KV_C038
Villez, K., Corominas, L., Vanrolleghem, P. A. (2015). Sensor placement by means of deterministic global optimization. 2nd IWA New Developments in IT & Water conference, Rotterdam, The Netherlands, February 8-10, 2015, appeared on USB-stick.

Sensor placement by means of deterministic global optimization

Kris Villez[1], Peter A. Vanrolleghem[2], Lluís Corominas[3]

[1] Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland (E-mail: *kris.villez@eawag.ch*)
[2] 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*)
[3] ICRA, Catalan Institute for Water Research, Emili Grahit 101, E- 17003 Girona, Spain (E-mail: *lcorominas@icra.cat*)

Abstract
Optimal placement of sensors is critical for the collection of valuable data in realistic environments. By means of a well-chosen sensor layout in wastewater treatment plants, one can simultaneously *(i)* suppress sensor installation and maintenance costs and *(ii)* increase the amount of obtained information. In addition, this also allows *(iii)* securing the possibility to automatically detect faulty data on the basis of mass balancing methods. This work evaluates how one can optimize the location of sensors in a wastewater treatment process. This is done so that the highest degree of structural observability (of variables) and redundancy (of measurements) is attained. Currently available results demonstrate that the optimization strategy is effective and efficient for simple wastewater treatment plant configurations. Case studies with larger-scale process network are currently executed to evaluate the method's potential for more realistic plant configurations.

Keywords
Data reconciliation, Experimental Design, Fault detection and identification, Global optimization, Graph theory, Sensor placement, Wastewater treatment

INTRODUCTION

Obtaining high quality on-line data from aquatic environments is a long-standing challenge in environmental engineering. Classic research in data quality evaluation, including fault detection and identification schemes consider produced data as a given. Indeed, data reconciliation and fault detection and identification (FDI) is usually executed after data is collected. In contrast, sensor placement prior to data collection can substantially improve the quality of collected data by placing sensors such that data reconciliation and FDI can be executed with increased statistical power and for more sensor fault types and locations. Therefore, this work evaluates how one can optimize the location of sensors in a wastewater treatment process so that the highest degree of observability and redundancy is attained.

The considered objectives for an optimal sensor layout or placement are *(i)* a minimal number of sensors, *(ii)* a maximal number of observable variables, and *(iii)* a maximal fraction of redundant sensors. The sensor layout optimization problem is a multi-objective problem since these objectives cannot be attained simultaneously. The actual trade-off between these objectives is considered unknown a priori and for this reason it is of interest to compute the Pareto front which describes all possible sensor layouts optimal in a certain sense.

To find the Pareto front, a novel optimization strategy is proposed. To this end, the pre-existing implementation of observability and redundancy labelling algorithms (Villez *et al.*, 2013b) is combined with a global deterministic optimization algorithm. Importantly, this optimization can be executed prior to the installation of any sensor, thus being an effective planning tool of sensor layouts and/or measurement campaigns.

METHODS

The presented analysis assumes that one knows *(i)* the spatial configuration of a plant, hereafter called the **plant configuration**, *(ii)* all potential streams in which one can place sensors, and *(iii)* the variables for which a sensor can be placed in each available stream. Any particular selection of sensors for all considered locations is further referred to as a **sensor layout**. In this study, optimality of a sensor layout is defined by means of three competing objectives which are *(i)* the cost associated with sensor installation and maintenance, *(ii)* the number of observable variables, and *(iii)* the fraction of the installed measurements or sensors which are redundant. Observability and redundancy are only evaluated in a structural sense, meaning that measurement uncertainty or the probabilities of faults are not accounted for. Optimizing these three competing objectives means that the Pareto front is searched for by means of placing sensors in the available sensor locations. The following paragraphs describe the studied academic example studied so far as well as the applied methodology.

Case study: A textbook wastewater treatment plant

A simple wastewater treatment plant 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 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 that effects of reactions and storage on the TSS concentrations are ignored for sensor layout evaluation and optimization. Each stream is considered to consist of two fractions, the TSS fraction and the remainder fraction, which is mainly water. The remainder fraction also includes all other components of the streams, which are not measured or considered of particular interest for estimation within this study.

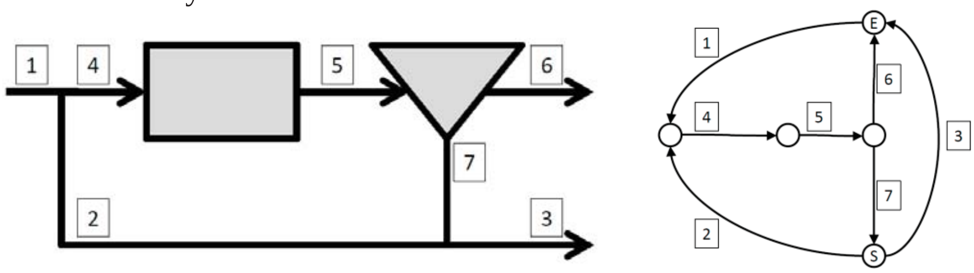

Figure 1. Scheme (left) and graph-theoretic representation (right) of a text-book wastewater treatment plant.

Definitions

Structural observability is defined as follows (e.g., Kretsovalis and Mah, 1987): *A <u>variable</u> 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.*

Importantly, structural observability merely corresponds to a feasibility of computation. Indeed, quality aspects such as precision of the measurement or estimate are not considered. This is equivalent to structural identifiability for model parameter identification (e.g., Dochain et al., 1995). We are thus merely concerned to evaluate whether a given sensor layout provides information on the variables of interest. Note that the evaluation of structural observability does not require that actual measurements are available. For this reason, structural observability is a useful objective for design of sensor layouts prior to actual installation of sensors.

Structural redundancy is defined as follows (e.g., Kretsovalis and Mah, 1987): *A <u>measurement</u> is considered structurally redundant if the measured variable remains observable when the considered measurement is removed from the set of available measurements.*

A classic interpretation of redundancy follows when one considers that one can evaluate the difference of any structurally redundant measurement with the corresponding estimate in the absence of this measurement. If the considered measurement is biased than the expected value for this difference reflects the apparent bias. As with structural observability, structural redundancy excludes quantitative measures for quality of data or estimates. Note that *reliability* (of estimates) as considered in Ali and Narasimhan (1993) is excluded as a criterion as this study focuses on structural, parameter-free objectives for sensor layouts.

Topological graph representation

In principle, any mathematical relationship, including dynamic equations, can be used for estimation and thus also for labelling of variables as observable or unobservable. In the presented methodology, only four types of equations are considered. These are *(i)* total mass flow balances, *(ii)* component mass flow balances, *(iii)* component concentration equality equations, and *(iv)* equations equating component mass flows to the product of the corresponding volumetric flow and component concentration. The mass balances and concentration equality equations *(i-iii)* are linear whereas the products in *(iv)* result in a bilinear equation. To evaluate criteria of observability and redundancy, one can represent the considered plant as a graph, G, leading to a so called graph theoretic representation of the considered plant (Deo, 2004). Fig. 1 shows the graph for the studied plant. Each physical location in the plant where two or more wastewater streams connect is represented by a vertex (indexed with i). Such vertices are visualized as a dot or circle. Each stream is represented as an arc (indexed with j) which is displayed as an arrow (directed arc). For a graph to be complete, every stream needs to connect two vertices. For this reason, every stream which connects the environment outside of the considered system boundaries is connected to a so-called environmental node, which is indicated in this work by means of an inscribed E. The total number of arcs is given as m. The number of vertices is noted as n. The resulting graph is a topological graph (as it describes relationships in a physical space). Splitters are locations in a plant where a given incoming stream is split into two or more streams without changing the concentrations (i.e., there is no reaction or separation process occurring in this location). In this paper, such splitters are visualized as circles (nodes) with an inscribed S (see Fig. 1).

Evaluation of observability and redundancy (Labelling)

Mathematical definitions

Available sensor locations are given as a set, S, consisting of pairs, (j, k), with j indicating the stream and k the considered flow rate or stream component. In this study, only flow rates ($k=1$) and TSS concentrations ($k=2$) are considered. The total number of measured variable types (flow rates and concentration of each considered stream component) is given as l. The total number of available sensor locations is given by m_S ($m_S \leq m{\cdot}l$). A vector of Boolean indicators, $\boldsymbol{\delta}_C$ (dimensions $m_S \times l$), is used to indicate whether a particular candidate location in S is equipped with a sensor (1) or not (0). An equally dimensioned matrix of Boolean indicators, $\boldsymbol{\delta}_R$ (dimensions $m_S \times l$), indicates whether the candidate location has a redundant sensor (1) or not (0, i.e., non-redundant sensor or no sensor at all). Boolean indicators for observability are grouped in a matrix, $\boldsymbol{\Delta}_O$ (dimensions $m \times l$), with $\boldsymbol{\Delta}_O(j,k)=1$ ($\boldsymbol{\Delta}_O(j,k)=0$) if the variable k for stream j is observable (non-observable). As explained further in the text, the indicators in $\boldsymbol{\Delta}_O$ and $\boldsymbol{\delta}_R$ are computed algorithmically as a function of the systems' graph representation, G, and the sensor layout specified by $\boldsymbol{\delta}_C$. For the present discussion, we represent these algorithmic procedures as two functions (f_O and f_R):

$$\boldsymbol{\Delta}_O = f_O(G, \boldsymbol{\delta}_C) \tag{1}$$

$$\boldsymbol{\delta}_R = f_R(G, \boldsymbol{\delta}_C) \tag{2}$$

The three objectives described above are formulated as follows. All objectives are constructed so that optimality is defined as the minimal values for these objectives. The sensor cost, z_C, is a weighted sum of the binary indicators for sensor presence with w_C the vector of weights for each sensor candidate location:

$$z_C = \Sigma_{s \in S}\, \mathbf{w}_C(s) \cdot \boldsymbol{\delta}_C(s) \tag{3}$$

Similarly, the observability objective is the following weighted sum of binary indicators for observability of each of the considered variables with $\mathbf{W}_O$ the weight matrix:

$$z_O = \Sigma_{j=1..m}\ \Sigma_{k=1..l}\ \boldsymbol{W}_O(j,k) \cdot (\ 1 - \boldsymbol{\Delta}_O(j,k)\ ) \quad (4)$$

By minimization of this objective one obtains a maximal degree of observability. Finally, the objective for redundancy is defined as 1 minus the ratio between the weighted sum of redundant sensors to the weighted sum of all installed sensors, with the weights, w_R, the same in the nominator and denominator:

$$z_C = 1 - (\ \Sigma_{s\in S}\ \boldsymbol{w}_R\ (s) \cdot \boldsymbol{\delta}_R(s)\ )\ /\ (\ \Sigma_{s\in S}\ \boldsymbol{w}_R\ (s) \cdot \boldsymbol{\delta}_C(s)\ ) \quad (5)$$

Once more, this formulation ensures that minimization leads to optimal (i.e., more redundant) sensor layouts. The weights $\boldsymbol{w}_C$, $\boldsymbol{W}_O$, and $\boldsymbol{w}_R$ respectively represent *(i)* the weights for the cost associated with each individual candidate sensor, *(ii)* the weights for each observability indicator, and *(iii)* the weights for each redundancy indicator. All weights are necessarily non-negative. The next section describes how the indicators in $\boldsymbol{\Delta}_O$ and $\boldsymbol{\delta}_R$ can be computed.

Algorithmic labelling of variables and sensors

The graph theoretic representation of a plant is especially suitable for further analysis in terms of observability and redundancy. Given the topological graph and a sensor layout, one can compute which variables are observable by means of the observability algorithm of Kretsovalis and Mah (1987). This algorithm is further referred to as NETOBS. Kretsovalis and Mah (1987) also provide an algorithm to label measurements or sensors as redundant. This algorithm is further referred to as the NETRED algorithm. Application of the NETOBS and NETRED algorithms implies that no reactions are taking place anywhere in the plant. Similarly, concentration equalities as implied by splitters are not taken into account either. As such, application of the NETOBS and NETRED algorithms is limited to the analysis of sensor layouts for mixing and separation process sequences only. Because of this limitation, the GENOBS and GENRED algorithms of Kretsovalis and Mah (1988a, 1988b) were developed specifically to permit labelling of variables and sensors in plants with reactions occurring and splitters present. By using the links between solvability of equations systems and graph-theoretical concepts such as biconnected components, components, cutsets, cycles, and paths, the NETOBS, NETRED, GENOBS, and GENRED algorithms provide an efficient manner to label variables as observable or non-observable and sensors as redundant or non-redundant. These algorithms are based on topological graphs as described above. All results presented in this work were obtained by application of the GENOBS and GENRED algorithms as explained in detail in Kretsovalis and Mah (1988b). Concretely, this means that Eq. 1-2 can be rewritten as follows:

$$\boldsymbol{\Delta}_O = f_O(G, \boldsymbol{\delta}_C) = GENOBS(G, \boldsymbol{\delta}_C) \quad (6)$$

$$\boldsymbol{\delta}_R = f_R(G, \boldsymbol{\delta}_C) = GENRED(G, \boldsymbol{\delta}_C) \quad (7)$$

Deterministic Pareto front optimization

Within the field of environmental engineering, stochastic algorithms such as simulated annealing and genetic algorithms are popular in the context of global optimization. These algorithms are efficient ways to obtain one or more good solutions by avoiding many local minima in an efficient manner. The disadvantage of stochastic optimization is however that no guarantee to the global optimality of a solution is given, except as a bound for the number of iterations going to infinity. Interestingly though, a class of deterministic algorithms exists which have become popular in civil engineering (e.g., logistics and transport engineering) and chemical engineering (e.g., for system design and optimization). In this case, the search algorithm leads to a guaranteed global optimum in a finite number of steps. Deterministic optimization has remained rather unpopular in environmental engineering so far. The branch-and-bound algorithm is an example. This algorithm is selected here for integer optimization (Nemhauser and Wolsey, 1988). The particular version of the branch-and-bound algorithm applied in this work is set up for multi-objective optimization (Ehrgott and Gandibleux, 2002). This is the only difference from its conventional formulation for integer programming (e.g., Nemhauser and Wolsey, 1988). Critical to the application of branch-and-bounding optimization schemes is that one can obtain provable bounds to the objective functions. In the interest of space, these bounds are not discussed in detail here. For the optimization specialist, we add here that a breadth-first approach is taken for branching and the branching variable is always the presence of the first sensor for which a placement is undecided. For an extensive introduction to deterministic global optimization for nonlinear problems we refer to Floudas (1999).

Implementation details

All computations were executed with Matlab (8.0.0.783, R2012b) on a dedicated desktop machine (CPU: Intel(R) Core(TM) i7-3770K (3.50 GHz), RAM: 8.00 GB). To this end, a structural observability and redundancy toolbox (SOAR) was created. It includes *(i)* the NETOBS and NETRED algorithms for observability and redundancy classification as reported in Kretsovalis and Mah (1987), *(ii)* the GENOBS and GENRED algorithms for observability and redundancy classification as reported in Kretsovalis and Mah (1988b) and implemented in Matlab (Villez *et al.*, 2013b), *(iii)* a pre-existing open-source graph theory toolbox by Iglin (2011), *(iv)* additional supporting code for graph analysis (e.g., identifying biconnected components) and visualization, *(v)* tools to compute the bounds for optimization, and *(vi)* specific code for this case study. The multi-objective extension of the branch-and-bound algorithm (originally implemented by Villez *et al.* (2013a) is available in the Spike_O toolbox (v1.2).

RESULTS

In what follows, the main features of the applied methodology are demonstrated. Without loss of generality, all produced results in this study are obtained as follows:

1. A single candidate sensor is available for all streams and both considered variables (flow rate and TSS concentration). Each stream can have at most one flow rate and/or one TSS concentration measurement. This leads to $m \cdot l$ potential sensor locations listed in S:
   $$S = \{ (1,1), (1,2), (2,1), \ldots (i,j), \ldots (m,l) \} \tag{8}$$
2. Observability of each flow rate and TSS concentration is of interest.
3. Redundancy of each installed sensor is of interest.
4. All weights for the sensor layout cost objective, observability objective, and redundancy objective are set to one. This means each sensor is considered to cost the same, the estimates of the considered variables are all equally valuable, and redundancy of all sensors is equally valuable. Mathematically, one writes:
   $$\forall s \in S: \mathbf{w}_C(s) = 1 \tag{9}$$
   $$\forall i \in 1 \ldots m,\ \forall k \in 1 \ldots l: \mathbf{W}_O(i,j) = 1 \tag{10}$$
   $$\forall s \in S: \mathbf{w}_R(s) = 1 \tag{11}$$

Naturally, any particular information regarding the cost of a sensor, preference for observability of particular variables, and preference for redundancy of a particular sensor can be incorporated by adjusting these weights. We proceed considering absence of any such information by means of uniform weighting. The following paragraphs showcase the results obtained so far regarding *(i)* sensor layout evaluation and *(ii)* sensor layout optimization.

Stream index (j)	Measurement $\delta_C(s)$ $s=(j,k)$		Observable? $\Delta_O(j,k)$		Redundant? $\delta_R(s)$ $s=(j,k)$	
	Flow rate (k=1)	TSS (k=2)	Flow rate (k=1)	TSS (k=2)	Flow rate (k=1)	TSS (k=2)
1	+	+	+	+	o	o
2	+	+	+	+	o	o
3			o	+		
4			+	+		
5			+	+		
6			o	o		
7			o	+		

Stream index (j)	Measurement $\delta_C(s)$ $s=(j,k)$		Observable? $\Delta_O(j,k)$		Redundant? $\delta_R(s)$ $s=(j,k)$	
	Flow rate (k=1)	TSS (k=2)	Flow rate (k=1)	TSS (k=2)	Flow rate (k=1)	TSS (k=2)
1	+	+	+	+	+	+
2	+	+	+	+	+	+
3			o	+		
4		+	+	+		+
5			+	+		
6			o	o		
7			o	+		

Stream index (j)	Measurement $\delta_C(s)$ $s=(j,k)$		Observable? $\Delta_O(j,k)$		Redundant? $\delta_R(s)$ $s=(j,k)$	
	Flow rate (k=1)	TSS (k=2)	Flow rate (k=1)	TSS (k=2)	Flow rate (k=1)	TSS (k=2)
1	+	+	+	+	o	o
2	+	+	+	+	o	o
3			+	+		
4			+	+		
5			+	+		
6	+		+	+	o	
7			+	+		

Stream index (j)	Measurement $\delta_C(s)$ $s=(j,k)$		Observable? $\Delta_O(j,k)$		Redundant? $\delta_R(s)$ $s=(j,k)$	
	Flow rate (k=1)	TSS (k=2)	Flow rate (k=1)	TSS (k=2)	Flow rate (k=1)	TSS (k=2)
1	+	+	+	+	+	+
2	+	+	+	+	+	+
3			+	+		
4			+	+		
5			+	+		
6	+		+	+	+	
7		+	+	+		+

Figure 2. Variable and sensor labelling by means of the GENOBS and GENRED algorithms for four layouts. Top-left: Layout A, Top-right: Layout B, Bottom-left: Layout C, Bottom-right: Layout D. Installed sensors are indicated by + and blue shading. Observable variables and redundant sensors are indicated by + and green shading. Non-observable variables and non-redundant sensors are indicated by o and orange shading.

Evaluation of observability and redundancy criteria

To demonstrate the use of the GENOBS and GENRED labelling algorithms, four (4) distinct sensor layouts are selected, referred to as layout A, B, C, and D. Layout A has four sensors (z_C=4): both a flow rate sensor and a TSS concentration sensor in streams 1 and 2 (influent and sludge recycle). Layout B has five sensors (z_C=5) and is the same as layout A plus one TSS concentration sensor for stream 4 (reactor entry). Layout C also has five sensors (z_C=5), and is equivalent to layout A plus one flow rate sensor in stream 6 (effluent). Layout D has six sensors (z_C=6) and is the same as layout C with an additional TSS concentration sensor in stream 7 (settler bottom).

Fig. 2 visualizes the results obtained with the GENOBS and GENRED algorithms for the considered layouts. It can be seen that layout A leads to observability of 10 variables. Four variables are left unobservable and none of the sensors is redundant ($z_O=4$, $z_R=100\%$). Layout B, with one extra sensor compared to layout A, leads to the same observability labelling ($z_O=4$). Interestingly, all installed sensors are now redundant, however ($z_R=0\%$). In contrast, layout C makes all

considered variables to be observable without any redundancy ($z_O=0$, $z_R=100\%$). With layout D, all variables are observable and all installed sensors are redundant ($z_O=0$, $z_R=0\%$).
It can be verified that the selected layouts are mutually non-dominant, i.e., none of these layouts has objective values which are simultaneously lower than those for any of the other layouts. Moreover, the selected sensor layouts are part of the Pareto front discussed below. This means that no sensor layout exists which exhibits objective values which are simultaneously better than those for any of these four exemplary sensor layouts.

Multi-objective optimization of sensor layouts

The possibility to individually decide to install any of the 14 candidate sensors leads to a total of 16384 (2^{14}) 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 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 were not evaluated, signifying a 49% reduction in explored sensor 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. 3. 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, $z_C=0$, $z_O=0$, $z_R=100\%$), *(ii)* 707 solutions with five (5) sensors, all variables observable, and without redundant sensors (green circle, cfr. layout C, $z_C=5$, $z_O=0$, $z_R=100\%$), and *(iii)* 148 solutions with six (6) sensors, all variables observable, and all sensors redundant (blue circle, cfr. layout D, $z_C=6$, $z_O=0$, $z_R=0\%$). Of the remaining solutions, *(iv)* 311 sensor layouts do not exhibit any redundant sensor (white circles, cfr. layout A, $z_R=100\%$) and *(v)* 72 sensor layouts exhibit at least one redundant sensor while some variables remain unobservable (grey circles, cfr. layout B, $0<z_O<m\cdot l$; $z_R<100\%$).

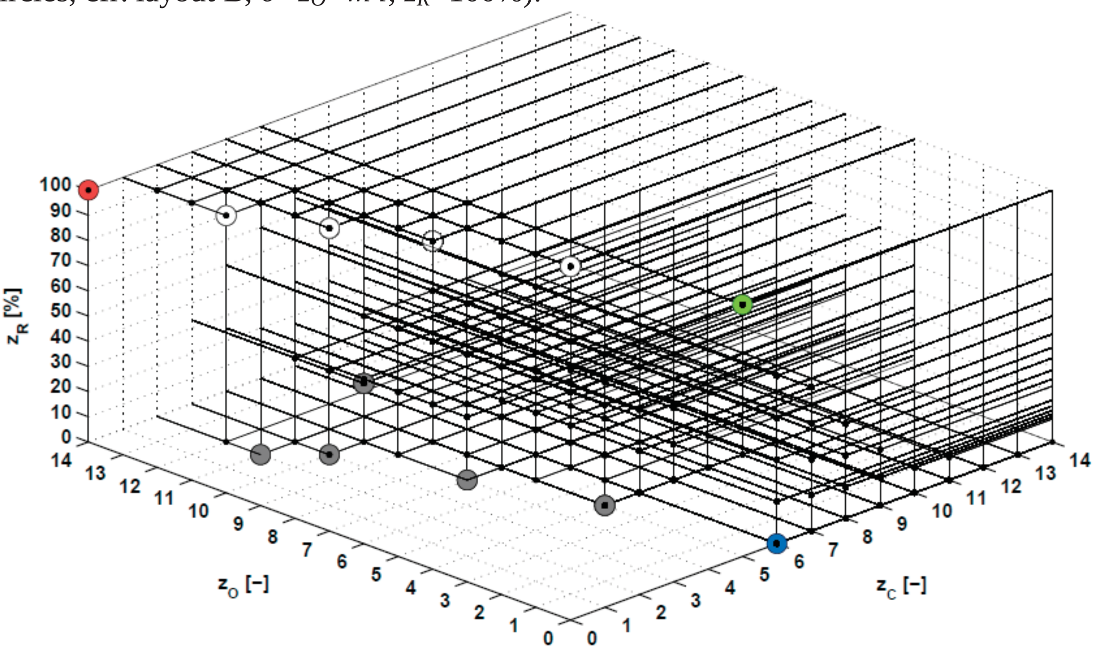

Figure 3. Visualization of the Pareto front by means of the considered objective values. Circles indicate the Pareto front sensor layouts whereas the dots indicate all evaluated sensor layouts. The red circle indicates the no-sensor layout. The green circles indicate layouts with all variables observable and no 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.

Most of the sensor layouts on the Pareto front can be found by existing greedy algorithms (e.g., Ali and Narasimhan, 1995). Indeed, by iteratively adding sensors one can track a path in the Pareto front from the red circle (no sensors) via the white circles to the green circle (all variables observable, no redundancy). A continued greedy search for redundancy will lead to the blue circle (all variables observable, all sensors redundant). The 72 sensor layouts with some redundant sensors and some variables unobservable (grey circles) cannot be obtained in this way, however. These layouts are however valuable when increasing the number of observable variables does not outweigh the value of redundancy for automated detection and identification of faulty sensor data.

CONCLUSIONS AND PERSPECTIVES

With this contribution, a deterministic global optimization strategy for sensor layout optimization has been presented. Currently available results demonstrate that multi-objective optimization of sensor placements is possible thanks to a combination of graph-theoretic labelling algorithms and branch-and-bound optimization. For the time being, the available results are limited to an academic example. Ongoing work is particularly focused on more complex and realistic wastewater treatment plant configurations, such as the Benchmark Simulation Model No. 1 (Jeppsson *et al.*, 2004).

ACKNOWLEDGEMENT

This work is possible thanks to a toolbox for graph theory application by Iglin (2011) and a toolbox for optimization developed by Villez *et al.* (2013). 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

Ali, Y. and Narasimhan, S. (1993). Sensor network design for maximizing reliability of linear processes. AIChE Journal, 39(5), 820-828.

Ali, Y. and Narasimhan, S. (1995). Redundant sensor network design for linear processes. AIChE Journal, 41(10), 2237-2249.

Deo, N. (2004). Graph theory with applications to engineering and computer science. PHI Learning Pvt. Ltd.

Dochain, D., Van Daele, M., and Vanrolleghem, P. A. (1995). Structural identifiability of biokinetic models of activated sludge respiration. Water Research, 29, 2571-2578.

Ehrgott, M. and Gandibleux, X. (2002). Multiple criteria optimization: State of the art annotated bibliographic surveys. Multiobjective combinatorial optimization – Theory, methodology, and applications, 369-444. Springer US.

Floudas, C. (1999). Deterministic Global Optimization: Theory, Algorithms and Applications. Kluwer Academic.

Iglin, S. (2011). grtheory - graph theory toolbox, January 2011, available online.

Jeppsson, U., Pons, M.-N. (2004). The COST 624 benchmark simulation model – Current state and future perspective. Editorial. Control Engineering Practice, 12, 299-304.

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.

Nemhauser G. L., and Wolsey, L. A. (1988). Integer and combinatorial optimization. Wiley.

Villez, K., Rengaswamy, R., and Venkatasubramanian, V. (2013a). Generalized shape constrained spline fitting for qualitative analysis of trends. Computers and Chemical Engineering, 58, 116-134.

Villez, K., Vanrolleghem, P. A., Corominas, Ll. (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, FR, Sept. 18-20, 2013, appeared on USB-stick (IWA-12741).