Villez, K., Corominas, L., Vanrolleghem, P. A. (2013). Structural observability and redundancy classification for sensor networks in wastewater systems. Proceedings of the 11th IWA conference on instrumentation control and automation (ICA2013), Narbonne, FR, Sept. 18-20, 2013, Appeared on USB-stick (IWA-12741).

Structural observability and redundancy classification for sensor networks in wastewater systems

K. Villez[1], P.A. Vanrolleghem[2], and Ll. Corominas[3]
[1] Spike, Eawag, Überlandstrasse 133, 8600 Dübendorf, Switzerland
[2] model*EAU*, Université Laval, 1065, Avenue de la Médecine. Québec G1V 0A6, QC, Canada
[3] ICRA, Catalan Institute for Water Research, Carrer Emili Grahit, 101, E- 17003 Girona, Spain

Abstract. The design of sensor networks in terms of sensor number, sensor locations and sensor types is a challenging task in the context of both waste water collection and treatment systems. Because in practice the reliability of sensors can be low it is useful to have some level of redundancy in the sensor network to detect and diagnose sensor faults on-line. However, such redundancy comes at a cost and means that under budget constraints certain variables cannot be measured or estimated. In this work, we provide the initial step towards optimization of sensor networks. In particular, we report on the application of a graph-theoretical method for the classification of variables as observable or unobservable and sensors as redundant or non-redundant for a given sensor configuration. Importantly, this analysis is based on the structural properties of the monitored system which imply always-valid flow and mass balances and does not rely on data obtained from the sensors themselves or on detailed understanding of the process dynamics.

Key words: Bilinear mass balancing, Fault detection and identification, Sensor network design, Structural observability and redundancy

1. Introduction

So far, fault detection and diagnosis in environmental engineering has focused on the characterization of systematic faults in sensing equipment. This includes detection (presence of a fault), isolation (which sensor is faulty) and characterization of sensor faults (identification, e.g. type and magnitude) (Aguado & Rosén, 2008; Puig et al., 2008; Corominas et al., 2011; Spindler & Vanrolleghem, 2012; amongst others). Limited attention has been given to the placement of sensors in such a way that fault detection and diagnosis tasks become easier. Known methods for sensor placement are based on direct hardware redundancy. Indeed, by placing two sensors of the same type in one location one can detect faults in either sensor as long as the start of a fault does not coincide with the start of a fault in another sensor. Isolation and identification are also possible as soon as one places three sensors of the same type in one location, again assuming that the simultaneous onset of sensor faults in multiple sensors is impossible. Unfortunately, such placement aimed at redundancy by means of replication does not take advantage of spatial relationships between measured variables such as those defined by mass balances. Furthermore, budget constraints may mean that the redundancy obtained for sensors in one location, signifies the loss of observability of other variables or in different locations. In contrast, fault detection and identification methods are often based on spatial relationships. Most commonly known are those based on flow, mass and energy balances in which measurements of mass flow, concentrations and temperatures can all be considered in one framework (Schraa & Crowe, 1998). However, this remains a new paradigm in the context of wastewater engineering.

Figure 1 shows the strategy we intend to develop for optimal sensor placement prior to data collection. An automated, computer-based sensor and variable classification method is used to assess which sensors are redundant and which variables are observable. To this end, a variable is considered observable if it is measured directly or if it can be computed by means of system equations (e.g. mass balances) from other available measurements. A sensor is considered redundant if the variable which it measures remains observable if that particular sensor is removed.

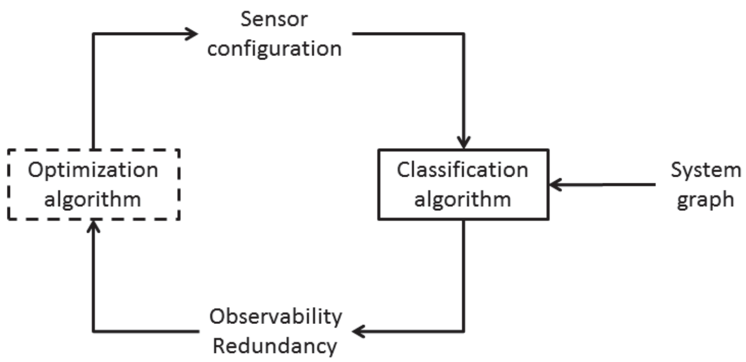

Figure 1: Schematic overview of integrated methods. A classification algorithm determines which variables are observable and which sensors are redundant given a graph-theoretical representation of the monitored system and a given configuration of sensors. Based on such classification algorithm, the sensor configuration can be optimized so to have desirable properties in terms of observability and redundancy. The current work deals with the classification algorithm only.

Such redundancy is equivalent to fault detectability as long as a sensor fault appears in a single sensor at a single time. Indeed, if a fault appears in the considered sensor then the difference between the measurement and the estimate of the same variable obtained using all sensors except the faulty one is a good indicator for the presence of a fault. The above classification of variables and sensors is fixed for a given sensor configuration, i.e. the location and type of sensors in a given process or system. If such classification method exists, then it can be used in an iterative optimization scheme by which the sensor configuration is improved. Optimality can be defined as maximum observability or maximum redundancy or a mixture of both.

In this work, we focus on the classification method, and not the optimization, based on existing methods in the literature (Kretsovalis and Mah, 1987, 1988a, 1988b). They are based on a graph-theoretical representation of process systems and enable labelling of sensors as either redundant or non-redundant as well as the labelling of variables as either observable or unobservable for a given sensor configuration (number, type, and location of sensors). While the algorithms to do this are not trivial, they can be implemented to allow fully automated classification for any plant and sensor configuration and do not rely on any actual measurements so that sensor network design prior to any data collection campaign is feasible. In this study, we apply the classification rules to the long term Benchmark Simulation Model (BSM_LT) for a given configuration of the sensing equipment in a WWTP (known number of sensors, measured variables and measurement locations).

2 Method

2.1 Studied system

The simulation platform used is the Long-Term Benchmark Simulation Model (BSM1_LT) (Rosen *et al*., 2004), which was developed to provide a framework to objectively evaluate process monitoring, diagnostic and automation strategies of WWTP. This platform includes model, process configuration (pre-denitrification plant with five activated sludge units in series, two anoxic –ASU1 and ASU2- and 3 aerobic -ASU3 to ASU5), inputs (influent, temperature, inhibition), control period, control systems, benchmarking procedures and evaluation criteria for process and controller performance. In Corominas et al. (2011) this platform was extended with a procedure for evaluation of monitoring performance. It comprises a one year evaluation period with a dynamic influent, includes temperature-dependent and inhibition kinetics.

2.2 Graph theoretical method

The rules as provided by Kretsovalis and Mah (1987, 1988a, 1988b) for classification of variables and sensors were applied to the BSM1_LT platform. This allows classifying flow and concentration variables as *observable* which means that the considered variable is either measured directly or can be computed from the available data. Similarly, a fault in a particular measurement is classified as (*structurally) detectable* if the corresponding variable remains observable when the considered measurement is removed from the data set, i.e. there is redundant information about this variable.

For purpose of variable and sensor classification, the BSM1_LT model is represented as a graph. This graph consists of nodes (also: vertices) representing all flow junctions and unit processes (reactor, settler) in the plant and edges (also: links) representing the flows between the nodes. Importantly, the graphical representation of the plant implies a steady-state assumption on all mass balances. For this reason, it is typical in fault detection and identification applications to average hydraulic and mass flows over a pre-determined time interval so to attenuate temporary imbalances over the nodes. In the particular case of conventional WWTPs, this is due to accumulation of mass in the reactor or settler. Alternatively, it is possible to account for reactions by expanding the existing graphical model with imaginary flows to the out-of-system environment. This allows mass balancing at the frequency of measurement and not at a time scale that is sufficiently long to attenuate any temporary imbalance in the system. Both approaches are studied and compared here.

3. Results

Figure 2 (top) shows the BSM1_LT plant as presented in a typical simulator. The system consists of 5 tank units which together represent the reactor system as well as a settler. Influent flow, sludge recycle flow, mixed liquor recycle flow and carbon dosage flow are all joined at a single point in front of the reactor system. The reactor system is followed by the settler. The underflow from the settler is separated into the recycle flow (RAS) and the wastage flow (WAS). The system is equipped with flow and TSS sensors as indicated in the graph. No other sensors are considered. Below this scheme is the graph-theoretical representation of the same system. In such representation, each unit process is represented as a node (also: vertex) and each flow between unit processes as an edge (also: arc). Importantly, splitters and junctions are also considered unit processes in this representation. For instance, node 4 represents the reactor system while node 6 represents the settler. For the BSM1_LT plant, temporary accumulation of TSS is considered to happen in the reactor and in the settler. For this reason, imaginary flows from the environment to the reactor and settler and vice versa are added (edges 13 to 16). One considers a fixed flow rate for each and assumes the TSS concentrations unknown (but possibly observable). Moreover, the flow rate for each edge of a pair of imaginary flows between the same two nodes must be equal to ensure hydraulic balance satisfaction (flow rates for edges 13 and 14, resp. 15 and 16, are the same).The complete graph consists of 8 nodes (indexed 1 to 8), 12 real edges (indexed 1 to 12) and 4 imaginary edges (indexed 13 to 16). Note that one of the nodes (node 8) represents the environment (outside of the system boundaries) and thereby represents the mass balances over the whole plant. If one ignores the imaginary edges, one can write 8 independent instances of for flow balances (flow only) and mass balances (flow x concentration) involving 24 variables (flow rates and TSS concentrations). The splitter represents two more independent equations which equal the TSS concentrations in (i) the settler underflow, (ii) the WAS flow and (iii) the recycle flow.

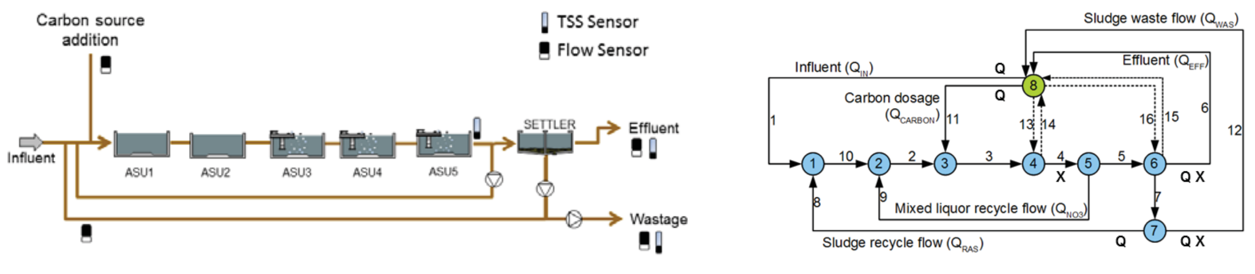

Figure 2: BSM_LT plant representations. Left: Scheme as typically shown in simulator software with flow and TSS sensors as icons. Right: Graph-theoretical representation with flow and TSS sensors as letters Q and X.

Table 2 displays the variable and sensor classification results. The classification of the flow rates is insensitive to the chosen approach (with/without imaginary flows), in contrast to the classification for TSS. All flow rate measurements (indexes 1, 6, 8, 11 and 12) are redundant and all flow rates are observable. All TSS measurements are observable without use of imaginary flows (column TSS(1)), thus assuming steady-state operation. The measurements for the installed sensors (flow indexes 4, 6 and 12) are redundant and faults in these sensors can thus be detected. With imaginary flows (column TSS(2)), thus allowing accumulation of TSS in the reactors and settler, TSS remains observable only in flows 5, 7, 8, 9 and 11 and none of the installed sensors are redundant.

Table 2. Classification of sensors and variables. For non-measured variables the classification is either observable (O) or non-observable (empty cell). All measured variables are automatically observable. For measured variables, R indicates the sensor to be redundant and N indicates that it is non-redundant. For flow rates, the classifications are the same irrespective of whether one adds imaginary flows to the graph or not. For TSS, results are presented for the case without imaginary flows (1) and with imaginary flows (2).

Flow index	1	2	3	4	5	6	7	8	9	10	11	12
Flows (Q)	O, R	O	O	O	O	O, R	O	O	O	O	O, R	O, R
TSS (1)	O	O	O	O, R	O	O, R	O	O	O	O	O	O, R
TSS (2)				O, N	O	O, N	O	O	O		O	O, N

4. Conclusions and perspectives

This work presents the early results obtained with a graph-theoretical approach to structural observability and redundancy classification. The BSM1_LT functions as benchmark and test platform for the proposed algorithms, which allow fully automated classification of sensors and process variables for any plant and sensor configuration. Future work includes further testing of other sensor configurations for the BSM1_LT plant and global optimization of its sensor configuration. Beyond this, we foresee the application of these methods to larger networks, including waste water collection systems and water distribution networks.

5. Acknowledgments

P. A. Vanrolleghem holds the Canada Research Chair in Water Quality Modeling. Lluís Corominas acknowledges funding from Marie Curie Career Integration Grant (PCIG9-GA-2011-293535), ENDERUS (CTM-2009-13018) project and “Juan de la Cierva” grant (jci-2009-05604).

6. References

Corominas, Ll., Villez, K., Aguado, D., Rieger, L., Rosén, C. and Vanrolleghem, P.A. (2011). Performance evaluation of fault detection methods for wastewater treatment processes. Biotechnol. Bioeng. 108(2), 333-344.

Aguado, D. and Rosén, C. (2008). Multivariate statistical monitoring of continuous wastewater treatment plants. Engineering Applications of Artificial Intelligence, 21, 1080–1091.

Kretsovalis, A., and Mah, R.S.H. (1987). Observability and redundancy classification in multicomponent process networks. AIChE Journal, 33(1), 70-82.

Kretsovalis, A., and Mah, R.S.H. (1988a). Observability and redundancy classification in generalized process networks - I. Theorems. Computers and Chemical Engineering, 12(7), 671-687.

Kretsovalis, A., and Mah, R.S.H. (1988b). Observability and redundancy classification in generalized process networks - II. Algorithms. Computers and Chemical Engineering, 12(7), 689-703.

Puig, S., van Loosdrecht, M.C.M., Colprim, J. and Meijer, S.C.F. (2008). Data evaluation of full-scale wastewater treatment plants by mass balance. Water Research 42 (18), 4645–55.

Rosen, C., Jeppsson, U. and Vanrolleghem, P.A. (2004). Towards a common benchmark for long-term process control and monitoring performance evaluation. Water Science and Technology, 50(11):41–49.

Schraa, O.J. and Crowe, C.M. (1998). The numerical solution of bilinear data reconciliation problems using unconstrained optimization methods. Computers and Chemical Engineering, 22, 1215-1228.

Spindler, A. and Vanrolleghem, P. A. (2012). Dynamic mass balancing for WWT data quality control. Water Science and Technology, 65(12), 2148-2153.