

## Appendix C. Supplementary Materials

### *Appendix C.1. Case study 2*

This second case study is added as supplementary material upon suggestion by reviewers. It explains a case study on a simulation test bed which has been developed with the aim of fault detection and identification (FDI) benchmarking in mind. The results presented in this case study provide further support to the conclusions made in the main article and also show how methods for qualitative analysis of time series can assist in the automated identification of faulty or abnormal process conditions. As such, we refer to the main article for detailed explanations of the provided methods. The next sections explain (1) the additional materials and methods for this case study, (2) qualitative analysis results and (3) fault detection and identification results.

### *Appendix C.2. Materials and methods*

#### *Appendix C.2.1. Benchmark data: Penicillin batch fermentation*

Noise-free simulated data used here for benchmarking is taken from the study of [17]. In this previous study, the Penicillin batch fermentation model from [3] is used for comparison of fault detection and identification methods based on Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs). Hundred batches were simulated for normal operation conditions (NOC) and 50 for each of three different faulty conditions, leading to 250 noise-free simulations in total. The three faulty conditions correspond to a reduction of the agitation power (Fault 1), an increase of the saturation constant from 0.15 g/l (nominal value) to values ranging between 0.3 and 0.9 g/l (Fault 2) and a reduction of the substrate feed rate in the fed-batch stage to values ranging from 0.001 to 0.01 l/h (Fault 3). A visual inspection of the simulated penicillin profiles in [17, Figure 3] shows that the Penicillin profiles for NOC, Fault 2 and Fault 3 can be distinguished easily on the basis of their qualitative description. In particular, normal operation corresponds to a DA profile, Fault 2 to a DADA profile and Fault 3 to a DABC. Fault 1 batches are hardly distinguishable from the NOC

batches. As such, it appears interesting to use qualitative analysis for fault detection and identification (FDI) of Fault 2 and Fault 3. Such an approach would avoid the use of a complicated mathematical model for data analysis and also provides a strong link with visual cues to the behavior of the process.

To investigate the potential of the QRT method for this purpose, noise-free data are used to determine the exact, true location of inflection points and maximal values. The first 50 NOC batches are selected for analysis as well as all Fault 2 and Fault 3 batches (150 batches in total). Each simulated batch lasts 400h and the Penicillin concentration is sampled at every time interval of 4.8 minutes (288 seconds), leading to 5000 samples for each simulated batch. Figure C.16 shows the resulting noise-free Penicillin profiles. For normal operation and Fault 3, simple differentiation of these profiles indicates the location of the extrema and inflection points. In the case of Fault 2, differentiation results in multiple inflection points which are considered spurious (i.e. due to numerical simulation error). In this case, the most important ones are selected by selecting the two inflection points with maximal slope and one inflection point in between those with a minimal slope. The selected extrema and inflection points determine the target qualitative representation. As discussed above, the target representations are DA (NOC), DADA (Fault 2) and DABC (Fault 3).

For analysis, each noise-free simulation was subjected to Gaussian noise at different noise levels. To this end, 10 different values for the standard deviation were used, ranging from 0.005 g/l to 0.1581 g/l, in equal steps of 0.5 on a logarithm 10 scale (i.e. each time multiplying by approximately 3.16). Averaged over the whole data set, this delivers values for the Signal to Noise Ratio (SNR, as ratio of standard deviations) ranging from  $1.25 \cdot 10^5$  down to 4. The ratio of the noise standard deviation to the signal range increases from  $3 \cdot 10^{-4}\%$  to 10%. As such, the simulated range of noise levels is from extremely low to high. As in the main article, hundred repetitions are generated at each level for the standard deviation. This leads to a total of 150.000 measurement profiles (150x10x100). Each of these were analyzed by the original method as well as the three extended methods proposed in the main article.

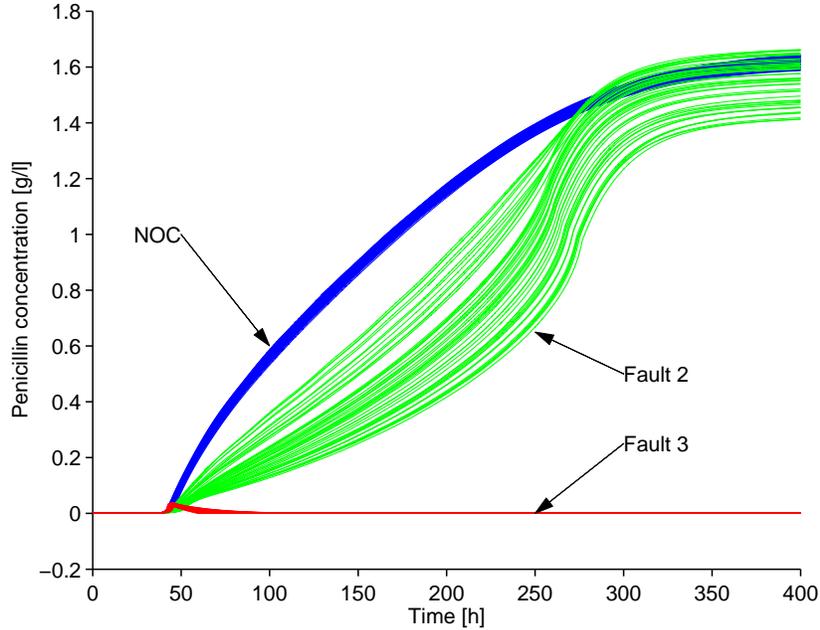


Figure C.16: Noise-free Penicillin concentration profiles throughout the 150 selected batches.

### Appendix C.2.2. Performance evaluation

In this case study, we investigate both the accuracy of the resulting qualitative representations as well as the accuracy of a fault detection and identification (FDI) system which could be based on it. The selected performance metrics for accuracy of qualitative representations are the same as in the main study (fraction of correct representations, fraction of representations with unwanted features and fraction of representation with missing features). For Fault Detection and Identification performance assessment, we set up a classifier which identifies a batch as normal if the resulting representation is DA and faulty if another representation results. The performance of the fault detection mechanism is computed as the fraction of batches which are correctly classified as normal or abnormal. This fraction is named fault detection accuracy. A similar strategy is used for fault identification. Here, the fault diagnosis system iden-

tifies normal operation if the obtained representation is DA; it identifies fault 2 if the obtained representation is DADA and identifies fault 2 is the obtain representation is DABC. If another representation is obtained, the batch is not assigned to any condition (class). The fault identification accuracy is then equal to the fraction of batches which are assigned to the correct condition (class).

*Appendix C.3. Qualitative analysis results*

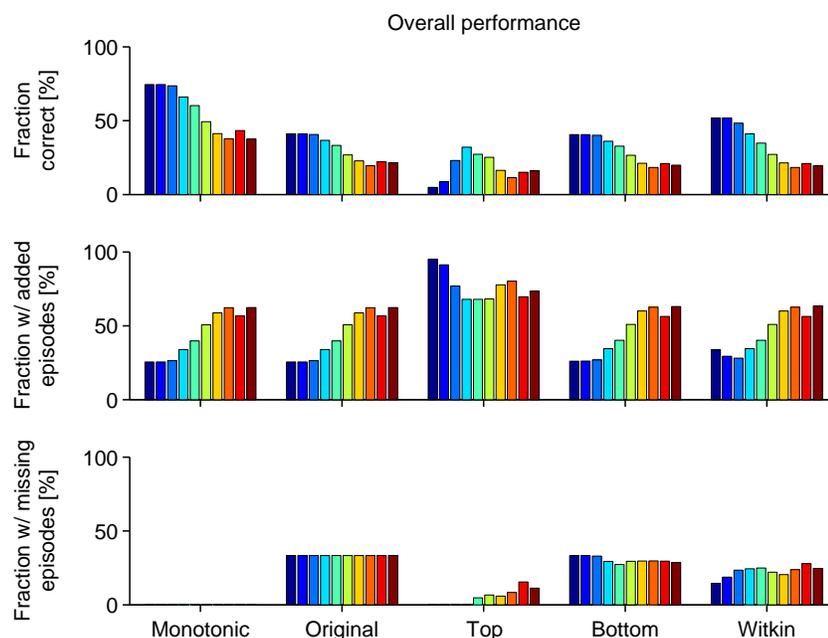


Figure C.17: Accuracy of identified qualitative sequences as function of method and noise level (increasing noise level from left to right in each group). Fraction of correct qualitative sequences (top), fraction of representations with unwanted episodes (middle) and fraction of representations with missing episodes (bottom).

Figure C.17 shows the overall performance of the original method and the proposed alternative methods for the fermentation benchmark data set. In the top panel, one can see that the best performance of about 65% is obtained for monotonic representations at very low noise levels ( $\sigma = 0.005g/l$ ). This

fraction drops to about 40% at the 8<sup>th</sup> noise level ( $\sigma = 0.15g/l$ , noise to signal ratio of standard deviations: 2.5%). If a triangular representation is desired (Original method and alternatives Top, Bottom and Witkin), then the a similar decreasing trend is seen for each of the applied methods, except for the Top alternative. The fractions are lower however, ranging from 50% for the Witkin method at the lowest noise level to 5% for the Top method, also at the lowest noise level. The Witkin method performs best at all noise levels. The original method and the Bottom alternative deliver very similar but lower performances. The Top alternative performs worst of all at all noise levels. The middle and bottom panel indicate that the misidentified qualitative representations are to a large extent due to the identification of representations which are too complex in the sense that these representations exhibit features which are not present in the target representation. This effect is fairly similar across all applied methods including Monotonic representations yet excluding the Top alternative. This method leads to a higher fraction of representations with unwanted features at each noise level than all other methods. Differences among the other methods are due to fractions with missing episodes. In the Monotonic representations, none have missing episodes as opposed to the Original method, the Bottom method and the Witkin method. This explains the higher fraction of correctly identified monotonic representations. For the Original method, the fraction of representations with missing episodes amounts to about 33%. This fraction (one third) is almost entirely due to the batches with DADA representation (Fault 2). These representations cannot be identified by default with the original method, as explained in the main article. Both the Bottom and Witkin alternative do better at identification of this set of representations. The Witkin alternative leads to the lowest fraction of representations with missing episodes. This explains why the Witkin method fares better in correctly identifying the triangular representations.

Figure C.18 shows the average time shifts of identified extrema and inflection points for each of the applied methods. These averages are computed over the set of representations which are correctly identified. For Monotonic repre-

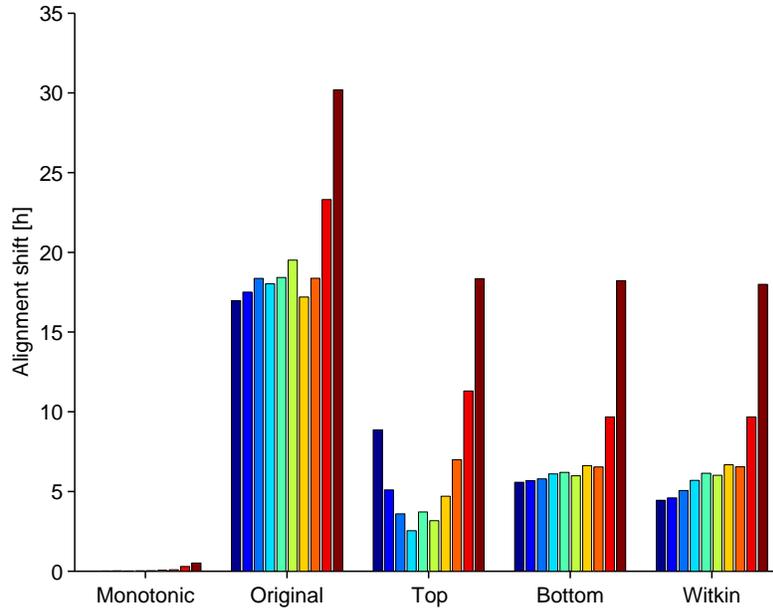


Figure C.18: Misalignment as function of method and noise level (increasing noise level from left to right in each group). Crosses (×) indicate where misalignment is not computed.

sentations, the shift is only affected by the maximum in the batches of fault 3 (with triangular representations DABC and monotonic representation UL). In this case, this average shift is very small (less than 0.6 hours or 36 minutes), indicating that the maximum can be located quite precisely provided its presence is identified correctly. Identifying the location of inflection points appears more difficult. Indeed, for the triangular representations obtained by the original method and the proposed alternatives, the average shift ranges from 2.5 to about 30 hours. Except for the Top method, the shift generally increases from the lowest noise level to the highest. Importantly, the average shift for the Witkin method is always significantly lower than for the original method. The ratio of shifts ranges from about 15% at the lowest noise level to 61% at the highest level. As such, the Witkin method is not only preferred on the basis

of the identified representations but also on the basis of the shift in identified locations of extrema and inflection points with respect to their true locations.

#### *Appendix C.4. Fault detection and identification results*

Figure C.19 shows the accuracy of the methods as applied for Fault Detection and Identification. We show only results on the basis of triangular representations since one cannot distinguish between normal operation (DA) and Fault 2 (DADA) on the basis of monotonic representations (both U). The top panel in Figure C.19 shows the fault detection accuracy for each of the methods and as function of the noise level. This fault detection accuracy ranges from 65% to 85%. Interestingly, fault detection generally increases for all methods at higher noise levels. This effect is due to the fact that a representation which does not match any of the true ones (DA, DADA or DABC) is classified as abnormal. This occurs most frequently for Fault 3, leading to increasing positive detection as a result of a wrongly identified qualitative representations. The Witkin method delivers the best detection rates at very low noise levels. The Bottom method delivers best detection rates at low to high noise levels (0.0005 g/l and higher). This is also explained by increasing positive detection rates due to wrongly identified qualitative representations.

The situation is different when comparing the fault identification results. In this case, the Witkin method performs best whereas the original method and Bottom method deliver a similar performance. The Top method delivers the worst performance. Note however that the Witkin method delivers a rather low performance of 52% at the lowest noise level which is reduced further to about 20% at the highest noise levels.

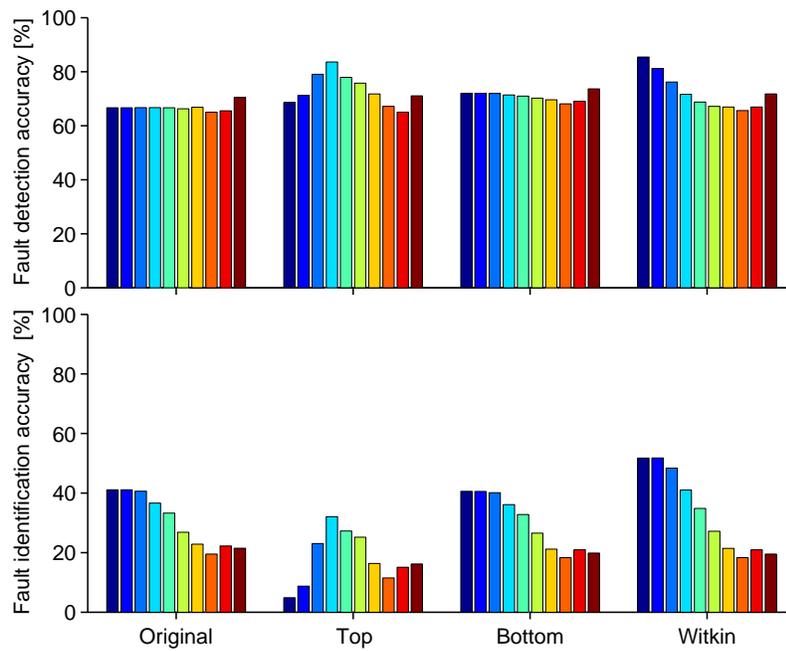


Figure C.19: Accuracy of QRT based Fault Detection (top) and Fault Identification (bottom) for each of the proposed methods.