Water Research 101 (2016) 75–83

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Optimal flow sensor placement on wastewater treatment plants

Kris Villez [a, *], Peter A. Vanrolleghem [b], Lluís Corominas [c]

[a] Eawag, Department Process Engineering, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland
[b] modelEAU, Université Laval, Pavillon Adrien-Pouliot, 1065, avenue de la Médecine, Québec, G1V 0A6 Québec, Canada
[c] ICRA, Catalan Institute for Water Research, Scientific and Technological Park of the University of Girona, Emili Grahit, 101, E-17003 Girona, Spain

ARTICLE INFO

Article history:
Received 13 February 2016
Received in revised form
20 May 2016
Accepted 22 May 2016
Available online 24 May 2016

Keywords:
Fault detection
Mass balancing
Multi-objective optimization
Redundancy
Sensor placement
Wastewater treatment

ABSTRACT

Obtaining high quality data collected on wastewater treatment plants is gaining increasing attention in the wastewater engineering literature. Typical studies focus on recognition of faulty data with a given set of installed sensors on a wastewater treatment plant. Little attention is however given to how one can install sensors in such a way that fault detection and identification can be improved. In this work, we develop a method to obtain Pareto optimal sensor layouts in terms of cost, observability, and redundancy. Most importantly, the resulting method allows reducing the large set of possibilities to a minimal set of sensor layouts efficiently for any wastewater treatment plant on the basis of structural criteria only, with limited sensor information, and without prior data collection. In addition, the developed optimization scheme is fast. Practically important is that the number of sensors needed for both observability of all flows and redundancy of all flow sensors is only one more compared to the number of sensors needed for observability of all flows in the studied wastewater treatment plant configurations.

1. Introduction

Asset management of urban water infrastructures is the set of practices that utilities execute to ensure that infrastructure performance corresponds to service targets over time, that risks are adequately managed, and that the corresponding costs are as low as possible. Water asset management cannot be understood without proper data collection, either from routine off-line measurements or online data gathered from sensors and actuators. Modern small wastewater treatment plants (WWTPs) generate up to 500 signals, whereas larger ones typically register over 30,000 (Olsson et al., 2014). Despite the data-rich nature of modern WWTPs, limited useful information is typically generated, often due to improper sensor placement, installation, and/or maintenance. As a result, the resources – both in terms of time and capital – spent on installing and maintaining on-line sensors, and on collecting and storing on-line data in databases, are to a large extent lost since the collected data is not transformed into actionable knowledge for system optimization. Despite the historical recognition of the need for high quality data (Rieger et al., 2010), sub-optimal operation of a WWTP is still the norm rather than the exception. This situation is unacceptable given the need for an efficient, resilient, and sustainable use of available resources.

The focus of this paper is on sensor placement as a way to improve fault detection performance. This has been studied for safety-critical systems such as drinking water supply systems (Hart and Murray, 2010), but is not addressed for WWTPs. The placement of sensors on a WWTP typically responds to arbitrary choices, regulatory needs, or controller performance (Rehman et al., 2015). Furthermore, practically all studies focussing on data quality and fault detection in WWTPs consider the available sensor signals a given (e.g., Corominas et al., 2011; Maere et al., 2012; Puig et al., 2008; Rosén and Lennox, 2001; Spindler and Vanrolleghem, 2012; Spindler, 2014; Villez et al., 2012; Villez and Habermacher, 2016; Yoo et al., 2004, 2006a,b,c). However, sensor placement can also be conducted wisely to increase redundancy and hence, improve the overall reliability of the plant. Indeed, the placement of sensors affects *(i)* the ability to reduce random errors, i.e. improve noise reduction and estimation accuracy (*What is the best estimate?*) and *(ii)* fault detection performance (*Is there a fault?*). Our vision is therefore that prior optimal sensor placement can be conducted to increase redundancy and hence, facilitate (posterior) fault detection for WWTP measurements.

For the placement of flow sensors, methods based on structural observability and redundancy criteria are applicable (Meyer et al., 1994). Such methods only consider whether a particular variable

* Corresponding author.
E-mail address: kris.villez@eawag.ch (K. Villez).

http://dx.doi.org/10.1016/j.watres.2016.05.068
0043-1354/

can be estimated (structural observability) and whether a sensor fault can be detected (structural redundancy). Structural criteria allow evaluating a sensor layout without detailed sensor information, such as failure rates or measurement uncertainty. We therefore adopt this approach and evaluate to which extent these structural criteria alone permit sensor layout optimization in typical WWTPs.

The proposed method is inspired by earlier studies focused on drinking water supply systems (Hart and Murray, 2010) and complex plants (Meyer et al., 1994). As in earlier works we make use of graph-based methods for evaluation of observability and redundancy criteria (Kretsovalis and Mah, 1987, 1988a,b; Ponzoni et al., 1999, 2004). In contrast to earlier sensor placement studies, the sensor placement problem *(i)* is posed as a multi-objective optimization problem, *(ii)* is solved to guaranteed global optimality, and *(iii)* leads to an enumeration of all Pareto-optimal sensor layouts. In addition, cost, observability, and redundancy are optimized simultaneously and optimal sensor layouts for typical WWTP configurations are discussed for the first time.

2. Materials and methods

2.1. Studied systems

The three studied WWTP configurations are shown in Fig. 1a/c/e and are *(i)* a simple organics removing WWTP (Tchobanoglous et al., 2003) (WWTP1), *(ii)* a WWTP for nitrogen removal (WWTP2, the so-called modified Ludzack Ettinger, MLE) which is also a popular benchmark WWTP (Gernaey et al., 2014), and *(iii)* a WWTP for biological nitrogen and phosphorus removal, namely the modified University of Cape Town (MUCT) system (Tchobanoglous et al., 2003) (WWTP3). In all cases, each flow is considered a feasible candidate for sensor placement.

2.2. Sensor placement as a multi-objective optimization problem

2.2.1. Problem description

The sensor placement problem consists of finding sensor layouts which are optimal in terms of cost, observability, and redundancy. A variable is *structurally observable* when *(i)* a measurement of this variable is available or *(ii)* a unique value for the variable can be computed by means of a set of balance equations and other measurements. A sensor is considered *structurally redundant* if the measured variable remains observable when the considered sensor is removed. For optimization purposes, a sensor layout is described by means of binary *decision variables*, x_j, which represent the absence (0) or presence (1) of a sensor for flow j. Given values for all flows, $\boldsymbol{x}$, one can compute which variables are observable and which of the placed sensors are redundant. Practically, y_j indicates whether flow j is observable (1) or not (0). Similarly, z_j indicates whether flow j is equipped with a redundant sensor (1) or not (0). The set of feasible sensor layouts consists of all considered vectors $\boldsymbol{x}$ and is further referred to as the root set. All symbols used in what follows are given in Table 1.

For the purpose of optimization, the cost, observability, and redundancy objectives are:

$$f_C(\boldsymbol{x}) = \sum_j w_{x,j} \cdot x_j = \boldsymbol{w}_x^T \cdot \boldsymbol{x} \tag{1}$$

$$f_O(\boldsymbol{x}) = \sum_j w_{y,j} - \sum_j w_{y,j} \cdot y_j = \boldsymbol{w}_y^T \cdot (\mathbf{1} - \boldsymbol{y}) \tag{2}$$

$$f_R(\boldsymbol{x}) = \sum_j w_{z,j} - \sum_j w_{z,j} \cdot z_j = \boldsymbol{w}_z^T \cdot (\mathbf{1} - \boldsymbol{z}) \tag{3}$$

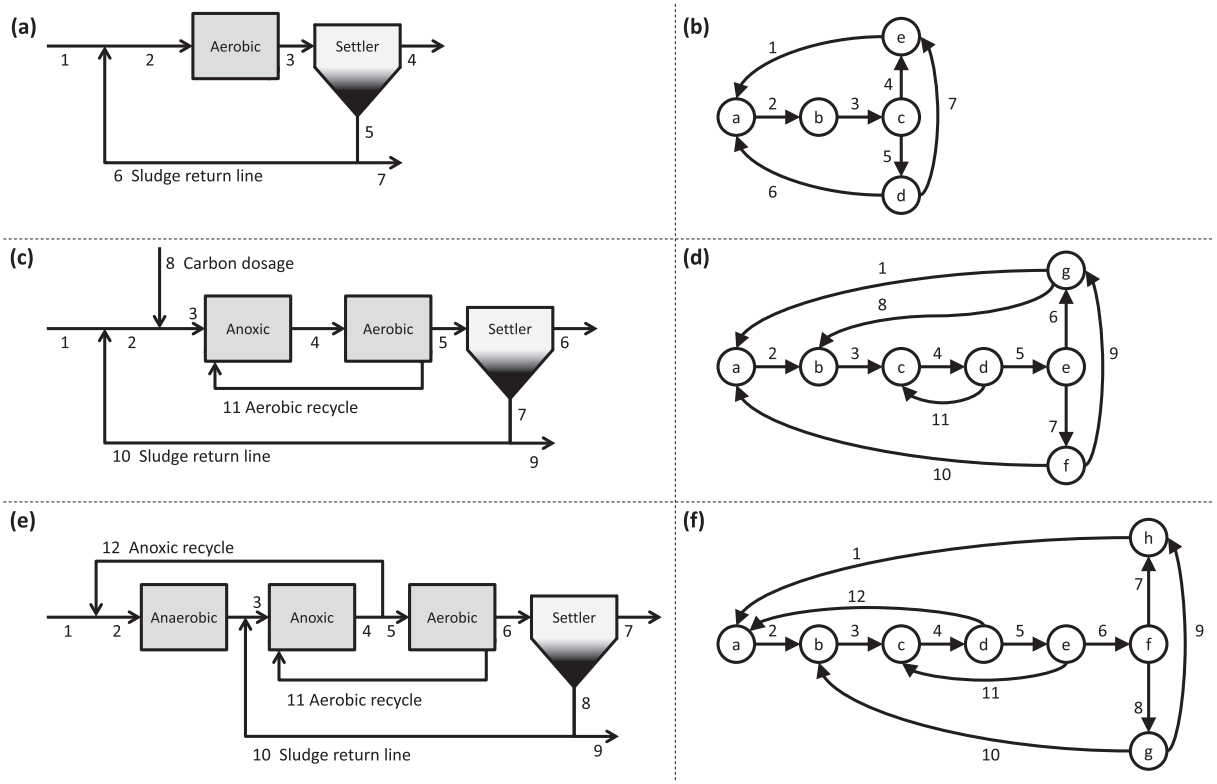

Fig. 1. Schemes and graph representations of the studied plant layouts. WWTP1: *(a)* scheme, *(b)* graph. WWTP2 (MLE): *(c)* scheme, *(d)* graph. WWTP3 (MUCT): *(e)* scheme, *(f)* graph.

Table 1
List of mathematical symbols.

Symbol	Description
j	Flow index
$f_C(\boldsymbol{x})$	Cost objective
$f_O(\boldsymbol{x})$	Observability objective
$f_R(\boldsymbol{x})$	Redundancy objective
n_j	Number of flows
$\boldsymbol{w}_o(j)$, $w_{o,j}$	Weight for flow j
$\boldsymbol{x}(j)$, x_j	Presence of a sensor for flow j
$\boldsymbol{y}(j)$, y_j	Observability of flow j
$\boldsymbol{z}(j)$, z_j	Presence of a redundant sensor in flow j
$\underline{o_{\mathscr{X}}}$, $\overline{o_{\mathscr{X}}}$	Lower/upper bound over the set $\mathscr{X}$
$\mathscr{X}$	Set of vectors x

with $w_{x,j}$, $w_{y,j}$, and $w_{z,j}$ non-negative weights. The optimization problem consists of finding vectors $\boldsymbol{x}$ which minimize $f_C(\boldsymbol{x})$, $f_O(\boldsymbol{x})$, and $f_R(\boldsymbol{x})$. Without loss of generality, all results are obtained with the weights set equal to one:

$$\forall j \in \{1, 2, \ldots, n_j\} : \{w_x(j) = 1 \wedge w_y(j) = 1 \wedge w_z(j) = 1\} \tag{4}$$

2.2.2. Nature of the problem

The optimization problem described above is an integer program. Despite this complexity, the following rules are easily verified:

$$\forall a, b : \boldsymbol{x}_a \leq \boldsymbol{x}_b \Rightarrow f_C(\boldsymbol{x}_a) \leq f_C(\boldsymbol{x}_b) \tag{5}$$

$$\forall a, b : \boldsymbol{x}_a \leq \boldsymbol{x}_b \Rightarrow f_O(\boldsymbol{x}_a) \geq f_O(\boldsymbol{x}_b) \tag{6}$$

$$\forall a, b : \boldsymbol{x}_a \leq \boldsymbol{x}_b \Rightarrow f_R(\boldsymbol{x}_a) \geq f_R(\boldsymbol{x}_b) \tag{7}$$

Equations (5)–(7) are used below to prove the lower bounds.

2.3. Labeling observable variables and redundant sensors

To evaluate $\boldsymbol{y}$ and $\boldsymbol{z}$, graph-based labeling procedures are used. The graphs (Deo, 2004) for the studied WWTPs are given in Fig. 1b/d/f. Edges correspond to flows in the WWTP. *Cycles* and *components* are critical concepts for our method. A cycle is defined here as a non-empty combination of edges in a graph which enables to connect a given node to itself by traversing one or more edges of the graph in any direction and without repetition. A component or connected component is a subgraph (Deo, 2004) within which each pair of nodes can be connected by traversing the edges of the graph.

Labeling is executed as follows. First, label all edges with measurements as observable and remove them from the graph. In the reduced graph, label the edges which are not on a cycle as observable. Label all other edges as unobservable. For redundancy labeling, identify all components of the reduced graph. For each of the removed edges, evaluate whether the start and end node are in different components and label the measurement as redundant when so. All remaining measurements are non-redundant. For proofs we refer to the existing literature (Stanley and Mah, 1981; Kretsovalis and Mah, 1988a,b).

2.4. Optimization

Optimization is executed by multi-objective branch-and-bound optimization in a breadth-first manner (Nemhauser and Wolsey, 1988; Ehrgott and Gandibleux, 2002). This is a deterministic algorithm for nonlinear global optimization as explained in the Supplementary Information (S.I.). The necessary bounds are given in the next section. The specific Matlab-based software needed to reproduce our results is published under the GPL v3 license and is added in the S.I. All computations were executed with Matlab R2014b (64-bit, win64) on a Lenovo ThinkPad X240 (Applied CPU: 1.40 GHz, Available RAM: 4.00 GB).

2.5. Bounds

At every branch-and-bound iteration, one evaluates the bounds to the objectives for a set of sensor layouts, $\mathscr{X}$. Each set consists of layouts for which some, none, or all values x_j are the same. For a remaining set of flows, the value for x_j can be 0 or 1. Consider two extremal sensor layouts within the set. The first of these layouts, $\underline{\boldsymbol{x}_{\mathscr{X}}}$, is the solution corresponding to placing sensors only in the locations for which every sensor layout within the set includes a sensor:

$$\underline{\boldsymbol{x}_{\mathscr{X}}}(j) = \begin{cases} 1 & \text{if } \forall \boldsymbol{x} \in \mathscr{X} : \boldsymbol{x}(j) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The second extremal sensor layout, $\overline{\boldsymbol{x}_{\mathscr{X}}}$, is obtained by placing a sensor in every location for which at least one sensor layout in the set includes a sensor:

$$\overline{\boldsymbol{x}_{\mathscr{X}}}(j) = \begin{cases} 1 & \text{if } \exists \boldsymbol{x} \in \mathscr{X} : \boldsymbol{x}(j) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

2.5.1. Upper bounds

In this study, upper bounds are obtained for the $\overline{\boldsymbol{x}_{\mathscr{X}}}$ solution:

$$\bar{f}_C = f_C\left(\overline{\boldsymbol{x}_{\mathscr{X}}}\right) = \boldsymbol{w}_x^T \cdot \overline{\boldsymbol{x}_{\mathscr{X}}} \tag{10}$$

$$\bar{f}_O = f_O\left(\overline{\boldsymbol{x}_{\mathscr{X}}}\right) = \boldsymbol{w}_y^T \cdot \left(\mathbf{1} - \overline{\boldsymbol{y}_{\mathscr{X}}}\right) \tag{11}$$

$$\bar{f}_R = f_R\left(\overline{\boldsymbol{x}_{\mathscr{X}}}\right) = \boldsymbol{w}_z^T \cdot \left(\mathbf{1} - \overline{\boldsymbol{z}_{\mathscr{X}}}\right) \tag{12}$$

The values for $\overline{\boldsymbol{y}_{\mathscr{X}}}$ and $\overline{\boldsymbol{z}_{\mathscr{X}}}$ are determined with the provided labeling procedures. These are valid upper bounds since the computed objective values are simultaneously attained with a single solution ($\overline{\boldsymbol{x}_{\mathscr{X}}}$) in the considered set.

2.5.2. Lower bounds

The lower bounds are obtained as follows. The lowest value of the cost within a set is obtained for the $\underline{\boldsymbol{x}_{\mathscr{X}}}$ solution. Based on Eq. (5) one obtains:

$$\forall \boldsymbol{x} \in \mathscr{X} : \underline{\boldsymbol{x}_{\mathscr{X}}} \leq \boldsymbol{x} \Rightarrow \underline{f_C} = f_C\left(\underline{\boldsymbol{x}_{\mathscr{X}}}\right) \leq f_C(\boldsymbol{x}) \tag{13}$$

The lowest value of the observability function within a set is obtained for the $\overline{\boldsymbol{x}_{\mathscr{X}}}$ solution. This follows from Eq. (6):

$$\forall \boldsymbol{x} \in \mathscr{X} : \left\{ \overline{\boldsymbol{x}_{\mathscr{X}}} \geq \boldsymbol{x} \Rightarrow \overline{\boldsymbol{y}_{\mathscr{X}}} \geq \boldsymbol{y} \Rightarrow \underline{f_O} = f_O\left(\overline{\boldsymbol{x}_{\mathscr{X}}}\right) \leq f_O(\boldsymbol{x}) \right\} \tag{14}$$

The lowest value of the redundancy objective within a set is obtained for the $\overline{\boldsymbol{x}_{\mathscr{X}}}$ solution. This follows from Eq. (7):

$$\forall \boldsymbol{x} \in \mathscr{X} : \left\{ \overline{\boldsymbol{x}_{\mathscr{X}}} \geq \boldsymbol{x} \Rightarrow \overline{\boldsymbol{z}_{\mathscr{X}}} \geq \boldsymbol{z} \Rightarrow \underline{f_R} = f_R\left(\overline{\boldsymbol{x}_{\mathscr{X}}}\right) \leq f_R(\boldsymbol{x}) \right\} \quad (15)$$

3. Results

The basic methodological concepts are demonstrated first by means of WWTP1. Afterwards, the results for WWTP2 and WWTP3 are discussed.

3.1. WWTP1

3.1.1. Sensor layout examples

Each of the seven flows in Fig. 1(a) is given as an edge in Fig. 1(b). The nodes *a*, *b*, *c*, and *d* correspond to physical locations. A fifth node, *e*, is the *environmental node* which represents the system boundary. This makes it a mathematically consistent graph, i.e. there are no loose ends.

To demonstrate the labeling procedures, consider that F1 and F6 are measured (layout A). Flows F1 and F6 are then removed for observability labeling (Fig. 2, top-left). Edges not lying on any cycle correspond to observable flows (F2, F3). All other flows are on cycles and thus unobservable (F4, F5, F7). The resulting graph consists of a single component so that all sensors are non-redundant. The situation changes when a flow sensor is added for the excess sludge flow (F7, layout B). Now there are no cycles (Fig. 2, top-right). Therefore, all flows are observable. Once more, a single-component graph results and the sensors remain non-redundant. Another sensor is now placed at the exit of the reactor (F3, layout C). All flows remain observable (Fig. 2, bottom-left). Two graph components result after removal of measured flows. The start and end nodes of edges 1, 3, and 6 correspond to different components, so F1, F3, and F6 are redundant. This is because these flow measurements are tied through a balance equation. In contrast, edge 7 has start and end nodes on the same component of the graph. Equivalently, no balance equation without unmeasured flows and including the F7 measurement exists. Therefore, F7 is non-redundant. Adding another sensor for the WWTP effluent (F4, layout D) makes all sensors redundant (Fig. 2, bottom-right). These results are summarized in Table 2.

Table 2
WWTP1. Observability and redundancy labeling for example layouts. Note that absence of a sensor ($x_j = 0$) logically implies the absence of a redundant sensor ($z_j = 0$).

Flow	Layout A			Layout B			Layout C			Layout D		
j	x	y	z	x	y	z	x	y	z	x	y	z
1	1	1	0	1	1	0	1	1	1	1	1	1
2	0	1	0	0	1	0	0	1	0	0	1	0
3	0	1	0	0	1	0	1	1	1	1	1	1
4	0	0	0	0	1	0	0	1	0	1	1	1
5	0	0	0	0	1	0	0	1	0	0	1	0
6	1	1	0	1	1	0	1	1	1	1	1	1
7	0	0	0	1	1	0	1	1	0	1	1	1

3.1.2. Pareto-optimal sensor layouts

Deterministic optimization is now used to evaluate the Pareto optimal sensor layouts. The complete set of feasible solutions consists of $2^7 = 128$ solutions. Ten distinct Pareto-optimal combinations for the objectives are shown in Fig. 3 with bubble sizes reflecting the number of layouts for each combination. These combinations and all individual solutions are listed in S.I. (Tables S1 and S2). Layouts A, B, and D discussed above are part of the Pareto front. Layout C is not part of this Pareto front. A total of 70 layouts lie on the Pareto front. Of these layouts, 37 layouts exhibit no

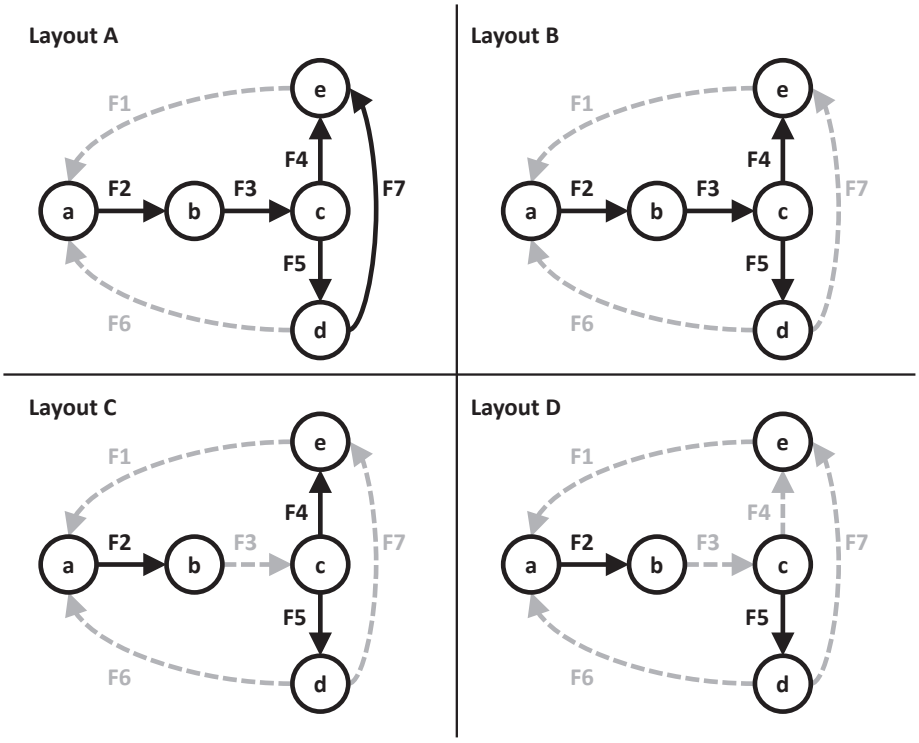

Fig. 2. WWTP1. Modified graphs for four sensor layouts. Removed edges are dashed and grey. Layout A: observable: F2, F3, redundant: none. Layout B: observable: all, redundant: none. Layout C: observable: all, redundant: F1, F3, F6. Layout D: observable: all, redundant: all installed sensors.

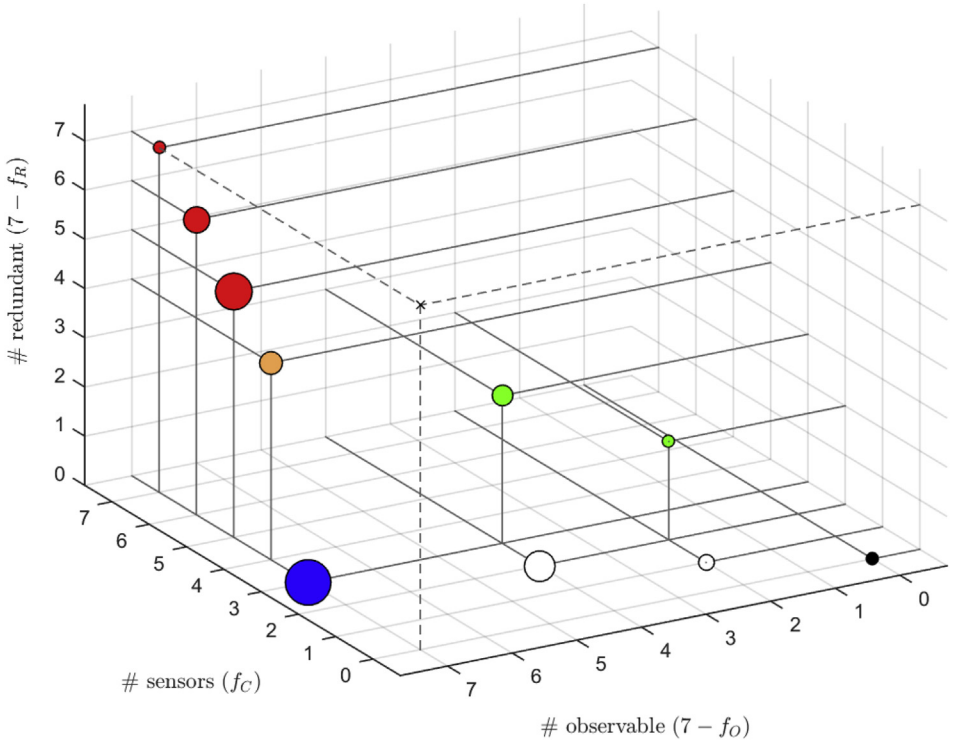

Fig. 3. WWTP1. Visualization of the Pareto front. The size of the bubbles reflects the number of layouts in each point on the Pareto front. The ideal solution is indicated with dashed lines. Black: trivial sensor layout without any sensor. Blue: 23 layouts, observable: all, redundant: none. Green: five layouts, observable: some, redundant: some. Orange: five sensor layouts, cost: minimal, observable: all, redundant: all. Red: 33 layouts, observable: all, redundant: all. White: 13 layouts, observable: some, redundant: none. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

redundant sensors (layouts 1–37), 38 layouts render all variables observable and all sensors redundant (layouts 43–70), and five layouts exhibit some redundant sensors while not all variables are observable (layouts 38–42). Note that these 5 layouts cannot be obtained through conventional optimization methods since these methods (e.g., Ali and Narasimhan, 1998) add sensors first until all variables are observable and proceed to increase the number of redundant sensors only after this is achieved. The Pareto front contains the layout without any sensors as a trivial solution (layout 1) and 23 layouts with three non-redundant sensors making all variables observable (layout 14–37). Four sensors are sufficient to make all variables observable and all sensors redundant. Five such layouts exist (layout 43–47).

Table 3
WWTP2 (LME). Observability and redundancy labeling for example layouts. Note that absence of a sensor ($x_j = 0$) logically implies the absence of a redundant sensor ($z_j = 0$).

Flow	Layout A			Layout B			Layout C			Layout D		
j	x	y	z	x	y	z	x	y	z	x	y	z
1	0	1	0	0	1	0	0	1	0	1	1	1
2	0	1	0	0	1	0	0	1	0	0	1	0
3	1	1	0	1	1	0	1	1	1	0	1	0
4	0	1	0	0	1	0	0	1	0	1	1	1
5	0	1	0	0	1	0	0	1	0	0	1	0
6	0	0	0	0	1	0	1	1	1	0	0	0
7	0	0	0	0	1	0	0	1	0	0	0	0
8	1	1	0	1	1	0	1	1	1	1	1	1
9	0	0	0	1	1	0	1	1	1	0	0	0
10	1	1	0	1	1	0	1	1	1	1	1	1
11	1	1	0	1	1	0	1	1	1	1	1	1

3.2. WWTP2 (MLE)

3.2.1. Sensor layout examples

Fig. 1(d) shows the graph representation of WWTP2. In Table 3, the results for a few layouts on the Pareto front are displayed in detail. Layout A is a layout with measurements for F3, F8, F10, and F11. In this case, eight flows are observable and none of the sensors is redundant. F6, F7, and F9 remain unobservable. Adding a sensor for F9 makes all flow rates observable (layout B). None of the sensors is redundant. Adding another sensor in F6 makes all placed sensors redundant (layout C). Layout D is a layout where all sensors are redundant but not all variables observable. Sensors are placed in F1, F4, F8, F10, and F11. F6, F7, and F9 remain unobservable as is the case for layout A.

A typical plant may already have sensors for the influent flow (F1) and flows equipped with pumps (F8-11). This makes all flows observable based on the labeling procedures discussed above. This is typical and means that all flows are observable in most WWTPs. In this case, at least two more sensors (seven sensors in total) are necessary to make all sensors redundant if the already installed sensors are kept in their positions. However, layout C has only six sensors while all flows are observable and all sensors are redundant. There are no minimal sensor layouts with all sensors redundant which include sensors for F1 and F8-11 simultaneously. The Pareto optimal set of sensor layouts thus depends on previously installed sensors. This can be taken into account easily by setting the corresponding cost weights ($w_{x,j}$) to zero. This is however not studied in detail.

3.2.2. Pareto-optimal sensor layouts

In the WWTP2 (MLE) case, there are eleven distinct flows for which a sensor can be placed. This leads to a total of $2^{11} = 2048$

feasible sensor layouts. The resulting Pareto front consists of 609 layouts in 15 unique objective combinations (Fig. 4). All combinations and solutions are given in the S.I. (Tables S3 and S4). Six combinations represent the 383 solutions without any redundant sensors (layout 1–383). Five sensors are sufficient to make all flow rates observable. There exist 209 such layouts (layout 175–383). With one to four sensors, one obtains at most one, three, six, and eight observable flow rates (layouts 2–174). Placing six sensors is sufficient to make all flow rates observable while also making all sensors redundant. Interestingly, there exist only seven such layouts (layout 402–408). Beyond six installed sensors, each additional sensor placement will increase the cost and the number of redundant sensors with one unit (201 layouts, layout 409–609). There are 18 sensor layouts for which not all flow rates are observable while all installed sensors are redundant (layout 384–401). The number of observable variables is three, six, and eight for three, four, and five installed sensors.

Table 4
WWTP3 (MUCT). Observability and redundancy labeling for example layouts. Note that absence of a sensor ($x_j = 0$) logically implies the absence of a redundant sensor ($z_j = 0$).

Flow	Layout A			Layout B			Layout C			Layout D		
j	x	y	z	x	y	z	x	y	z	x	y	z
1	0	0	0	0	1	0	0	1	0	1	1	1
2	0	0	0	0	1	0	0	1	0	0	1	0
3	0	0	0	0	1	0	0	1	0	0	1	0
4	0	0	0	1	1	0	1	1	1	1	1	1
5	0	0	0	0	1	0	0	1	0	0	1	0
6	0	0	0	0	1	0	0	1	0	0	1	0
7	0	0	0	0	1	0	1	1	1	0	0	0
8	0	1	0	0	1	0	0	1	0	0	0	0
9	1	1	0	1	1	0	1	1	1	0	0	0
10	1	1	0	1	1	0	1	1	1	1	1	1
11	1	1	0	1	1	0	1	1	1	1	1	1
12	1	1	0	1	1	0	1	1	1	1	1	1

3.3. WWTP3 (MUCT)

3.3.1. Sensor layout examples

Fig. 1(f) shows the graph representation of WWTP3. Table 4 lists four layouts on the Pareto front. The first (layout A) consists of four sensors placed in the excess sludge flow and the recycle flows (F9-12). Five flows are observable and there is no redundancy. Adding a sensor for F4 (anoxic reactor outlet) makes all variables observable (layout B). Adding another sensor for F7 (effluent) makes all sensors redundant (layout C). Layout D involves a flow sensor in the influent (F1), the anoxic reactor outlet (F4), and the recycle flows (F10-12). Nine flow rates are observable and all sensors are redundant. Note that adding a sensor for F4 might be unrealistic for many plants as F4 typically exists between two tank zones separated with a baffle. To account for this, one can set the corresponding value for x to zero in the root solution set or one can set the corresponding cost weight ($w_{x,j}$) to infinity. The effect of such changes is however not studied in detail.

3.3.2. Pareto-optimal sensor layouts

WWTP3 exhibits twelve distinct flows leading to $2^{12} = 4048$ feasible layouts. The Pareto front in Fig. 5 contains 1154 of these. Sixteen unique points exist on this front and are described in detail in S.I. (Tables S5 and S6). Six of the points represent 577 solutions without redundancy (layout 1–577). These include the trivial solution without any sensors (layout 1) and 336 layouts with five sensors which make all flows observable (layout 242–577). Adding sensors delivers layouts with additional redundant sensors (layouts

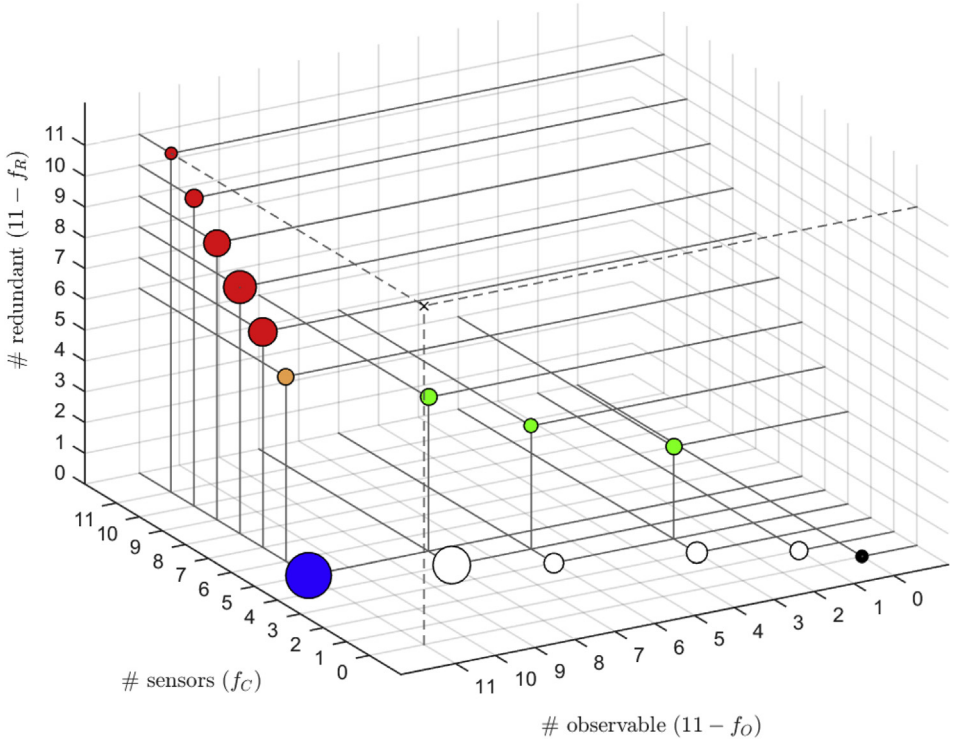

Fig. 4. WWTP2 (MLE). Visualization of the Pareto front. The size of the bubbles reflects the number of layouts in each point on the Pareto front. The ideal solution is indicated with dashed lines. Black: trivial sensor layout without any sensor. Blue: 209 layouts, observable: all, redundant: none. Green: 18 layouts, observable: some, redundant: some. Orange: seven sensor layouts, cost: minimal, observable: all, redundant: all. Red: 201 layouts, observable: all, redundant: all. White: 173 layouts, observable: some, redundant: none. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

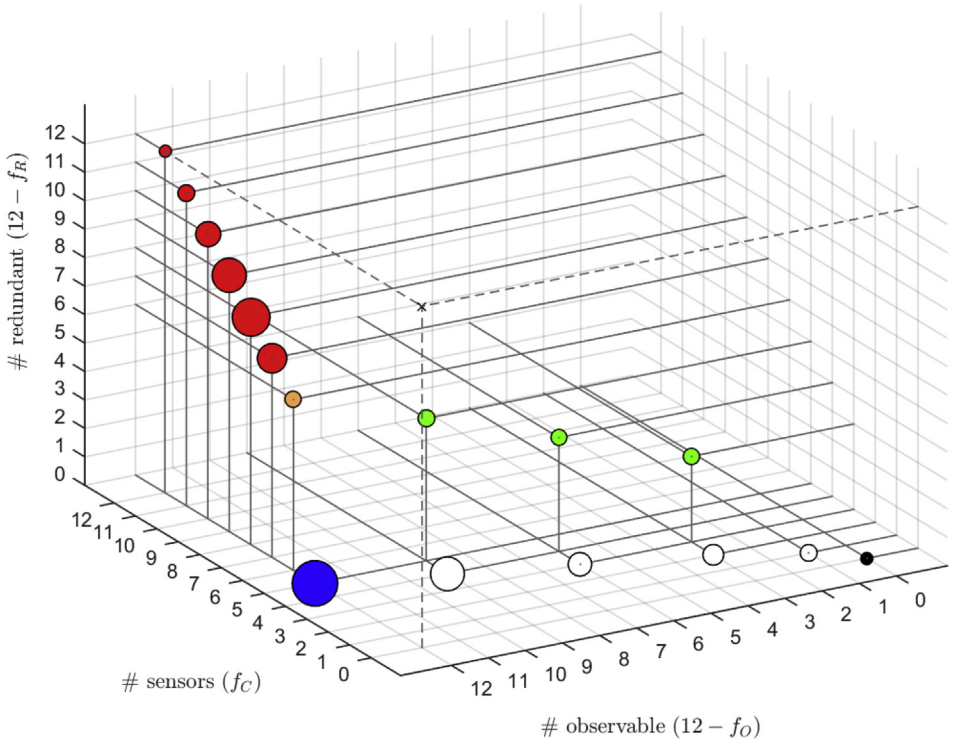

Fig. 5. WWTP3 (MUCT). Visualization of the Pareto front. The size of the bubbles reflects the number of layouts in each point on the Pareto front. The ideal solution is indicated with dashed lines. Black: trivial sensor layout without any sensor. Blue: 336 layouts, observable: all, redundant: none. Green: 31 layouts, observable: some, redundant: all. Orange: ten sensor layouts, cost: minimal, observable: all, redundant: all. Red: 536 layouts, observable: all, redundant: all, White: 240 layouts, observable: some, redundant: none. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

609–1154). Interestingly, only ten layouts exist with six sensors for which all variables are observable and all sensors are redundant. The Pareto front also contains 31 layouts for which all sensors are redundant while not all flow rates are observable (layout 578–608).

4. Discussion

In what follows, the main benefits of the applied optimization method as demonstrated above are given, followed by more general observations regarding the solved sensor placement problem and a few considerations regarding future work.

4.1. Benefits of the optimization method

In this study, Pareto optimal flow sensor layouts are found by means of a deterministic optimization scheme combined with graph-based evaluation of structural observability and redundancy. These are the most important benefits of this method:

- Using structural objectives for optimization means that sensor layout optimization *(i)* can be executed for any plant as well as at the design stage of a new plant, *(ii)* does not require information regarding the performance of candidate sensors, and *(iii)* does not require collecting measurements prior to optimization. This is the most important benefit of the method.
- The provided Pareto front solutions are guaranteed to be globally optimal thanks to bounds on the objectives developed specifically for this work. At the same time, the time needed to compute the Pareto front was less than 20 min for all studied cases. This indicates that automated screening of candidate sensor layouts on the basis of structural criteria is possible with conventionally available computational resources.
- All Pareto-optimal sensor layouts can be found by the proposed method, including sensor layouts with redundant sensors that do not make all flows observable. This is not guaranteed by conventional sequential optimization of observability and redundancy criteria. This of particular interest when only a few flows should be estimated with high accuracy and reliability while other flows are not of interest at all.

4.2. General observations

Based on the three case studies, the following observations can be made:

- Although the number of Pareto-optimal layouts tends to be large (e.g., 1154 for MUCT), the fraction of optimal layouts among all feasible layouts is reduced considerably (28.2% and 29.7% for the realistic MLE and MUCT plants). Most likely, the resulting Pareto front is most valuable to assist the end-user with further elucidation of sensor layout preferences.
- Increasing the complexity of a WWTP configuration, particularly by increasing the number of flows, leads to a smaller fraction of feasible sensor layouts lying on the Pareto front. This is thought to be due to an increased presence of recycle flows.
- For the MLE and MUCT plants, installing five sensors is sufficient to achieve observability for all flows without any redundant sensors while installing six sensors is sufficient to achieve the same observability with this time with all sensors redundant. This shows that the additional investments to move from complete observability to complete observability and redundancy are small in many WWTPs.

- All flows were considered feasible sensor locations. However, some of the computed solutions require that a sensor is placed at the inlet of a reactor tank or at the inlet of the settler, which is not always feasible. Our method however allows accounting for this by setting the corresponding weights ($w_{x,j}$) to infinity or by redefining the root set so it does not contain infeasible sensor layouts prior to optimization. Similarly, already installed sensors or legally required instruments can be included automatically by redefining the root set so it contains only sensor layouts which include these sensors. Importantly, the physical configuration of a WWTP and existing sensor placements can restrict the possibilities for sensor layout optimization. Plant design and sensor layout optimization is thus best considered simultaneously.

4.3. Outlook

The following enhancements are considered for future work:

- The inclusion of isolability criteria is considered valuable if one aims not only to facilitate data reconciliation (*What is the best estimate?*) and fault detection (*Is there a fault?*) but also to automatically execute fault isolation (*Which sensor signal(s) cause the detection?*), and *(iv)* improve diagnosis performance (*What is the root cause?*). Unfortunately, there is no algorithm yet for isolability labeling based on topological graphs as used in this work. Note however that an alternative approach based on a bipartite graph is available (Raghuraj et al., 1999).
- The addition of practical criteria for observability (Waldraff et al., 1998; Chmielewski et al., 2002) or redundancy (Ali and Narasimhan, 1998) will enable further refinement of the Pareto front. This however requires that detailed information about the candidate sensors is available.

5. Conclusions

In this study, a systematic search for optimal sensor layouts is solved as a deterministic multi-objective optimization problem involving cost, observability and redundancy objectives. The general applicability of the developed method is demonstrated by means of three popular wastewater treatment plant configurations. Our results indicate that the additional investments to move from complete observability to complete observability and redundancy requires only one additional flow sensor in the studied configurations. Finally and most importantly, the applied optimization method is fast and is applicable for any plant configuration without prior data collection or specification of sensor characteristics.

Acknowledgements

The authors thank Sergii Iglin for his graph theory toolbox and Juan Pablo Carbajal, Eli Duenisch, Pascal Getreuer, Timothy E. Holy, and Will Robertson for code facilitating visualisation and reporting of our results. Funding for this study was provided by Funding for this study was provided by Peter Vanrolleghem's Discovery Grant awarded by NSERC (Natural Sciences and Engineering Research Council of Canada). Peter Vanrolleghem holds the Canada Research Chair on Water Quality Modelling. The authors acknowledge the Ramon and Cajal grant RYC-2013-14595 and the Marie Curie Career Integration Grant PCIG9-GA-2011-293535. ICRA is recognized as consolidated research group by the Catalan Government with code 2014-SGR-291.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2016.05.068.

References

Ali, Y., Narasimhan, S., 1998. Sensor network design for maximizing reliability of linear processes. AIChE J. 39 (5), 820–828.
Chmielewski, D.J., Palmer, T., Manousiouthakis, V., 2002. On the theory of optimal sensor placement. AIChE J. 48, 1001–1012.
Corominas, L., Villez, K., Aguado, D., Rieger, L., Rosén, C., Vanrolleghem, P.A., 2011. Performance evaluation of fault detection methods for wastewater treatment processes. Biotechnol. Bioeng. 108 (2), 333–344.
Deo, N., 2004. Graph Theory with Applications to Engineering and Computer Science. PHI Learning Pvt. Ltd.
Ehrgott, M., Gandibleux, X., 2002. Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys. Springer US, Ch. Multiobjective combinatorial optimization – theory, methodology, and applications, pp. 369–444.
Gernaey, K.V., Jeppsson, U., Vanrolleghem, P.A., Copp, J.B. (Eds.), 2014. Benchmarking of Control Strategies for Wastewater Treatment Plants. IWA Publishing.
Hart, W.E., Murray, R., 2010. Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. J. Water Resour. Plan. Manag. 136 (6), 611–619.
Kretsovalis, A., Mah, R.S.H., 1987. Observability and redundancy classification in multicomponent process networks. AIChE J. 33 (1), 70–82.
Kretsovalis, A., Mah, R.S.H., 1988a. Observability and redundancy classification in generalized process networks – I. Theorems. Comput. Chem. Eng. 12 (7), 671–687.
Kretsovalis, A., Mah, R.S.H., 1988b. Observability and redundancy classification in generalized process networks – II. Algorithms. Comput. Chem. Eng. 12 (7), 689–703.
Maere, T., Villez, K., Marsili-Libelli, S., Naessens, W., Nopens, I., 2012. Membrane bioreactor fouling behaviour assessment through principal component analysis and fuzzy clustering. Water Res. 46 (18), 6132–6142.
Meyer, M., Le Lann, J.M., Koehret, B., Enjalbert, M., 1994. Optimal selection of sensor location on a complex plant, using a graph oriented approach. Comput. Chem. Eng. 18, S535–S540.
Nemhauser, G.L., Wolsey, L.A., 1988. Integer and Combinatorial Optimization. Wiley, New York.
Olsson, G., Carlsson, B., Comas, J., Copp, J., Gernaey, K., Ingildsen, P., Jeppsson, U., Kim, C., Rieger, L., Rodríguez-Roda, I., Steyer, J.-P., Takács, I., Vanrolleghem, P.A., Vargas, A., Yuan, Z., Åmand, L., 2014. Instrumentation, control and automation in wastewater – from London 1973 to Narbonne 2013. Water Sci. Technol. 69 (7), 1373–1385.
Ponzoni, I., Sánchez, M.C., Brignole, N.B., 1999. A new structural algorithm for observability classification. Industrial Eng. Chem. Res. 38, 3027–3035.
Ponzoni, I., Sánchez, M.C., Brignole, N.B., 2004. Direct method for structural observability analysis. Industrial Eng. Chem. Res. 43, 577–588.
Puig, S., Van Loosdrecht, M.C.M., Colprim, J., Meijer, S.C.F., 2008. Data evaluation of full-scale wastewater treatment plants by mass balance. Water Res. 42 (18), 4645–4655.
Raghuraj, R., Bhushan, M., Rengaswamy, R., 1999. Locating sensors in complex chemical plants based on fault diagnostic observability criteria. AIChE J. 45 (2), 310–322.
Rehman, U., Vesvikar, M., Maere, T., Guo, L., Vanrolleghem, P.A., Nopens, I., 2015. Effect of sensor location on controller performance in a wastewater treatment plant. Water Sci. Technol. 71 (5), 700–708.
Rieger, L., Takács, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P.A., Comeau, Y., 2010. Data reconciliation for wastewater treatment plant simulation studies – planning for high-quality data and typical sources of errors. Water Environ. Res. 82, 426–433.
Rosén, C., Lennox, J., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. Water Res. 35 (14), 3402–3410.
Spindler, A., 2014. Structural redundancy of data from wastewater treatment systems. determination of individual balance equations. Water Res. 57, 193–201.
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.
Stanley, G.M., Mah, R.S.H., 1981. Observability and redundancy classification in process networks: theorems and algorithms. Chem. Eng. Sci. 36 (12), 1941–1954.
Tchobanoglous, G., Burton, F., Stensel, H. (Eds.), 2003. Wastewater Engineering: Treatment, Disposal and Reuse. McGraw-Hill, New York.
Villez, K., Habermacher, J., 2016. Shape anomaly detection for process monitoring of a sequencing batch reactor. Comput. Chem. Eng. Accept. http://dx.doi.org/10.1016/j.compchemeng.2016.04.012 (in press).
Villez, K., Rieger, L., Keser, B., Venkatasubramanian, V., 2012. Probabilistic qualitative analysis for fault detection and identification of an on-line phosphate analyzer. Int. J. Adv. Eng. Sci. Appl. Math. 4, 67–77.
Waldraff, W., Dochain, D., Bourrel, S., Magnus, A., 1998. On the use of observability measures for sensor location in tubular reactor. J. Process Control 8 (5), 497–505.

Yoo, C.K., Villez, K., Lee, I.-B., Rosén, C., Vanrolleghem, P.A., 2006a. Multi-model statistical process monitoring and diagnosis of a sequencing batch reactor. Biotechnol. Bioeng. 96 (4), 687–701.
Yoo, C.K., Villez, K., Lee, I.-B., Vanrolleghem, P.A., 2006b. Multivariate nonlinear statistical process control of a sequencing batch reactor. J. Chem. Eng. Jpn. 1, 43–51.
Yoo, C.K., Lee, D.S., Vanrolleghem, P.A., 2004. Application of multiway ICA for online process monitoring of a sequencing batch reactor. Water Res. 38 (7), 1715–1732.
Yoo, C.K., Villez, K., Lee, I.B., Van Hulle, S., Vanrolleghem, P.A., 2006c. Sensor validation and reconciliation for a partial nitrification process. Water Sci. Technol. 53 (4–5), 513–521.