in the model, making these operations less necessary. This also explains the interaction between scatter correction, baseline correction, smoothing and the number of PCs chosen.

3.2. Demonstrator process II: real-world industrial biochemical production

An overview of the validation results for ENDBOSS applied to the second demonstrator process is given in Table 2. Results are shown for both monitoring scenarios (Q- and Hotelling \( T^2 \)-statistic), and for optimizing the parameter settings on only the primary reactions, only the secondary reactions or all reactions together.

For this process, using the Q-statistic during the monitoring step gives in most cases a higher endpoint detection accuracy for the test set than using the Hotelling \( T^2 \)-statistic does. A cause for this is that there is one parameter less to optimize when the Q-statistic is used (the control space duration). This causes less opportunity for overfitting and thus a higher performance on an independent test set. A secondary reason is that when the Hotelling \( T^2 \)-statistic is used, not only data from the preparation stage but also data from the reaction stage is used to define the control space and the noise threshold for the statistic. As the batch-to-batch variation of the reaction stages is higher than that of the preparation stages, it is more difficult to optimize the parameter settings batch-invariantly. This lowers the performance on the independent test set.

However, using the Hotelling \( T^2 \)-statistic does give higher performance when all reactions are considered at once. A reason for this is that the
The optimal parameter settings are more similar for all reactions when the Hotelling T²-statistic is used then when the Q-statistic is used. This is confirmed by the fact that the parameter optimization shows less overfit (a lower difference between calibration and testing performance) for the Hotelling T²-statistic.

The performance of ENDBOSS is higher for the primary reactions than for the secondary reactions. One reason for this is that the sample size for the secondary endpoints is lower than for the primary endpoints (8 versus 19, respectively), which makes ENDBOSS optimization more prone to overfitting and thus having a lower testing performance for the secondary reactions. Another reason is that data from the primary reaction stage is used as control space for detecting each secondary endpoint, rather than data from the preparation stage. As mentioned earlier, the batch-variation of the reaction stage is likely higher than that of the preparation stage, lowering the detection accuracy for the secondary reaction.

The highest endpoint detection accuracies for the data from this demonstrator process are obtained when the Q-statistic is used during the monitoring step. Plots showing the detected versus reference endpoints using this scenario are given in Fig. 8a–b for the two reactions separately. These results are found using double cross-validation, and correspond to the results shown in Table 2 in the top two rows and rightmost column.

The optimal parameters settings for the primary endpoints are to use both AsLS and SNV, model two PCs, smooth the Q-statistic with a window of 0.2 times the average production duration and to use a noise threshold of 0.05 times the standard deviation of the Q-statistic in the control space. The optimal settings for the secondary endpoints are to not use AsLS or SNV, use only one PC and use a smoothing and noise factor of 0.05 and 0.25, respectively. The monitoring results from ENDBOSS using these settings for the primary reaction of the batch of which the raw data is shown in section 2.5 are visualized in Fig. 9a, analogous to Fig. 4 for the first demonstrator process. A similar visualization for the secondary reaction is given in Fig. 9b, which clearly shows that the spectroscopic data suffers from variation near the end of the batch that cannot be modelled by ENDBOSS, resulting in a low detection performance.

Remarkable is that removing fluorescence artefacts using AsLS and SNV is only optimal for the primary endpoints. A likely reason for this is that PCA itself can model the artefacts for this dataset particularly well. However, these artefacts are not yet present in the control spaces used for primary endpoint detection. They are therefore not modelled by PCA, and have to be removed explicitly during the monitoring step using data pre-processing. For the control spaces used for secondary endpoint detection, the artefacts are present in the control spaces and are therefore modelled by PCA. During the monitoring step, explicitly removing them using data pre-processing is not required.

The detection for one secondary endpoint could not be validated: only 7 out of 8 endpoints are shown in Fig. 8b. The batch for which no endpoint could be detected has only very little production data available after its endpoint. The Q-statistic for this batch did not reach the endpoint threshold within this period, resulting in the batch being discarded. In other words: ENDBOSS cannot detect the endpoint this short after the
true endpoint, even though this would still be an accurate prediction considering the detection accuracy for the other batches and the uncertainty of the reference endpoints. This does however indicate that for optimizing the ENDBOSS parameters, it is important to have historical data for batches with a considerable measurement period after the endpoint.

The main effects of the parameters on the accuracy of primary and secondary endpoint detection using the Q-statistic are visualized in Fig. 10a–b, respectively. The parameters corresponding to the data pre-processing (baseline and scatter correction) have the highest effect on the detection accuracy, except for SNV for the primary endpoints. For both endpoints, increasing the noise factor decreases the detection accuracy, while increasing the number of PCs increases the accuracy. The smoothing parameter shows an optimum at 0.1 for the primary endpoints. For the secondary endpoints, the results suggest that increasing this parameter beyond 0.2 would increase the detection accuracy even more. However, using very wide windows is undesirable as the detection will be late by half of the window width by design.

Fig. 8. a-b: Detection versus reference plots for double cross-validated primary (a) and secondary (b) endpoint detection of the first demonstrator process using the Q-statistic.

Fig. 9. a-b: Trajectories of the differentiated Q-statistic found using ENDBOSS for the primary and secondary reactions of the second demonstrator process, for the batch of which the raw data is shown in section 2.5.

Table 2
Overview of validated endpoint detection accuracies using either the Q-statistic or Hotelling $T^2$-statistic during the monitoring step, for primary reactions alone, secondary reactions alone and both reactions together. For the permutation testing results, the mean and 95% confidence limits over all 100 permutations are given.

| Monitoring scenario | Industrial reaction | Permutation r(det, ref) | Calibration r(det, ref) | Validation r(det, ref) | Testing r(det, ref) |
|---------------------|---------------------|-------------------------|-------------------------|------------------------|--------------------|
| Q-statistic         | Primary             | -0.023 ± 0.060          | 0.842                   | 0.797                  | 0.796              |
|                     | Secondary           | -0.007 ± 0.092          | 0.935                   | 0.284                  | 0.175              |
|                     | All                 | 0.066 ± 0.056           | 0.787                   | 0.377                  | 0.371              |
| Hotelling $T^2$-statistic | Primary             | 0.008 ± 0.061          | 0.683                   | -0.033                 | -0.106             |
|                     | Secondary           | 0.028 ± 0.084           | 0.984                   | 0.225                  | 0.141              |
|                     | All                 | 0.008 ± 0.053           | 0.721                   | 0.414                  | 0.394              |

Fig. 10a–b: Detection versus reference plots for double cross-validated primary (a) and secondary (b) endpoint detection of the first demonstrator process using the Q-statistic.

The main effects of the parameters on the accuracy of primary and secondary endpoint detection using the Q-statistic are visualized in Fig. 10a–b, respectively. The parameters corresponding to the data pre-processing (baseline and scatter correction) have the highest effect on the detection accuracy, except for SNV for the primary endpoints. For both endpoints, increasing the noise factor decreases the detection accuracy, while increasing the number of PCs increases the accuracy. The smoothing parameter shows an optimum at 0.1 for the primary endpoints. For the secondary endpoints, the results suggest that increasing this parameter beyond 0.2 would increase the detection accuracy even more. However, using very wide windows is undesirable as the detection will be late by half of the window width by design.

Most remarkable is that the main effects visualized in Fig. 10a–b suggest different optimal settings than the ones mentioned before found using double cross-validation. For primary endpoint detection, using two PCs is optimal according to cross-validation, while Fig. 10a shows that the average detection accuracy is highest when three PCs are used. Similarly, cross-validation shows that using AsLS for secondary endpoint detection is not optimal, while Fig. 10b suggests that it would be. These apparent disagreements result from the interaction effects that the parameters have on the endpoint detection, and will be further discussed later on as part of the ANOVA-results.

The results of the ANOVA performed to quantify the main and interaction effects of the parameters on the endpoint detection accuracy are shown in Fig. 11. This figure, as before, only shows the point-biserial correlations $r_{pb}$ [35]; the full ANOVA results are given in Table S4 in the
supplemental material. Both Fig. 11 and Table S4 show the results for the primary and the secondary endpoint detection using the Q-statistic, and correspond to the results shown in Fig. 8a–b. All of the ENDBOSS parameters have a significant main effect on the accuracy of both primary and secondary endpoint detection, as their p-values are below 0.05, and are thus important to optimize.

The ANOVA results also indicate significant two- and three-level interactions of quite some parameters for this demonstrator process, which explains the disagreement between the results in Fig. 10a–b and the optimal parameter settings found using double cross-validation. For instance, the number of PCs interacts with all other parameters for primary endpoint detection. Although using three PCs might increase the detection accuracy on average (Fig. 10a), due to these interactions there can be a parameter design that uses only two PCs that has the highest detection accuracy. Performing both AsLS and SNV might for example increase detection accuracy, but only when two components are used instead of three. Likewise, for the secondary endpoint detection, scatter correction interacts significantly with all other parameters. Performing SNV might give better detection performance on average, but the correct combination of all other parameters without performing SNV might still give the highest performance. Explanations for interactions between scatter correction, baseline correction, smoothing and the number of PCs chosen follow the ones given for demonstrator process I.

### 3.3. Comparing ENDBOSS to an alternative method

The results of applying the endpoint detection method reported by Svensson et al. [13] with optimization of the method’s parameters using the same approach as used for ENDBOSS are given in Table 3 for all demonstrator reactions. Only the results of the permutation test and the double cross-validation of the endpoint detection are discussed for conciseness. Note that for each demonstrator process, no endpoint is detected for the golden batch as it is used as modelling control space.

The results in Table 3 show that the alternative endpoint detection method has virtually no performance for the secondary reaction of demonstrator process II. The performance of ENDBOSS for this reaction, given in Fig. 8b, is higher, but this performance is also too low to apply it to this reaction. The alternative endpoint detection has also low performance for demonstrator process I, but has reasonable performance for the primary reactions of demonstrator process II. For both demonstrator cases, the detection accuracy as obtained by ENDBOSS, given in Figs. 5 and 8a, is higher and significant. As the performances presented for both methods are found after leave-one-out cross-validation, this difference in performance would indicate that ENDBOSS is indeed more robust against batch-to-batch variations than the alternative method.

Fig. 10. a-b: Effects of the ENDBOSS parameters on detection accuracy for primary (a) and secondary (b) endpoints for demonstrator process II. Whiskers refer to the 95% confidence intervals.
4. Conclusion

In this study, we presented ENDBOSS, a novel strategy for endpoint detection of industrial batch reactions, and demonstrated its potency on two demonstrator production processes of which one may feature both a primary and secondary reaction per production. ENDBOSS is based on monitoring changes in the major sources of variation in spectroscopic data measured in real-time, using PCA. ENDBOSS uses batch-specific control spaces to increase robustness against batch-to-batch variation, but the optimal settings for certain steps have to be optimized per production process. Expert-knowledge can and should be used for this optimization if available, but for the worst-case scenario that none of such knowledge is available we have proposed and demonstrated a generic data-driven optimization approach using a full-factorial experimental design. Double leave-one-out cross-validation illustrated that ENDBOSS and the proposed optimization approach has very high performance for the first demonstrator process, high performance for the primary endpoints of the second demonstrator process and low performance for the secondary endpoints of the same process. The low performance for the secondary reactions can be largely attributed to the low number of productions available to optimize ENDBOSS on. Using ANOVA applied to the models with significant performance we showed that for each of the steps that are optimized, changing the settings has a significant effect on the endpoint detection accuracy. This stresses the importance of a good optimization approach for ENDBOSS, as is supplied in this work. Finally, it has been showed that ENDBOSS has higher endpoint detection accuracy for the demonstrator processes than the most similar alternative method reported in literature. We expect ENDBOSS and the associated optimization routine to be a viable endpoint detection method for other production processes, and recommend it for consideration.

CRediT authorship contribution statement

Tim Offermans: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. Terry-Hale Wijker: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft. Rita Folcarelli: Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration. Rudi Heemskerk: Conceptualization, Methodology, Resources, Data curation, Writing - review & editing, Supervision, Project administration. Packo P. Lamers: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft. Maria Proença: Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration. Thanh N. Tran: Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration. Lutgarde M.C. Buydens: Conceptualization, Methodology, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Acknowledgements

This project is co-funded by TKI-EI with the supplemental grant 'TKI-Toeslag' for Topconsortia for Knowledge and Innovation (TKI’s) of the Ministry of Economic Affairs and Climate Policy. The authors thank all partners within the project 'Integrating Sensor Based Process Monitoring and Advanced Process Control (INSPMC)', managed by the Institute for Sustainable Process Technology (ISPT) in Amersfoort, The Netherlands.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2020.104229.

References

[1] E.N.M.v. Sprang, et al., Batch process monitoring using on-line MIR spectroscopy, The Analyst 128 (2003) 98–102.
[2] P. Frake, et al., Process control and end-point determination of a fluid bed granulation by application of near infra-red spectroscopy, Int. J. Pharm. 151 (1997) 75–80.
[3] S. Craven, J. Whelan, B. Glennon, Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller, J. Process Contr. 24 (2014) 345–357.
[4] A. Hagsten, et al., Identifying sources of batch to batch variation in processability, Powder Technol. 179 (2018) 292–299.
[5] T.A. Catelani, et al., Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: a feasibility study, Talanta 179 (2018) 292–299.
[6] B. Boeselhao, X.Z. Wang, Multivariate statistical process control for on-line monitoring size of milling process, TNI J. Eng. Technol. 1 (2) (2013).
[7] K.A. Esmonde-White, et al., Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing, Anal. Bioanal. Chem. 409 (2017) 657–649.
[8] O. Svensson, M. Josefsson, F.W. Langkilde, The synthesis of metoprolol monitored using Raman spectroscopy and chemometrics, Eur. J. Pharm. Sci. 11 (2000) 141–155.
[9] S. Yin, J.J. Rodriguez-Andina, Y. Jiang, Real-time monitoring and control of industrial cyberphysical systems, IEEE Industrial Electronics Magazine 13 (4) (2019) 38–47.
[10] J.P. Gorden, J.A. Westerhuis, A.K. Smilde, Monitoring of batch processes using spectroscopy, Am. Inst. Chem. Eng. 48 (10) (2002) 2283–2297.
[11] L.W. Ward, et al., On-line determination of reaction completion in a closed-loop using NIR spectroscopy, Appl. Spectrosc. 52 (1) (1998) 17–21.
[12] L.R. Snyder, J.J. Kirkland, J.W. Dolan, Introduction to Modern Liquid Chromatography, 3 ed, John Wiley & Sons, Inc, 2009.
[13] S. Smit, et al., Assessing the statistical validity of proteomics based biomarkers, Anal. Chim. Acta 592 (2007) 210–2107.
[14] P.I. Good, Resampling Methods: A Practical Guide to Data Analysis, third ed., Birkhäuser, Boston, 2006.
[15] R.M. Mickey, O.J. Dunn, V.A. Clark, Applied Statistics: Analysis of Variance and Regression, Wiley, New York, 2004.
[16] S. Goldrick, et al., The development of an industrial-scale fed-batch fermentation simulation, J. Biotechnol. 193 (2015) 70–82.
[17] S. Goldrick, et al., Moden day monitoring and control challenges outlined on an industrial-scale benchmark fermentation processes, Comput. Chem. Eng. (2019) 130.
[18] L.R. Snyder, J.J. Kirkland, J.W. Dolan, Introduction to Modern Liquid Chromatography, 3 ed, John Wiley & Sons, Inc, 2009.
[19] P. Vandenabeele, Practical Raman Spectroscopy: an Introduction, John Wiley & Sons Inc, 2013.
[20] V.B. Kampenes, et al., A systematic review of effect size in software engineering experiments, Inf. Software Technol. 49 (2007) 1073–1086.

11