Summary

This thesis deals with the development, application and validation of techniques for data analysis in view of supervisory control of cyclic systems, including and integrating aspects of monitoring, diagnosis and control. Two so far largely separated tools for data mining of process data are used as a basis for the presented developments. These are Principal Component Analysis (PCA) and Qualitative Representation of Trends (QRT). A pilot-scale sequencing batch reactor (SBR) for wastewater treatment is used as a case study for the major parts of the work presented. Another application is pursued regarding the analysis of flow measurement time series derived from an urban drinking water network.

The pilot-scale SBR setup studied throughout the work has served as a valid source of data-rich information-poor data sets, which are consequently used for evaluation of several data analysis tools for process supervision and control. It is observed that the design of the given experimental unit is suboptimal in view of the desire to discriminate between physical and biological failures of the system. In fact, an impending need exists to detect and diagnose biological faults, and the intent was to develop and evaluate techniques for this purpose. Several remarks and suggestions for improved design of experimental setups are given.

The first of the evaluated approaches for data analysis, Principal Component Analysis, is popularly reported as a straightforward method to deal with correlated measurements in the context of process monitoring and diagnosis. In view of fundamental limitations of the original (linear) PCA technique, a myriad of extensions and adjustments that deal with non-linearity, dynamics, natural changes of processes and the typical three-dimensional nature of batch process data (i.e. batch index, time-in-batch and measured variable), are already presented in literature. However, a lack of proper evaluation, validation and comparative studies is observed. In addition, the concepts of Maximum Likelihood and bias-variance trade-off, already theoretically presented for PCA modelling excercises, have so far been left unattended at large by the process monitoring research community.

A conventional extension for batch process data, Multi-way PCA (MPCA), is evaluated for process monitoring of both the hydraulic parts of the system only as well as for the multivariate data of the complete SBR system. For many types of faults, good performance rates are obtained by means of the MPCA models. In this, the performance proves to be rather insensitive to choices in the modelling stage such as the chosen approach for scaling of the data. Peculiar types of faults, which are of a time-local or frequency-local nature, are not likely detected by the presented approach. It is therefore concluded that process monitoring should integrate techniques that allow the detection of such events, e.g. by making use of the wavelet framework already presented and applied for process monitoring. Also, suggestions for improvement of PCA modelling practice, not limited to the case study, are added.

Modelling by means of MPCA is also used as a basis for process diagnosis. Explorative treatment of the diagnosis problem showed that MPCA is a straightforward tool to visualize and recognize faults. Despite this result, automatisation of this recognition process by means of combination of MPCA and fuzzy clustering is found to be applicable to a limited extent and therefore remains suitable for further research. Suggested modifications may lead to increased interpretability of the models, more generalized and appropriate representation of reality and, overall, improvement of diagnosis performance.

SBR systems are characterized by their cyclic operation and the recognition of several phases that constitute the cycles. The time length of cycles and phases is not necessarily the same for each cycle which allows large flexibility in design and operation. Generally speaking, the determination of the optimal lengths is not an easy task. Therefore, a new control scheme for phase length optimization is proposed, applied in real-time and evaluated as successful. In the proposed scheme, the Hotelling's T^2 statistic, which is also defined in the context of PCA, is used to define a region in the data space, corresponding to the targeted completion of biochemical reactions. If sufficient consecutive multivariate data samples are found within this region, the necessary reactions are judged to be complete signifying that the running phase can be shut down safely. The control scheme proved technically successful and is shown to lead to significant improvements in effluent quality of the studied SBR system, hereby representing a clear proof of concept.

The second approach used for automated data analysis is taken from the field of qualitative analysis. Qualitative analysis aims at the qualitative description of data and has so far largely been concentrated on the qualitative representation of time series. Following a review and evaluation of available techniques, it is concluded that proper identification of inflection points is not possible on the basis of the most generic techniques available in literature. One technique, based on the cubic spline wavelet decomposition, is selected and improved in such a way that inflection point detection is possible in a more reliable and consistent way. Suggestions for other observed problems, such as the inability to appropriately detect jump changes by means of the cubic spline wavelet method and the lack of robustness associated with the interval-halving approach found in literature, are proposed.

The qualitative analysis framework is succesfully evaluated for data mining of time series typical for urban drinking water networks. Qualitative analysis proved to reveal details of the studied time series that are not obvious from the more classic wavelet spectrum analysis technique. In an application to the studied SBR system, the qualitative analysis technique is integrated into a preliminary design for closed loop supervisory control on the basis of qualitative analysis. To this end, two simple tables, one connecting the qualitative analysis results with corresponding diagnostics and another connecting the diagnostics with proposed control actions, are devised.

In the closing chapter *Conclusions and perspectives*, the decision to develop monitoring schemes in the absence of knowledge, which was fundamental to the earlier choice for data-driven approaches, is evaluated critically. It is indicated that the presumed absence of knowledge proves not to correspond to the now historical reality. Based on the latter observation, mixed approaches, taking the best of knowledge-driven and data-driven approaches, are motivated for future practice in process analysis and supervision. In addition, several suggestions are given with respect to multivariate and qualitative analysis, including improvements and extensions of currently available techniques as well as new applications and opportunities for data analysis.