Doctoral Thesis

QUALITY CONTROL OF OPERATIONAL DATA FROM WASTEWATER TREAMTENT PLANTS

submitted in satisfaction of the requirements for the degree of
Doctor of Science in Civil Engineering
of the Vienna University of Technology, Faculty of Civil Engineering

---

Dissertation

ÜBERPRÜFUNG DER BETRIEBSDATEN VON ABWASSERREINIGUNGSANLAGEN

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaft
eingereicht an der Technischen Universität Wien Fakultät für Bauingenieurwesen
von

Dipl.-Ing. André Spindler
Matrikelnummer 0627171
Zur Quelle 9, 01731 Kreischa OT Saida, Deutschland

Gutachter: Em.O.Univ.Prof. Dipl.-Ing. Dr.techn. Dr.h.c. Helmut Kroiss
Institut für Wassergüte, Ressourcenmanagement und Abfallwirtschaft, TU Wien
Karlsplatz 13/226-1, 1040 Wien

Gutachter: Prof. Peter A. Vanrolleghem, MSc, PhD
Department of civil engineering and water engineering, Université Laval
Pavillon Adrien-Pouliot - 1065, Médecine avenue, Office 2974, Québec (Québec) - Canada - G1V 0A6

Wien, Jänner 2015

ABSTRACT

This thesis deals with questions of data quality control based on the principles of mass conservation. The focus is entirely on operational data from wastewater treatment plants. The goal was to provide a practically applicable method for the determination of well-balanced time periods associated with high data quality in historic data. CUSUM charts were found to be an appropriate way to evaluate the error vector of mass balances on a day-to-day basis. This method was called “Continuous mass balancing” and can also be applied for quasi-online monitoring of current operational data. Contrary to static mass balancing as commonly applied in the field of wastewater treatment, continuous mass balancing allows to incorporate the temporal redundancy contained in the data and can therefore detect even minor systematic errors. Flow dynamics (hydraulic retention), leading to delayed output of influent mass flows, have to be considered to achieve good balancing results. Accumulation would normally also need to be considered in short term balances. However, on the time scale relevant for continuous mass balancing its calculation was found to cause too much noise in the balancing error.

In addition to continuous mass balancing an algorithm was developed that allows the calculation of all possible balancing equations upon definition of the plant layout and measured and unmeasured variables in all streams. Flow is treated as an individual variable and therefore balancing equations are bilinear. The developed algorithm is based on structural redundancy analysis as known in data reconciliation.

There is hope that this thesis may help to close the existing gap between data quality evaluation in wastewater treatment and the powerful methods of data reconciliation developed in the field of process engineering.

CONTENTS

INTRODUCTION

Problem statement

Wastewater treatment is a key factor in modern water quality management. High legal standards in many countries require state-of-the-art technical solutions for municipal and industrial wastewater treatment. In sensitive areas of the EU, according to the Water Framework Directive 2000/60/EG, the "best available technique" has to be applied and in many regions of the world the necessary reuse of wastewater requires treatment to comparable technical standards.

This high standard of wastewater treatment fundamentally relies on well trained personnel. Experience shows that motivated personnel with a thorough understanding of the physical, chemical and biological processes are a key factor for reliable, sustainable and successful plant operation. To understand the behavior of the treatment plant in relation to the current requirements at any time, personnel depend heavily on measurement data. This is partly a consequence of the high level of automation but also of the fact that the characteristics of wastewater and sludge composition are not otherwise accessible for human perception. Reliable and correct measurement data therefore is an essential component of good wastewater treatment plant operation.

Besides plant operation there are, of course, more and equally important requirements for good measurement data quality. The design and especially the upgrading of wastewater treatment plants depend on sound measurement data. This applies to tank volumes, strongly correlating with construction costs, but also to equipment such as pumps and blowers. For the latter, adequate design is even more important as operating costs and even lifetime depend on optimal operation. So far, however, no scientific consensus has been reached upon the question which magnitude of (systematic) error is acceptable for wastewater treatment operational data.

Mathematical simulation of biological treatment processes has become a common tool for optimization of design and operation. The results of simulation studies follow the simple principle "garbage in - garbage out". As a consequence, the necessity of reliable input data is obvious. Until now, major simulation projects have usually relied on additional measurement campaigns to generate the input data. This is a very costly approach and probably has led to a considerable amount of simulation projects never being started. Additional measurement campaigns in most cases are only representative for a short time span and do not cover the year-round operational conditions of wastewater treatment plants. It follows that a method able to continuously ensure high reliability and correctness of all available operational data would strongly enhance the value of simulation tools for all purposes.

Last but not least, the documentation of the compliance with legal requirements has to be based on high quality data. Monitoring by the authorities can only be performed on a limited number of sampling occasions. This cannot be enough information for the continuous control of plant operation. If effluent quality is directly linked to fees for pollution loads (as is the case in Germany) this leads to the necessity of continuous control of the reported measurement data.

All these requirements on the quality of operational data from wastewater treatment plants are not met at the current situation. The characteristics of (municipal) wastewater, particularly the variability of flow and composition, make it difficult to obtain reliable and representative measurement data. In the reality of plant operation, the laboratory analysis itself can be one relevant source of error in measurement data. But even with the best level of quality assurance at the laboratory, sampling and sample treatment of wastewater and sludge still remain major sources of systematic and random errors. This is due to the unequal distribution and varying amount of solids. In addition automatic flow measurements, too, do not result in reliable and correct data all the time. From process benchmarking investigations in Austria (e.g. Lindtner 2008) it is well known that on municipal wastewater treatment plants an average of 5% - 10% of operating costs are spent for monitoring.

It can be concluded that the development of a method for continuous quality control of the monitoring data will contribute to better and more efficient design and operation of waste water treatment plants.

Quality evaluation for off-line data – an overview

The basis for good data quality is, once more, laid by reliable, well trained personnel. If workers on wastewater treatments plant know why a measurement is taken and how it can be biased, and if they are possibly even involved in the decision making that is based on their sampling and analysis, this is the perfect environment to achieve reliable data. According to the goals defined below, the following is mainly concerned with offline data, usually measured in laboratories.

Additional **parallel measurements** to verify data are not common on wastewater treatment plants due to two reasons. First, these measurements would in many cases be prone to the same type of errors as the original (operational) measurements. And secondly, with the considerable costs invested into monitoring already, additional measurements are difficult to convey. Therefore data quality should be assessed mainly within the existing data itself.

The trivial approach is simply by **plausibility testing**. Are the data values within a typical range? Is the temporal variability of data reasonable? Are there unexplained gaps in the data?

The more profound approach is to incorporate **redundancy** within the data. Redundancy can exist when similar measurements are taken in the same place, e.g. total nitrogen and ammonium in the influent or biological and chemical oxygen demand in any stream. Typical ratios between such measurements are known and can be verified. Another type of redundancy is derived from the **principals of mass conservation**. The total amount of an inert compound (e.g. an element) entering a system will either leave the system again or become accumulated in that system. If all mass flows of such a compound into and out of a system and the difference in the stored amount of that compound are measured over a certain time span, the sum of all mass flows into the system plus the possibly released load equals the sum of all mass flows out of the system plus the possibly accumulated load. This concept is known as mass balancing. If the laws of conservation are not obeyed by the measured mass flows, this indicates systematic measurement errors (or an erroneous system description, a case that should be ruled out at an early stage).

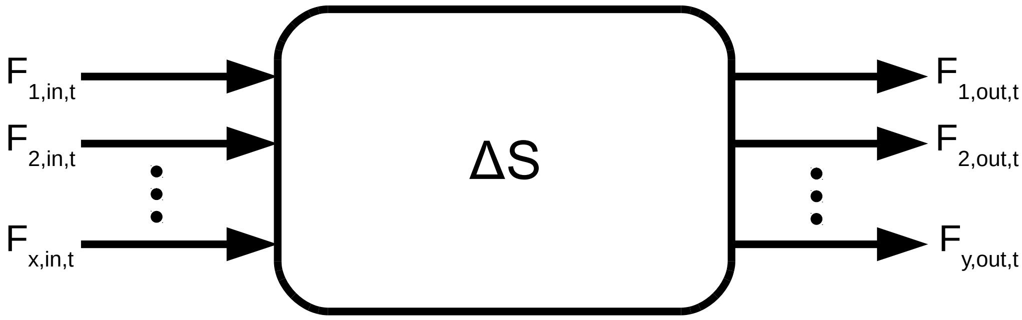

Figure 1: Simple balancing layout. Several fluxes may enter or leave a system, accumulation (ΔS) is possible.

Mass balancing is the basis for and therefore closely related to **data reconciliation**, which aims at improving overall data quality by finding the best approximate for each measurement so that all constraints (e.g. mass balances) are obeyed. This field has been widely investigated in process engineering and powerful methods have been developed. The question arises, as to which extend use could be made of these techniques regarding data from wastewater treatment. Three aspects relate to this question. First, data reconciliation requires relatively high quality data to begin with, e.g.

measurement variability is vital to be known. This is usually not a given fact in wastewater treatment facilities. Secondly, the dynamics of wastewater flow and composition as well as the bilinear nature of the data require more sophisticated approaches to reconcile data (nonlinear and dynamic methods). This relates to the third aspect. Operators of wastewater treatment plants are usually not experts in process engineering, but still need a profound understanding of the biological and physical processes taking place at their plants. Methods for data quality control therefore gain practical relevance with simplicity. The goal is not to provide operators with streamlined error-free data but to support fault localization and thus the process of understanding.

Simple mass balancing is an established and well known method in wastewater treatment, not so much data reconciliation. However, the implications and special requirements of mass balancing in wastewater treatment have hardly been investigated. Relevant contributions were made by Nowak (2000; see Spindler 2014) and Thomann (2002; see Spindler 2014). Both focus on static mass balances over long time spans that allow the assumption of steady state processes. Thomann (2002) also suggests including accumulation in order to allow balancing on a day-to-day basis, which he calls "dynamic balancing".

Goals

Even though the focus of this thesis changed slightly with time, it always remained concentrated on the interpretation of (real) operational data in regard to their quality. The original motivation came from the question, which deviation between input and output streams of a (static) mass balance would be admissible. Among experienced colleges, an error of 10% was widely accepted, up to 20% appeared reasonable. It soon became clear, that these assumed limits did not hold in light of the experience made during the course of this work. To achieve the maximum performance however, the temporal redundancy in data must not be neglected in mass balancing. This led to the main concern of this work changing towards a more continuously applicable method of data quality control.

This became indeed the main goal: to develop and proof the applicability of a method that allows both, the determination of error-free time periods in historic data and to continuously monitor the quality of data from wastewater treatment. "Continuous" in this context is restricted to the meaning of "on a day-to-day basis" because (offline) concentrations are usually measured in 24h composite samples giving one value per day. The naming "continuous mass balancing" was chosen mainly to convey the idea that operators at each given day, that means "always", could check on the quality of the data they are basing their decisions on.

Some aspects of so-called "continuous mass balancing" were investigated along with the development of the method itself. One fundamental aspect is to determine the complete set of theoretically possible and practically applicable balancing equations. While the latter part of this goal remains to be achieved, this type of structural (and hopefully later also practical) redundancy analysis also aids the determination of possibly sensible additional measurements. Another aspect is to investigate the possibilities of handling mass balances on a day-to-day basis, when the assumption of a steady state process cannot be maintained. While accumulation is the classical aspect that comes to mind in short term balancing, the effect of hydraulic retention also has to be considered on this time scale.

Methods

The work on this thesis started as a quest for a statistical basis for mass balancing wastewater treatment operational data. It wasn't clear in the beginning, what kind of methods would be used or which approaches would be followed. In fact, data reconciliation as such was totally unknown to the author. There had hardly been any applications of it in the scientific literature on wastewater treatment. And obviously no process engineer had taken on the challenge to establish a link between the two worlds. It wasn't until the author stumbled upon a paper by Van der Heijden (1994; see Spindler 2014) that he got in touch with this world. This paper wasn't actually very representative for process engineering and its state of research at the article's time, because it only translated the process oriented approach to elemental mass balances around a lab fermenter. It did, however, point to the determination of balanceability and calculability of measured and unmeasured data by matrix algebra. From there on it was clear that it would be worthwhile to further pursue the original question.

During the author's following stay with the model*EAU* group of the Canadian Research Chair on Water Quality Modeling it still wasn't clear for a long time, on which basis the (systematic) error of a mass balance should be evaluated. Only towards the end of this visit the application of CUSUM charts (Page 1954; see Spindler and Vanrolleghem 2012) led the way to answer this question. It had become clear, that the temporal redundancy of data in time series would have to be taken into account. With some knowledge about the classical methods of data reconciliation, however, there did not seem to be any way around the necessity of knowing all the measurements' variances. And because of the strong intention to use only readily available operational data, the possibility of multiple measurements for the determination of variances was ruled out. CUSUM charts are a method of statistical process control. Calculated as a special cumulative sum of consecutive measurement values, they signal when the process mean significantly deviates from its expected value. To apply CUSUM charts, balances had to be calculated on a daily basis which led to the introduction of the error vector (of day-to-day balances) and suddenly one was dealing with an expectedly stationary process, whose variability could be calculated easily.

Because CUSUM charts also consider past values, even minor deviations from the expected mean (in relation to the process's variance) can be detected reasonably fast and reliably. In the application to mass balances, the expected process mean (of the error vector) is always zero.

Because certain data (sludge concentrations of balanced components) are usually not measured in practice, some statistical assessments were required. The intention was to determine usually unmeasured variables from frequently measured data such as suspended solids. Data sets from three different large Austrian wastewater treatment plants where available for investigation, which was very important to remain in conformity with the intention of working with real data only. In these data sets all required sludge components had been measured at least weekly, in some cases as additional analyses in the author's institute's laboratory. Monte Carlo simulation was then applied to determine the minimum frequency of such measurements to ensure good approximations for the typically unmeasured data.

When the algorithm for an automated determination of balance equations was developed, the existing methods of data reconciliation were finally abandoned.

Individual balance equations are simply not necessary in data reconciliation. They do have the advantage of being more intuitively applicable for the practitioner who might not be a process engineer. Based on a matrix representation of all possible subsystem combinations of a given plant layout (the extended incidence matrix, Spindler 2014), individual equations were derived from the classification into redundant and non-redundant measured variables and calculable and non-calculable unmeasured variables. The necessary symbolic calculations to derive the individual balancing equations were executed by a computer algebra system, substituting calculable unmeasured variables with measured variables. Especially when concentrations of multiple compounds are measured, the resulting equations can be complex and therefore difficult to find otherwise.

Article summary

This thesis is composed of three articles, all written by the (first) author and supervised by the second author. The investigation of CUSUM charts for mass balancing of wastewater treatment operational data led to two articles, Spindler and Vanrolleghem (2012) and Spindler and Krampe (2015). A third article (Spindler, 2014) was written with the focus on structural redundancy of measurement data, providing a method for an automated setup of bilinear balancing equations. This article also intends to strengthen the connection between mass balancing as known and applied in wastewater treatment and the field of data reconciliation, broadly investigated in the process engineering domain.

The principal applicability of CUSUM charts for daily operational data from wastewater treatment was shown in Spindler and Vanrolleghem (2012). CUSUM charts were introduced and explained using a synthetic example. Practical application to two sets of flow data, one comprised of several influents and one effluent of a treatment plant, the other a flow balance over an anaerobic digester, revealed that measurement data that appears sufficiently well balanced on average over a long time period might very well consist of several poorly balanced shorter time periods with the single errors adding up to (almost) zero. The CUSUM chart, basically an integration of positive and negative errors, conveniently displays well balanced and poorly balanced time periods. The focus on flow data only allowed ruling out additional issues like accumulation or hydraulic retention. It also underlined the importance of well-balanced flow data, because these measurements are the basis for the calculation of mass flows from measured concentrations. Daily cumulative values for flow are usually available on virtually every wastewater treatment plant and mostly measured online. During this first investigation of CUSUM charts for mass balancing based on daily values it also turned out that the variability of the error vector (resulting from the single day-to-day balances) is an important indicator of data quality itself. A low variability of the error vector (with an expected mean of zero) indicates similar results for the single balances. This facilitates the detection even of small systematic errors by the method which inspires more confidence in overall data quality than wildly scattered random errors with a mean value of zero.

While the application of CUSUM charts for mass balancing was labeled "dynamic balancing" in the first article, this naming was subsequently changed to "continuous balancing". The term "dynamic" is strongly associated with biological modelling where "dynamics" are expressed by kinetic rates of microbial growth and chemical reactions. "Continuous" is also not quite exact as described above. However, the term "discrete step mass balancing" is likewise hardly suited to communicate an easily applicable method to the practitioner.

The second article on the application of CUSUM charts (Spindler and Krampe, 2015) was based on a research project financed by the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management. Several aspects of great practical relevance are investigated. Generally, this article is concerned with (bilinear) mass balances rather than (linear) flow balances and gives a number of real data examples. Typically balanced sewage sludge components such as COD, TP and TN are usually measured rarely, sometimes not at all. Statistically analyzing data sets for primary sludge, waste activated sludge and digested sludge from three different treatment plants, it was shown that in most cases these sludge components can be determined reliably from practically more convenient and therefore more regular measurements of (total or volatile) suspended solids. The precondition for this determination is the monthly measurement of the relevant sludge components which will usually have to be carried out by an external laboratory. A linear dependency between total or volatile suspended solids and the respective sludge component had to be superimposed by a seasonal component in most combinations of sludge types and components to give the best results.

The second aspect of continuous mass balancing covered in Spindler and Krampe (2015) deals with the influences from accumulation and hydraulic retention. Accumulation (release) in reactors with a fixed volume occurs, when a component's concentration rises (drops). Stemming from the first aspect introduced above, this often has to do with increased suspended solids concentrations in tanks. Surprisingly, the consideration of accumulation led to a deterioration of the error vector variability. This in turn made it more difficult to distinguish well balanced from poorly balanced time periods. It is assumed that this effect was caused by the daily accumulation being calculated from differentials, and their integration (by the CUSUM method) is known to amplify noise. And the measurement of suspended solids itself is quite likely to introduce that noise into the equation, as representative sludge samples are often difficult to obtain.

When hydraulic retention was regarded instead of accumulation, continuous balancing gave considerably better results. Owing to the nature of wastewater and sludge treatment (to a large extend based on phase separation) there are usually at least two output streams from a subsystem. Often one of those carries components that have a retention time well above one day. Therefore it is clear that component loads entering such a subsystem on one day will not necessarily leave it entirely on the same day but rather distributed over a long period depending on the retention time of that component in that subsystem. With the assumption of an ideal CSTR for the respective subsystem under evaluation, this behavior can be integrated into mass balancing and the expected output load (calculated from the measured input load and the starting concentration in the tank) is balanced against the measured output load.

The third article (Spindler, 2014), chronologically the second, covers an aspect of mass balancing independent from the measured data itself. Derived from the methods of structural redundancy analysis and based on a complete system description together with the information about the availability of measurements in each stream and reactor, a method is introduced that allows to automatically set up all theoretically possible mass balance equations for a system. Due to the bilinear nature of mass flows this can result in non-trivial solutions, especially when multiple components are allowed. In these cases, missing flow measurements can be substituted by available concentration measurements. The method also allows for a simple investigation about the effect of additional measurements on the overall balanceability (redundancy) of the system.

Scientific contribution

As of today, a clear distinction has to be made between data reconciliation in the world of process engineering and data quality evaluation in wastewater treatment. In process engineering the profit driven development has produced a vast amount of powerful techniques for the reconciliation of measurement values that supports ever more precise control of production. However, while the knowledge of each measurement's variability is a crucial element in most of these techniques and variability of (mass) flows is in most cases reduced to the minimum in most process engineering applications, the contrary is the case in wastewater treatment. The main disturbance to the whole system is the more or less uncontrollable influent (flow and composition!). Adding to this is the fact that wastewater treatment is a negligible economic factor, driven by legal requirements and the demand for environmental protection. The effect of these preconditions are simply less frequent and less reliable measurement values. This thesis has been an attempt to maximize the information contained in typically available operational data from wastewater treatment by aiding operators and other stakeholders to verify the data quality. It therefore belongs to the gross error detection part of data reconciliation.

There has so far not been a profound investigation on the application of CUSUM charts for mass balancing in the field of wastewater treatment. Zaher and Vanrolleghem (2003; see Spindler and Vanrolleghem, 2012) named this possibility among others without going into details. CUSUM charts have now been proven suitable for continuous mass balancing even though a number of open questions remain to be answered. Although not addressed directly in this thesis, the possibility of the application of CUSUM charts to characteristic values calculated from plant data is obvious. Characteristic values (sometimes also known as expert knowledge) such as the specific amount of volatile suspended solids in digested sludge (around 18 g VSS/pe/d) or the typical specific energy demand for aeration can be used as a target value (instead of the mean balancing error) of a CUSUM chart. Their usefulness is comparable to that of classical balances but they often require less input data.

The determination of component loads in sludges from total or volatile suspended solids, though regularly applied under the assumption of direct proportionality, has never been based on a thorough statistical examination. As it turned out, direct proportionality is sometimes given but cannot be expected in every case. The range of typical ratios (when direct proportionality is suitable) varies considerably which implies a low probability of free assumptions to be correct. Operators can clearly improve the general balanceability of their wastewater treatment plants by having samples of their sludges analyzed monthly in an external laboratory.

Regarding the automatic determination of bilinear balancing equations, to the author's knowledge no such algorithm has been published before in wastewater treatment literature or related fields. This probably results from process engineering's data reconciliation aiming at entire datasets at once, not at individual (subsystem) balances. Non-trivial balancing equations have until now hardly been used in wastewater treatment practice.

Conclusions

Continuous mass balancing has the potential to define a new standard in quality control of wastewater treatment operational data. It gives plant operators a possibility to evaluate the general integrity of their measurements on a daily basis. The real data analyzed so far allows the conclusion that continuous mass balancing can easily determine even minor systematic errors in data (well below 5% of the input load when hydraulic retention is considered). This means, the commonly assumed 10% permissible error (or more) should be abandoned. In scientific studies based on plant data the proof of good data quality should become a matter of course before any conclusions are drawn.

A considerable number of aspects remain to be dealt with. Until now, accumulation and hydraulic retention in continuous mass balances have been dealt with separately. Although the calculation of accumulation has been shown to increase random error, it might be feasible to include along with hydraulic retention if data are filtered in an appropriate way. Typically Kalman filtering would be used in this case. It remains to be shown if accumulation does play a significant role when data are analyzed on a daily basis. Negligible on long term balances, accumulation is likely to have its maximum significance in balancing periods of around one sludge retention time of the balanced subsystem.

More practical experience is needed, although continuous mass balancing gave good results with the real data it has already been applied to. It would be especially beneficial if detected faults in data could be confirmed by expert knowledge. The recently much intensified application of the Benchmark

Simulation Model (Gernaey et al., 2014) in many areas of wastewater treatment study would probably be an appropriate way to better assess the reliability of systematic errors detected by continuous mass balancing. It could also be applied to answer a number of additional questions.

Still missing is a general assessment of the practical possibility of quality control for the single variables measured in a wastewater treatment system. It appears quite likely, that a number of measurements remain practically not redundant. For example total phosphorus or COD in the effluent have such minor effect on their respective balances, that quality control of these data might remain inaccessible by the means of mass balancing. This question is similar to the determination of identifiability of individual parameters in modelling and could probably also be investigated using the Benchmark Simulation Model. The developed algorithm for automatic determination of balance equations could also be extended by an appropriate sensitivity analysis. Further improvement of this algorithm is probably possible by the application of graph theory (Deo, 1994) to determine the initial set of theoretically possible balance equations.

If, as expected, some typical measurement values in fact do remain non-verifiable by mass balancing, other means of verification should be applied regularly. In the case of effluent concentrations this is usually realized already through external control by the authorities. Further, the question of missing data has not yet been properly addressed. In smaller wastewater treatment plants operational data are typically not measured on a daily basis and therefore much information is missing. It might, however, still be feasible to ignore missing data and to find a compromise about the minimum time span that gives one data point for continuous mass balancing. A monthly average of available data values might actually proof a suitable input for the CUSUM chart and provide a still more detailed analysis than a static balance when the considered time span is long enough, maybe 2 years or more.

Finally, the influence of autocorrelation on CUSUM charts in their proposed application remains to be investigated. CUSUM charts are known to be sensitive to autocorrelation. Wastewater treatment data is clearly autocorrelated. However, it is not clear that the error vector of a continuous mass balance is autocorrelated, too. Even under consideration of hydraulic retention which itself is calculated in an autoregressive way, the error vector of a continuous mass balance should actually be only noise as long as no systematic error is present. The investigation into this question is probably best considered after the practical applicability of continuous mass balancing has been confirmed further.

In this thesis the intention was not to avoid the merits of data reconciliation. Obviously, the connection between two similar, though not equal, fields - wastewater treatment and process engineering - is not very strong at this time. There is hope that this work will help to bridge the existing gap. It would be a great success, too, if this work would stimulate contributions by scientists who are well familiar with data reconciliation but at the same time well aware of the special implications of wastewater treatment.

In the future, plant operators, administration, engineers and scientists should no longer be in doubt upon first contact with plant data. It is at the hands of operators to have their measurements organized and monitored in such a way, that reliable data quality can be proven at any time. This will considerably shorten the time of typical data evaluations including the corresponding cost savings. For simulation studies, virtually no additional effort should be necessary any more, once the simulation model has been initially set up and calibrated. Today we are still a considerable distance away from this situation. It is the firm conviction of the author, that the here described methods and approaches open a practically feasible way to achieve this scenario.

References not cited in articles

Deo, N. (1994) *Graph Theory: with Applications to Engineering and Computer Science*, New Delhi, Englewood Cliffs [N.J.], Prentice-Hall of India, Prentice-Hall International.

Gernaey, K. V., Jeppsson, U., Vanrolleghem, P. A. and Copp, J. B (eds) (2014) *Benchmarking of Control Strategies for Wastewater Treatment Plants (IWA Scientific and Technical Report)*. IWA Publishing, London.

Lindtner, S., Schaar, H., and Kroiss, H. (2008) Benchmarking of large municipal wastewater treatment plants greater than 100,000 pe in Austria. *Proceedings of the Water Environment Federation*, 2008(7), 7655–7657.

ARTICLES

This thesis is a cumulative work consisting of three publications in international peer-reviewed scientific journals. All articles where written by the author, who proposed the methods, developed and implemented the algorithms and evaluated the data. Article 1 and article 3 were supervised by the co-authors.

Article 1

Spindler, A. and Vanrolleghem, P. A. (2012) Dynamic mass balancing for wastewater treatment data quality control using CUSUM charts. *Water Science and Technology*, 65(12), 2148–2153.

Article 2

Spindler, A. (2014) Structural redundancy of data from wastewater treatment systems. Determination of individual balance equations. *Water Research*, 57, 193–201.

Article 3

Spindler, A. and Krampe, J. (2015) Quality control of wastewater treatment operational data by continuous mass balancing: Dealing with missing measurements and delayed outputs. *Water Quality Research Journal of Canada* (accepted 08/01/2015).