KV_C047
Villez, K., Vanrolleghem, P. A., Corominas, L. (2015). Optimizing observability and redundancy by means of globally optimal measurement selection. Proceedings of the 9th IWA Symposium on Systems Analysis and Integrated Assessment (Watermatex2015), Gold Coast, Queensland, Australia, June 14-17, 2015, appeared on-line.

Watermatex 2015

9th IWA Symposium on Systems Analysis and Integrated Assessment
14-17 June 2015, Gold Coast, Australia

Optimizing observability and redundancy by means of globally optimal measurement selection

Villez, K.*, Vanrolleghem, P. A. and Corominas, Ll.*****
* Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland (E-mail: kris.villez@eawag.ch)
** modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065, Avenue de la Médecine. Québec G1V 0A6, QC, Canada (E-mail: peter.vanrolleghem@gci.ulaval.ca)
*** ICRA, Catalan Institute for Water Research, Carrer Emili Grahit, 101, E- 17003 Girona, Spain (E-mail: lcorominas@icra.cat)

Keywords: deterministic optimization, graph theory, measurement networks, sensor placement, wastewater treatment

Summary of key findings

In this work, a deterministic global optimization approach to the selection of sampling locations and measurement types is tested on a small textbook example of a wastewater treatment plant (WWTP). This selection is aimed at the simultaneous optimization of (i) the number of selected measurements, (ii) the number of variables that can be estimated on the basis of these measurements (observability), and (iii) the fraction of selected measurements which are redundant. This approach allows to set up a measurement campaign or a sensor network before sampling, sensor installation, or data collection is started. Current results, while preliminary, suggest that this is possible thanks to graph-theory based algorithms for observability and redundancy evaluation and deterministic optimization schemes.

Background and relevance

Obtaining high quality data in environmental engineering systems is a challenging task. Thus far, many studies have focused on the evaluation of data quality a posteriori, i.e., after the data has been collected. This can be based on mechanistic models, such as mass balancing equations (e.g., Spindler & Vanrolleghem, 2012; Villez et al., 2013a) and dynamic models (Villez et al., 2011), or empirical data models. However, the success of these methods in detecting and identifying faulty measurements heavily relies on the availability of redundancy relationships between the measured variables. In practice, little is done to ensure the presence of such redundancy prior to data collection. However, the selection of measured variables and their sampling location can substantially affect the potential for automated fault detection and identification. For this reason, this study focuses on the optimal selection of measured variables and their sampling location in a WWTP. To this day, limited research has been conducted to evaluate how this can be executed best. In this work, it is explored how one can exploit mass balancing equations and deterministic optimization schemes to simultaneously minimize the number of taken measurements, maximize the number of variables (flow rates and concentrations) that can be estimated (observability) and maximize the fraction of chosen measurement locations which are redundant and for which faults are detectable on the basis of mass balances (redundancy).

Results and discussion

Methods. The above described problem is a multi-objective problem which is solved by means of deterministic global optimization. In particular, current results have been obtained with a branch-and-bound algorithm (Nemhauser & Wolsey, 1988) for multi-objective optimization (Ehrgott & Gandibleux, 2002). The considered criteria are explicitly described as follows:

1. Minimize the number of measurements, z_C.
2. Minimize the number of structurally unobservable variables (flow rates, concentrations), z_O, with the following definition: A variable is considered structurally observable when *(i)* a direct measurement is available as a value for the considered variable or *(ii)* other measurements are available which, in combination with a mathematical representation of the measured process or system, permit computation of a unique value (estimate) for the considered variable.
3. Minimize the fraction of measurements which are structurally non-redundant, z_R, in the following sense: A measurement is considered structurally redundant if the measured variable remains

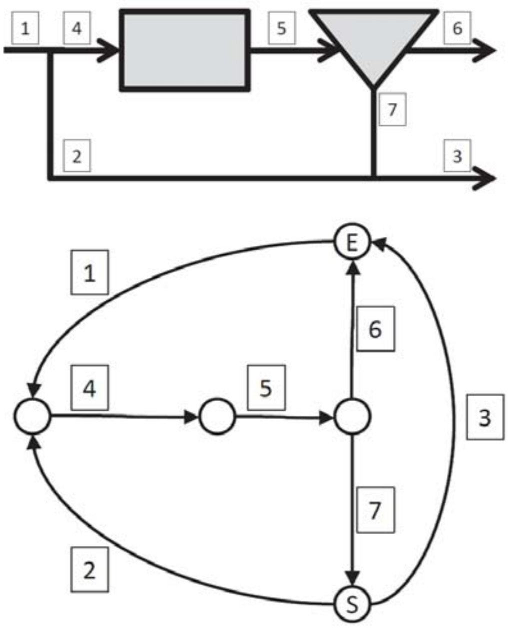

Figure 1. Scheme (left) and graph-theoretic representation (right) of a text-book WWTP.

observable when the considered measurement is removed from the set of available measurements. Evaluation of structural observability and redundancy for a given set of mass balancing equations and set of measurement locations is a challenge of its own (e.g., Ponzoni, et al., 1999). In this work, the algorithms of Kretsovalis & Mah (1988a, 1998b) have been applied to execute this task (Villez et al., 2013b). These algorithms are based on a topological graph representation of the network of unit processes and are applicable to any biological or chemical plant, including WWTPs.

Case study: A textbook wastewater treatment plant. A simple WWTP configuration has been selected for software development and demonstration purposes. This plant configuration consists of seven (7) streams connecting four (4) junctions as depicted in Fig. 1. It contains a single bioreactor, a settler, and one recycle stream. The considered variables are the flow rates and Total Suspended Solids (TSS) measurements. In this preliminary study, the plant is considered to be in steady state. In particular, the net effect of TSS production, TSS degradation, and TSS storage is zero in all plant locations. Practically, this means reactions with TSS and storage of the TSS component can be ignored for sensor layout evaluation and optimization. Each stream is considered to consist of two fractions, the TSS fraction and the remainder fraction. The remainder fraction represents all other components of the streams, including water, which are not measured or considered of particular interest for estimation.

Optimization results. The possibility to individually decide to install any of the 14 candidate sensors (7 flow meters and 7 TSS sensors) leads to a total of 16384 (214) sensor layouts. It is needless to say that even for a small example like this, enumeration and evaluation of all sensor layouts should be avoided in as much as possible. By means of the branch-and-bound algorithm, the GENOBS and GENRED algorithms were executed for 8352 distinct sensor layouts. This means 8032 sensor layouts had not to be evaluated, signifying a 49% reduction in explored measurement layouts compared to brute-force enumeration. The evaluation of the Pareto front was completed in just under 3 hours and 1257 distinct sensor layouts were retained in the Pareto set. This means that only 7.6% of all possible sensor layouts are part of the Pareto-optimal solution set. The Pareto front is visualized in Fig. 2. Quite clearly, many sensor layouts on the Pareto front lead to the same combination of objective values since only 12 unique objective value combinations can be found on the Pareto front. Three extreme combinations for the objective values can be found. This includes *(i)* the (trivial) solution without sensors and consequently no observable variables or redundant sensors (red circle, z_C =0, z_O

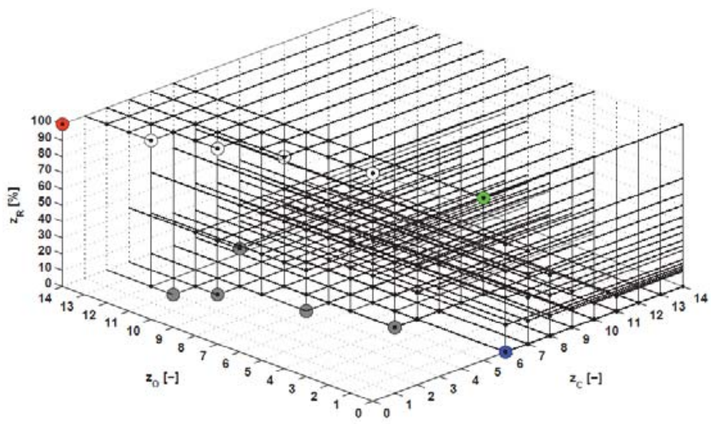

Figure 2. Visualization of the Pareto front by means of the considered objective values. Circles indicate the Pareto front sensor layouts whereas the dots indicate all evaluated sensor layouts which were evaluated during optimization. The red circle indicates the no-sensor layout. The green circles indicate layouts with all variables observable without redundancy. The blue circle indicates the sensor layouts with all variables observable and all-redundant sensors. White circles indicate sensor layouts without redundancy. Grey circles indicate sensor layouts with at least one redundant sensor.

=14, z_R =100%), *(ii)* 707 solutions with five (5) sensors, all variables observable, and without redundant sensors (green circle, z_C =5, z_O =0, z_R =100%), and *(iii)* 148 solutions with six (6) sensors, all variables observable, and all sensors redundant (blue circle, z_C =6, z_O =0, z_R =0%). Of the remaining solutions, *(iv)* 311 sensor layouts do not exhibit any redundant sensor (white circles, z_R =100%) and *(v)* 72 sensor layouts exhibit at least one redundant sensor while some variables remain unobservable (grey circles, $0 < z_R < 14$; $z_R < 100\%$).

Discussion. The currently available results indicate that global Pareto front optimization for measurement selection is possible prior to data collection on the basis of structural observability and redundancy criteria. Despite the advantages of using deterministic optimization schemes (guaranteed global optimality with minimal function evaluations), it remains to be seen whether the method can be scaled to larger WWTPs (e.g, the Benchmark Simulation Model No. 1, Gernaey *et al.*, 2014).

Acknowledgments

This work is possible thanks to a toolbox for graph theory applications (Iglin, 2003) and a toolbox for optimization (Villez et al., 2013c). Peter Vanrolleghem holds the Canada Research Chair in Water Quality Modelling. Lluís Corominas acknowledges the FP7 Marie Curie Grant (PCIG9-GA-2011-293535), and acknowledges MINECO from Spain for the "Ramon y Cajal" grant (RYC-2013-14595).

References

Ehrgott M.; Gandibleux X (2002). *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys.* Springer, Heidelberg, Germany.

Gernaey, K. V., Jeppsson, U., Vanrolleghem, P. A., & Copp, J. B. (Eds.). (2014). *Benchmarking of Control Strategies for Wastewater Treatment Plants.* Scientific and Technical Report No 23. IWA Publishing, London, UK.

Iglin S (2003). *grTheory - Graph Theory Toolbox.* http://www.mathworks.com/matlabcentral/fileexchange/4266-grtheory-graph-theory-toolbox .

Kretsovalis, A., & Mah, R. S. H. (1988a). Observability and redundancy classification in generalized process networks—I. Theorems. *Comp. Chem. Eng.*, 12(7), 671-687.

Kretsovalis, A., & Mah, R. S. H. (1988b). Observability and redundancy classification in generalized process networks—II. Algorithms. *Comp. Chem. Eng.*, 12(7), 689-703.

Nemhauser G. L.; Wolsey L A (1988). *Integer and Combinatorial Optimization.* Wiley, New York.

Ponzoni, I., Sánchez, M. C., & Brignole, N. B. (1999). A new structural algorithm for observability classification. *Ind. Eng. Chem. Res.*, 38(8), 3027-3035.

Spindler, A., & Vanrolleghem, P. A. (2012). Dynamic mass balancing for wastewater treatment data quality control using CUSUM charts. *Water Sci. Technol.*, 65(12), 2148-2153.

Villez, K., Srinivasan, B., Rengaswamy, R., Narasimhan, S., & Venkatasubramanian, V. (2011). Kalman-based strategies for Fault Detection and Identification (FDI): Extensions and critical evaluation for a buffer tank system. *Comp. Chem. Eng.*, 35(5), 806-816.

Villez, K., Corominas, L., & Vanrolleghem, P. A. (2013a). Sensor fault detection and diagnosis based on bilinear mass balances in wastewater treatment systems. In: *Proceedings of the 11th IWA Conference on Instrumentation, Control and Automation (ICA2013),* Narbonne, France, Sept. 18-20, 2013, appeared on USB-stick (IWA-12744).

Villez, K., Corominas, L., & Vanrolleghem, P. A. (2013b). Structural observability and redundancy classification for sensor networks in wastewater systems. In: *Proceedings of the 11th IWA Conference on Instrumentation, Control and Automation (ICA2013),* Narbonne, France, Sept. 18-20, 2013, appeared on USB-stick (IWA-12741).

Villez, K., Venkatasubramanian, V., & Rengaswamy, R. (2013c). Generalized shape constrained spline fitting for qualitative analysis of trends. *Comp. Chem. Eng.*, 58, 116-134.