



A general-purpose method for Pareto optimal placement of flow rate and concentration sensors in networked systems – With application to wastewater treatment plants

Kris Villez^{a,b,*}, Peter A. Vanrolleghem^c, Lluís Corominas^{d,e}

^aEawag, Department Process Engineering, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland

^bORNL: Oak Ridge National Laboratory, Oak Ridge, TN, USA

^cmodelEAU, Université Laval, Pavillon Adrien-Pouliot, 1065, avenue de la Médecine, Québec, G1V 0A6 Québec, Canada

^dICRA, Catalan Institute for Water Research, Emili Grahit, 101, E-17003 Girona, Spain

^eUniversitat de Girona, Girona, Spain

ARTICLE INFO

Article history:

Received 27 December 2019

Revised 14 April 2020

Accepted 20 April 2020

Available online 6 May 2020

Keywords:

Bilinear balance equations

Fault detection

Multi-objective optimization

Optimal experimental design

Redundancy

Sensor placement

ABSTRACT

The advent of affordable computing, low-cost sensor hardware, and high-speed and reliable communications have spurred ubiquitous installation of sensors in complex engineered systems. However, ensuring reliable data quality remains a challenge. Exploitation of redundancy among sensor signals can help improving the precision of measured variables, detecting the presence of gross errors, and identifying faulty sensors. The cost of sensor ownership, maintenance efforts in particular, can still be cost-prohibitive however. Maximizing the ability to assess and control data quality while minimizing the cost of ownership thus requires a careful sensor placement. To solve this challenge, we develop a generally applicable method to solve the multi-objective sensor placement problem in systems governed by linear and bilinear balance equations. Importantly, the method computes all Pareto-optimal sensor layouts with conventional computational resources and requires no information about the expected sensor quality.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, the advent of large-scale cost-effective computing and low-cost sensor hardware has stimulated great interest in the deployment of large-scale sensor networks. By combining large-dimensional data sets with computationally efficient yet flexible and predictive models, it is said, one can obtain system designs and process operations that are much more efficient than those that are rooted in domain expertise, heuristics, and mechanistic models (Corominas et al., 2018; International Water Association, 2019; Venkatasubramanian, 2019). Executing this kind of vision is however hampered by the fact sensors often produce signals of questionable quality due to exposure to a harsh environment (e.g., Ohmura et al., 2019). In turn, this leads to high sensor maintenance cost, to the point that this cost may subdue any benefit of ubiquitous sensor networks.

Poor data quality is known to challenge the use of both machine learning models and mechanistic process models (Hauduc et al., 2009; Rieger et al., 2005, 2006; Rosén et al., 2008). It is therefore no surprise that several studies provide methods and insights into the problem of data quality assessment and control in the water sector (Alferes et al., 2013; Corominas et al., 2011; Newhart et al., 2019; Rosén and Olsson, 1998; Spindler, 2014; Spindler and Vanrolleghem, 2012; Thomann, 2008; Thomann et al., 2002; Villez et al., 2008). Importantly, identification of many models requires that the available data provide a complete and precise picture of all phenomena that are relevant for the task at hand. For this, data sets ought to be representative, voluminous, and of high quality.

It is generally valuable to place redundant sensors for the purpose of computer-aided anomaly and fault detection, isolation, and diagnosis. At the same time, installing large numbers of sensors implies an operational cost of sensor ownership, associated with sensor cleaning, calibration, validation, and part replacements. We believe this cost will remain high for the foreseeable future, despite the advent of low-cost hardware and tremendous advances in the robustness of instruments measuring physical properties

* Corresponding author at: ORNL: Oak Ridge National Laboratory, Oak Ridge, TN, USA.

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

(e.g., flow rate, level, pressure, temperature) or water quality (e.g., ammonia, nitrite, nitrate, ortho-phosphate, total suspended solids). Thus, it is key to balance the costs of sensors against information-richness and redundancy.

Optimal sensor placement is a complex problem due to the presence of nonlinear relationships between the variables of interest and the networked nature of many systems. Moreover, to place sensors that measure extensive variables, like flow rates, as well as intensive variables, like concentrations and temperatures, special care must be taken to account for bilinear mass balance equations describing the effect of reactions, storage phenomena, and flow splitters. Several authors have addressed this optimization problem. E.g., Ali and Narasimhan (1996) provide an iterative approach to obtain an optimal sensor layout in terms of observability, followed by a second search for additional sensors to increase the number of redundant sensors (see below for definitions). In a similar spirit, Le et al. (2018) search for Pareto optimal sensor layouts that trade off the number of sensors against the relative improvement in precision, i.e. the ratio by which the standard deviation of the estimation errors is reduced, thanks to installation of redundant sensors. In the process, the search is limited to the set of sensor layouts that make a predefined set of variables of interest observable. The relative improvement of precision is one way among many to evaluate practical observability (e.g., Waldraff et al., 1998; Chmielewski et al., 2002; Chang et al., 2012; Serpas et al., 2013; Mukherjee et al., 2017; Rico-Ramirez et al., 2007; Joshi and Boyd, 2008; Soldevila et al., 2018; de Winter et al., 2019; Nahar et al., 2019). The approach of Le et al. (2018) is based on the analysis of systems of linear equations, meaning that measuring a concentration in a single stream implies measurement of the flow rate in the same stream. Because this is based on estimation precision, prior knowledge regarding the measurement error standard deviation of the candidate sensors is required. The first author of this study argues in Le (2019) that a better approach consists of placing individual flow rate and concentration sensors. This however implies that bilinear constraint equations must be dealt with carefully during the evaluation of observability. Le (2019) evaluates his own symbolic implementation of the method for identification of redundant relationships described in Spindler (2014) to be very time-consuming.

Villez et al. (2016) is a rare study where cost, information richness, and redundancy are considered as independent objectives for optimization. To this end, the objective for information-richness quantifies the number of variables that are structurally observable (definition below). Similarly, the redundancy objective is based on structural redundancy of the installed sensors. This means that no prior information regarding the precision of any sensor candidate is required. Unfortunately, this work is limited to the placement of flow rate sensors only, mainly to avoid accounting of bilinear equations during sensor layout optimization. We address this in this work by adopting the GENOBS and GENRED procedures of Kretsovalis and Mah (1988b) for evaluation of structural observability and structural redundancy in systems with arbitrary placement of sensors measuring flow rates, concentrations, and temperature sensors. We apply this to two typical wastewater treatment plant (WWTP) configurations for which information about hydraulic flow rates, concentrations and mass flows of wastewater contaminants is crucial for plant operation. As a result, and for the first time, it is possible to compute the complete set of Pareto optimal sensor layouts in engineered systems described by linear and bilinear balance equations. Special care has been taken to obtain fast implementations of the GENOBS and GENRED procedures as well as the optimization method itself. As a result, the developed method is both efficient and generally applicable to any system described by linear and bilinear balance equations.

2. Materials and methods

2.1. Studied systems

Sensor placement results are shown for two WWTP configurations: (i) a simple organics removing WWTP (Metcalf & Eddy, 2003) (WWTP1, Fig. 1a) and (ii) a WWTP for nitrogen removal (WWTP2, the so-called modified Ludzack Ettinger process, MLE, Fig. 1e). Both configurations consist of a set of tanks, including reactors and a settler, which are connected by closed pipes. All pipes carry wastewater, including a multitude of contaminants, further referred to as the components. The most important conversion processes are (i) growth and decay of biomass, which affects the concentration of particulate matter (solids), (ii) oxidation of organic matter, (iii) oxidation of ammonia (into nitrite and nitrate), (iv) and oxidation of nitrite (into nitrate). We consider the plants to have aerobic reactors only so that the concentration of ortho-phosphate is unaffected by any conversion process, i.e. ortho-phosphate can be treated as a component that is both conserved and soluble.

2.2. Sensor placement as a multi-objective optimization problem

2.2.1. Problem description

The sensor placement problem consists of finding sensor layouts that are optimal in terms of cost, observability, and redundancy for a system of material flows described by linear hydraulic balances (total mass flows), bilinear component balances (e.g., solids mass flow), and splitter constraints (i.e., to express that concentrations in front and at the back of a splitter box are equal). We apply the definitions of *structural observability* and *structural redundancy* as in prior work (Villez et al., 2016). 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 equations and other measurements. A sensor is considered structurally redundant if the measured variable remains observable when the considered sensor is removed.

The total number of components is given as C and the total number of physical streams, i.e. pipes, is given as S . The S streams are indexed with an integer s ($s = 1, \dots, S$). We consider five types of variables associated with every stream: the (total mass) flow rate (unit: $[m^3/d]$), concentrations of C components (unit: $[g/m^3]$), C component mass flow rates corresponding to the concentrations (unit: $[mol/d]$), temperature (unit: $[K]$), and heat flow (unit: $[J/d]$). Within each stream we index the considered variables with v , where $v = 0$ is the index of the (total mass) flow rate, $v = 1, 2, \dots, C$ are the indices of the concentrations of the C components (unit: $[mol/m^3]$), $v = C + 1$ is the index of temperature, $v = C + 2, C + 3, \dots, 2 \cdot C + 1$ index the component mass flow rates (in the same order as the concentration variables), and $v = 2 \cdot C + 2$ refers to the heat flow (unit: $[J/d]$). By default, we consider that a sensor can be placed for any the considered variables in any physical stream.

For optimization purposes, a sensor layout is described by means of binary *decision variables*, $X_{s,v}$, which represent the absence (0) or presence (1) of a sensor for variable v in stream s . The matrix \mathbf{X} is an $S \times V$ -dimensional matrix ($V = 3 + 2 \cdot C$) containing the values for all decision variables ($X_{s,v} = \mathbf{X}(s, v)$). Given such a matrix, one can compute which variables are observable and which of the placed sensors are redundant. The matrix \mathbf{Y} is an $S \times V$ -dimensional matrix with elements $Y_{s,v} = \mathbf{Y}(s, v)$, where $Y_{s,v}$ indicates whether variable v in stream s is structurally observable (1) or not (0). Similarly, the $S \times V$ -dimensional matrix \mathbf{Z} contains elements $Z_{s,v} = \mathbf{Z}(s, v)$ indicating whether stream s is equipped with a structurally redundant sensor for variable v (1) or not (0). The complete set of feasible sensor layouts consists of all feasible ma-

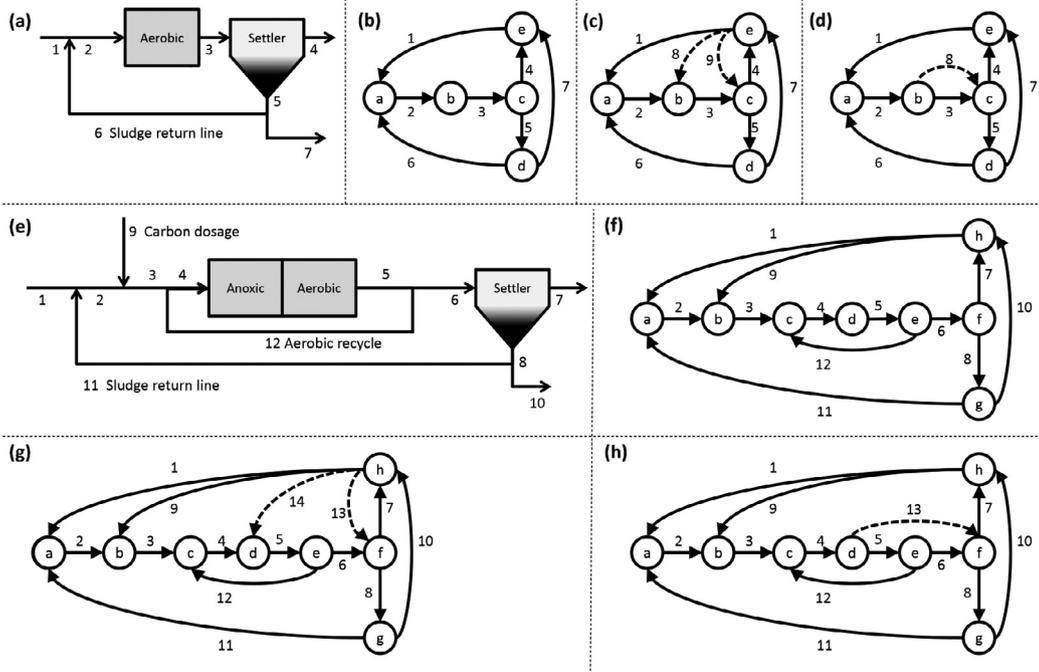


Fig. 1. Schemes and graph representations of the studied plant configurations. WWTP1: (a) scheme, (b/c/d) graph. WWTP2 (MLE): (e) scheme, (f/g/h) graphs. In the graphs, physical streams are depicted with full lines while reaction and storage phenomena are shown as dashed lines.

Table 1

List of mathematical symbols.

Symbol	Description
C	Number of components
S, S^*, \bar{S}	Number of streams (physical/imaginary/total)
s	Stream index
$f^C(\mathbf{X})$	Cost objective
$f^O(\mathbf{Y})$	Observability objective
$f^R(\mathbf{Z})$	Redundancy objective
V	Number of variables
v	Variable index
$W_{s,v}, \mathbf{W}(s,v)$	Weight for variable v in stream s
$X_{s,v}, \mathbf{X}(s,v)$	Presence of a sensor for variable v in stream s
$Y_{s,v}, \mathbf{Y}(s,v)$	Observability of variable v in stream s
$Z_{s,v}, \mathbf{Z}(s,v)$	Presence of a redundant sensor for variable v in stream s

trices \mathbf{X} and is named the root set. All remaining symbols are described in Table 1.

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

$$f^C(\mathbf{X}) = \sum_s \sum_v W_{s,v}^C \cdot X_{s,v} \quad (1)$$

$$f^O(\mathbf{Y}) = \sum_s \sum_v W_{s,v}^O (1 - Y_{s,v}) \quad (2)$$

$$f^R(\mathbf{Z}) = \sum_s \sum_v W_{s,v}^R (1 - Z_{s,v}) \quad (3)$$

with $W_{s,v}^C$, $W_{s,v}^O$, and $W_{s,v}^R$ weights. These weights are typically non-negative. However, negative weights can be used in principle, e.g. to express that removing an existing sensor comes with a non-negligible cost. The optimization problem consists of finding matrices \mathbf{X} which minimize f^C , f^O , and f^R .

2.3. Labeling observable variables and redundant sensors

2.3.1. Accounting for physical streams

To evaluate \mathbf{Y} and \mathbf{Z} , the graph-based labeling procedures GENOBS and GENRED provided by Kretsovalis and Mah (1988b) are adopted to produce the matrices \mathbf{Y} and \mathbf{Z} given any plant graph and any sensor layout \mathbf{X} . The graphs (Deo, 2004) for the studied WWTPs are given in Fig. 1b/c/d/f/g/h. These graphs contain at most two types of edges. The edges of the first type are shown as full lines and correspond to the physical streams of the studied system. These are the S streams mentioned above. Every node in the graph represents a total mass flow balance with the in- and outgoing edges representing in- and outgoing mass flows. For example, node a in Fig. 1b, corresponds to the following hydraulic balance:

$$0 = q_1 + q_6 - q_2 \quad (4)$$

where q_1 , q_2 , and q_6 the mass flow rates in stream 1, 2, and 6. The corresponding edges of the graph are called a cutset, as their removal cuts the graph in separate graphs. This means one can enumerate all possible hydraulic balances by enumerating the cutsets of the graph. For this, efficient algorithms exist (Deo, 2004). The maximal number of independent hydraulic balances among these balances is always one less than the number of nodes. One possible set of independent hydraulic balances can be found by computing the fundamental cutsets of the graph. An advantage is that this provides a precise definition for the concept of overlapping balances, which has only been defined loosely so far (see e.g., Rieger et al., 2010; Spindler, 2014; Le et al., 2018), namely: any two cutsets with a shared edge represent pair of overlapping balances.

2.3.2. Accounting for reactions

Edges of the second type are optional, are shown with dashed lines, and correspond to imaginary streams in the studied system. Each dashed line in Fig. 1c represents an edge between a physical node (node b and c) and the environment (node e). As discussed in Kretsovalis and Mah (1988a), this enables accounting for deficits in individual component balances due to the effect of reactions (Kretsovalis and Mah, 1988a; 1988b). Stream 8 is directed

from the environmental node e to the reactor node b . This provides a graphical representation of the fact that the balance equation for the j th component around node b is written as:

$$0 = q_2 \cdot c_{2,j} - q_3 \cdot c_{3,j} + \rho_8 \cdot v_{8,j} \quad (5)$$

with q_2 and q_3 the mass flow rates in stream 2 and 3, $c_{2,j}$ and $c_{3,j}$ the corresponding concentrations, ρ_8 the reaction rate (unit: $\text{mol}/\text{m}^3 \cdot \text{d}$), and $v_{8,j}$ the stoichiometric coefficient for the considered component (unit: $[-]$). The same holds for stream 9, which represents effects of reactions in the settler. We assume, as in Kretsovalis and Mah (1988b), that all stoichiometric coefficients are known.

2.3.3. Accounting for transport phenomena

A modified approach is taken in the graph Fig. 1d. In this case, an imaginary stream is added from node b to c . This particular modification allows to account for deficits in the balances around the reactor basin and the settler due to temporary storage of the considered component (e.g., solids) in the reactor or settler. In this case, reactions in the reactor or the settler are not accounted for. As an example, the balance equation around node b now has the following structure:

$$0 = q_2 \cdot c_{2,j} - q_3 \cdot c_{3,j} - M_{8,j} = q_2 \cdot c_{2,j} - q_3 \cdot c_{3,j} - \rho_8 \cdot u_{8,j} \quad (6)$$

where q_2 , q_3 , c_2 , c_3 have the same meaning as above and $r_{8,j}$ represents the net transport rate of component j (unit: $[\text{mol}/\text{d}]$). To match the balance equation structure above, we define $M_8 := r_8 \cdot v_8$ with v_8 a constant. $u_{8,j}$ indicates whether the component participates in the transport phenomenon. It is equal to one if the component is subject to the considered storage phenomenon (e.g., component j is a solid) and zero otherwise (e.g., component j is soluble). We refer to ρ_8 as the transport rate. In the present case, a positive (negative) transport rate means the considered component is accumulating in the reactor (settler). This is represented by the minus sign in the term $-M_8$ and as an edge going out from node b in the graph. More terms of the same form can be added when multiple components are stored and released with distinct rates, thus leading to the addition of multiple imaginary edges to the graph. This approach is different compared to existing work (Kretsovalis and Mah, 1988a; 1988b) in the sense that the added imaginary streams connect nodes that represent physical locations, in contrast to edges representing reactions. Importantly, this modified method implies that the total mass of every component in the system remains constant, i.e. we assume that there are no plant-wide component balance deficits.

As a result, Fig. 1b represents the situation for a conserved component (e.g., ortho-phosphate), Fig. 1c represents the situation for a reactive component (e.g., ammonia, nitrite, nitrate), and Fig. 1d represents the case of a component whose total mass in the plant is conserved (e.g., total solids). Similar graphs for WWTP2 (MLE) are shown in the same order in Fig. 1f/g/h. The total number of imaginary edges is given as S^* . The edges are indexed so that the first S edges among the complete list of $\bar{S} := S + S^*$ edges correspond to the physical streams. Similar to the cutsets discussed above, one can compute the cutsets of the complete graph, covering both physical and imaginary streams. By doing so, one can enumerate all possible component flow balances for any single component, regardless of whether the component is conserved, subject to conversions, or subject to storage phenomena. As before, pairs of cutsets with shared edges represent pairs of overlapping balances.

2.3.4. Implementation

While the original Structural Observability And Redundancy (SOAR) toolbox (Villez et al., 2016) could in principle be used for this work also, initial results indicated that its computational efficiency was insufficient to solve the multi-objective sensor layout

optimization problems within acceptable time windows (days to weeks). For this reason, a new version of the SOAR toolbox (v2.0) has been created. The most important change is that the sets of (a) all cycles of the complete graph, (b) all cutsets of the complete graph, and (c) all cutsets of the graph without imaginary edges are computed in advance and used as inputs to the GENOBS and GENRED procedures. This enables an efficient search for the Pareto front. The exact implementation is described in detail in Villez (2019). All code necessary to reproduce our results is added in the *Supplementary Information*.

2.4. Optimization

2.4.1. Concept of surplus redundancy

In any system equipped with a given sensor layout, one can add redundant sensors by installing sensors for variables that are already observable. In many cases, such an additional sensor will only increase the cost by one and increase the number of redundant sensors by one. When this is the case, we consider the resulting sensor layout to have surplus redundancy, i.e. one could remove at least one sensor without changing the number of observable variables or the number of non-redundant sensors. Surplus redundancy can bring a multitude of benefits. E.g., additional redundant sensors can improve the precision of reconciled measurements (e.g., Crowe, 1989; Le et al., 2018) or improve the ability to isolate sensor faults (e.g., Gertler and Singer, 1990). In the present study, we exploit surplus redundancy to obtain a more efficient multi-objective optimization strategy, as explained below.

2.4.2. Using surplus redundancy during multi-objective optimization

The complete Pareto front is computed with a two-step method. In step A, we solve a modified version of the multi-objective optimization problem by replacing the redundancy objective function with f^N , which computes the fraction of the objective f^R that is associated with only the non-redundant sensors among the installed sensors.

$$f^N(\mathbf{X}, \mathbf{Z}) = f^R(\mathbf{X} - \mathbf{Z}) \quad (7)$$

This new objective function computes a penalty for installing non-redundant sensors. The modified optimization problem then trades this penalty off against the cost and observability objectives. This problem is solved with the deterministic multi-objective branch-and-bound optimization method (Nemhauser and Wolsey, 1988; Ehrgott and Gandibleux, 2002; Villez et al., 2016).

In step B, we consider the original objective functions f^C , f^O , f^R . The complete Pareto front, now including sensor layouts with surplus redundancy, is obtained by executing the following steps for each layout on the existing Pareto front:

1. Take the computed matrices for the considered layout and name these \mathbf{X}_0 , \mathbf{Y}_0 , and \mathbf{Z}_0 . Compute the objective functions f^C , f^O , f^R and add this solution to the list of solutions.
2. List the K additional candidate sensors that have not been placed but for which the corresponding variable is already observable. Any of these additional sensors, when placed, will be redundant automatically.
3. Create $N = 2^K - 1$ additional sensor layouts by placement of every non-empty subset of the additional candidate sensors. Call the number of added sensors P .
4. For each of the N layouts:
 - (a) Check if the considered layout is already in the Pareto front. If not, continue.
 - (b) Compute \mathbf{X}_1 by changing every 0 in \mathbf{X}_0 in the location of the added sensors P to 1 to reflect the added sensor placement. \mathbf{Z}_1 is computed in the same fashion starting with \mathbf{Z}_0 .

- (c) Set $\mathbf{Y}_1 = \mathbf{Y}_0$. Indeed, no variables becomes observable because sensors are only placed for variables that are already observable.
- (d) Compute the objective functions f^C , f^O , f^R and add the obtained solution to the list of solutions.
5. Prune any solution that is not Pareto optimal in the original objective functions f^C , f^O , f^R from the list of solutions.

We prove that this two-step method effectively computes the Pareto front for the objectives f^C , f^O , f^R . To this end, we refer to the Pareto front obtained by direct optimization of f^C , f^O , and f^R as PFO. The Pareto front obtained in step A described above consists of all solutions that exhibit Pareto optimality for f^C , f^O , and f^N and is named PF1. Now consider any Pareto optimal layout in PFO and call it layout L0. If this layout has no surplus redundancy (Case 1), it means that one cannot remove any sensor from this layout without decreasing the value of f^O or increasing the value of f^N . This layout is therefore also part of PF1 and is added to the final set of sensor layouts in Step B.1 above. If this layout has surplus redundancy (Case 2), it is possible to remove a redundant sensor from this layout in such a way that the sensor layout is guaranteed to belong also to PFO. Indeed, one can remove a sensor among the redundant sensors for which the proportion of the increase of f^R to the decrease of f^C is minimal. In addition, removing any redundant sensor cannot change the value of f^O , given that all variables that were observable remain observable. One can continue this process of sensor removal to produce sensor layouts that are guaranteed to be in PFO until one has reached a sensor layout without surplus redundancy. At this point, one has obtained a sensor layout in PFO that is necessarily part of PF1 also. Call this layout L1. Conversely, this means that layout L0 can be produced by adding sensors to layout L1 for variables that observable already. Since Step B.2 until B.3 enumerate every possible sensor layout L0 by adding sensors to every layout L1 in PF1, it follows that the complete Pareto front PFO is produced by combining Step A and B as described above.

2.5. Illustration of the methods

2.5.1. Restrictions

The studied methods are general and enable placement regardless of (a) the number of components and (b) the chosen weight matrices (W^C , W^O , W^R). Without losing generality, we illustrate the methods with the following restrictions:

- We assume there are 3 or more components present in the system and that these components are present in every stream ($C \geq 3$).
- The cost weights ($W_{s,v}^C$) are 0 (sensor available at no cost), 1 (sensor available at a cost), or $+\infty$ (sensor not feasible).
- Sensors for direct measurement of component mass flow rates and heat flow rates are not available.
- We place sensors for two variables only: total mass flow rates and concentrations of one component (e.g., ortho-phosphate or total suspended solids, but not both). Practically, all cost weights are equal to infinity, except for $v = 0$ and $v = 1$.
- Unless mentioned otherwise, we assume that every stream can be equipped with a flow rate sensor, a concentration sensor, or both:

$$W_{s,v}^C = \begin{cases} 1, & \text{if: } s \in \{1, \dots, S\} \text{ and } v \in \{0, 1\} \\ +\infty, & \text{otherwise} \end{cases} \quad (8)$$

- Unless mentioned otherwise, we assume a uniform interest in estimates for all variables (all potential estimates are equally valuable), and assume that it is equally desirable for any sensor to be redundant:

$$W_{s,v}^O = W_{s,v}^R = 1 \quad (9)$$

Table 2
WWTP2. Applied weights for the observability objective function.

Stream (s)	Total mass flow rate	Component concentration	Component mass flow rate
1	1	1	1
2	0	0	0
3	0	0	0
4	0	0	1
5	0	0	0
6	1	1	0
7	1	1	1
8	0	0	1
9	1	1	1
10	1	1	0
11	1	0	0
12	1	0	0

2.5.2. Studied problems

We study a total of nine optimization problems. The first four optimization problems are based on the WWTP1 system (Fig. 1a) while the remaining five are based on the WWTP2 system (Fig. 1e). Among each of these sets of optimization problems, the first problem consists of placement of flow rate sensors only (problem 1 and 5). This problem is selected to compare the performance of the methods proposed in this work with the methods used in Villez et al. (2016) and is therefore solved with the original toolbox (SOAR v1.2) and the newest version (SOAR v2.0). Optimization problem 2, 3, and 4 concern placement of sensors for flow rates and one component in WWTP1. These correspond to the graphs in Fig. 1b/c/d and thus reflect a case with a conserved component, a reactive component, and a conserved solid component. Optimization problem 6 to 8 are the equivalent problems for WWTP2. In this case, the concentration in stream 9 (carbon dosage stream) is considered to be known. Optimization problem 9 consists of a modification of problem 8 by changing the weights for the observability objective. These weights are listed in Table 2 and reflect our judgment regarding the relevance of flow rates, concentrations, and component flow rates for sludge management (solids) in a typical WWTP. For example, only the component mass flow rates in stream 1 (influent), 6 (settler loading), 7 (effluent), 9 (carbon dosage), 10 (sludge removal), 11 (return sludge), and 12 (internal recycle) are of interest. In addition, we consider the flow rates for stream 9 to 12 known (carbon dosage, internal recycle, sludge recycle, sludge waste). This is based on the fact that these streams are often equipped with a pump, so that these mass flow rates or their setpoints can be known in principle. By doing so, we illustrate the generality of our method.

2.6. Implementation

All computations were executed with Matlab 9.6.0.1072779 (R2019a) on a desktop machine (System: Intel® Core™ i7-4770 CPU 3.40GHz, RAM: 16GB; OS: Microsoft Windows 10 Education, 64-bit). The Supplementary Information includes all software to reproduce our results, including the SOAR software toolbox developed for this purpose, and is published under the GPL v3 license.

3. Results

The WWTP1 system mainly serves a didactic purpose. For this reason, we refer the reader new to sensor placement, graph theory, or multi-objective optimization to Section S.2 of the supplementary information, where results with WWTP1 are described in detail. In the next paragraphs, we discuss on results obtained for WWTP2 (MLE), as the scale of this system matches realistic cases. At the end of the results section, we discuss computational re-

requirements. Note that all computed Pareto fronts are listed in tabular form in the *Supplementary Data*.

3.1. WWTP2 (MLE)

We discuss results for optimization problems 6 to 9 to illustrate the applicability of our approach in diverse systems at realistic scale. In Problem 6 to 8, mass flow rate sensors and component concentration measurements can be placed in every physical stream except stream 9. In stream 9, the component concentration is considered known and only a mass flow rate sensor can be placed. This means a total of 23 sensors can be placed, leading to a total of 8,388,608 candidate sensor layouts. Problem 9 is used to stress the generality of our method by demonstrating that one can easily modify the objective functions to account for a reduced set of sensor candidates, presence of already installed sensors, and prioritization of certain variables of interest over others.

3.1.1. Problem 6: WWTP2 (MLE) with conserved component

In problem 6, the component of interest is a conserved component, e.g. ortho-phosphate, so that the corresponding graph (Fig. 1f) does not contain any imaginary edge. Fig. 2 shows the obtained Pareto front. One can see 32 distinct bubbles describing the complete Pareto front. These contain a total of 1,565,072 sensor layouts (19% of all candidate layouts). Within this Pareto front, there are 1,531,864 sensor layouts with surplus redundancy (15 red bubbles, 18% of all candidate layouts) and 33,208 layouts without surplus redundancy (17 bubbles, 0.40% of all candidate layouts). In this second category, one can recognize (a) the trivial layout without any sensor (black), (b) 7,437 layouts with some but not all variables observable and one to six non-redundant sensors (white), (c) 24,690 layouts with all variables observable and seven non-redundant sensors (blue), (d) 400 layouts with all variables observable and eight sensors, all of which are redundant (yellow), and (e) 680 layouts with some but not all variables observable and two to seven sensors, some or all of which are redundant (green).

To understand the Pareto solutions better, Fig. 3 provides another look into the set of optimal sensor layouts. Rather than showing the number of redundant sensors on the vertical axis, it shows the fraction of concentration sensors among all installed sensors. This results in dividing most of the bubbles shown in Fig. 2 into multiple bubbles. For example, the blue bubble in Fig. 2, representing 24,690 sensor layouts with seven sensors making all variables observable is now split into four bubbles in Fig. 3. From bottom to top, these are (a) a bubble of 8,880 layouts with five mass flow rate sensors (71% mass flow rate sensors), (b) a bubble of 11,369 layouts with four mass flow rate sensors (57% mass flow rate sensors), (c) a bubble of 4,051 layouts with three mass flow rate sensors (43% mass flow rate sensors), and (d) a bubble of 390 layouts with two mass flow rate sensors (29% mass flow rate sensors). It is particularly striking that the linear profile of bubbles representing layouts with surplus redundancy in Fig. 2 has now been transformed into a set of bubbles appearing like scales on a fish, first with increasing spread (as if from mouth to body) and then with decreasing spread (as if from body to tail). This is a result of two competing effects. First, as the number of sensors is increased, this also increases the flexibility in choice of sensor type for these additional sensors. Second, increasing the number of sensors beyond 14 sensors starts to limit the number of additional sensor options that remain available. This continues until all (14) candidate sensors are selected and no flexibility remains at all. The same phenomenon occurs for every problem involving both flow rate sensors and concentration sensors (Problem 1–3 and 6–9, not shown). By setting the cost weights for flow sensors and concentrations equal to each other one obtains sensor layouts achieving the same levels of structural observability and redundancy while

exploiting a varying number of sensors for each variable. This ability to achieve the same sensor layout performance (observability and redundancy) with layouts with varying distributions of mass flow rate and concentration sensors is a result of giving equal weight to all sensors in the cost and redundancy objectives and to all variables in the observability objective. A simple modification, e.g. increasing the relative cost of concentration sensors, would reduce the number of solutions dramatically. We return to this observed flexibility in the discussion.

3.1.2. Problem 7: WWTP2 (MLE) with reactive component

Problem 7 concerns the case of a reactive component, like any nitrogen-containing component (e.g. ammonia, nitrite, nitrate, total nitrogen). This is reflected in Fig. 1g. Fig. 4 shows the resulting Pareto front. In this case, there are 48 bubbles representing 981,242 layouts (12% of all candidate layouts). The majority of these (909,132 layouts, 11% of all layouts) have surplus redundancy and are spread over 26 bubbles (red bubbles). These bubbles are split into 2 subsets of equal size (each 13 bubbles, each 454,566 layouts). One subset consists of layouts with all 42 variables observable and all sensors redundant except one. The other subset consists of Pareto optimal layouts that make only 37 out of 42 variables observable but exhibit redundancy for all sensors.

A minority of the Pareto optimal layouts (72,110 layouts, 0.86% of all layouts) do not exhibit surplus redundancy and are spread over 22 bubbles. This includes (a) the trivial layout without any sensor (black). This trivial layout produces one observable variable, namely the concentration in stream 9, because it is assumed known. One also recognizes (b) 38,629 layouts with some but not all variables observable and one to eight non-redundant sensors (white), (c) 30,214 layouts with all variables observable and nine non-redundant sensors (blue), (d) 972 layouts with all variables observable and ten sensors, nine of which are redundant (yellow), and (e) 2,294 layouts with some but not all variables observable and two to nine sensors, some or all of which are redundant (green).

Note that reaching complete observability without redundancy in Problem 7 requires two more sensors compared to Problem 6. This is not so surprising, considering that the reactive case includes two additional rate variables (the two reaction rates). For example, the available balances around the reactor and settler (nodes *d* and *e*, component mass flow balances) enable estimation of the unmeasured reaction rates (stream 13 and 14) when the mass flow rates and component concentrations in stream 4, 5, 6, 7, and 8 are measured. This is because the stoichiometric coefficients for stream 13 and 14 are known, so that the two component mass flow balances enable estimation of the reaction rates. This requires 10 sensors. Note that similar effects can be observed for WWTP1 (Problem 2 and Problem 3, see Supplementary Information for details).

A curious diagonal pattern can be seen at the top-left of the figure. One can see two series of bubbles representing layouts with surplus redundancy. The first series starts with a bubble representing 8,001 layouts with 11 sensors that make all (42) variables observable and 10 of the 11 installed sensors redundant. The second series starts with a bubble of the same size (8,001 layouts) with 10 sensors that make all but five variables (37 total) observable and all (10) sensors redundant. The remaining bubbles in each series can be traced by increasing the number of sensors incrementally while simultaneously increasing the number of redundant sensors with the number of sensors added. Importantly, for each increase of the number of installed sensors, the number of layouts in the corresponding bubbles in both series remains the same.

The interesting repeated pattern observed in the previous graph can be explained as follows. The most important property of Problem 7 is that it is impossible to place a redundant concentration measurement in stream 7. More specifically, it is impossible to

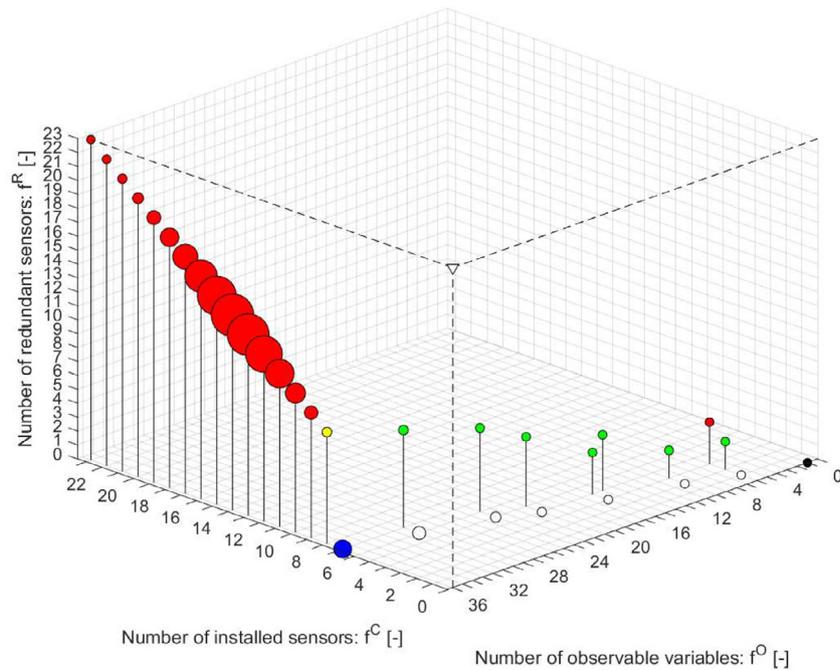


Fig. 2. Problem 6 – Placing sensors for flow rate and concentration of a conserved component (e.g., ortho-phosphate) in WWTP2. Visualization of the Pareto front. The ideal solution is indicated as a white triangle. The size of the bubbles reflects the number of layouts in each point on the Pareto front: (black) zero sensors; (white) incomplete observability, no redundancy; (green) incomplete observability, some or complete redundancy; (yellow) complete observability, some or complete redundancy; (red) surplus redundancy. Reaching observability for all variables and redundancy of all sensors requires only seven sensors. Many other Pareto optimal choices are available, including layouts without observability and layouts with surplus redundancy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

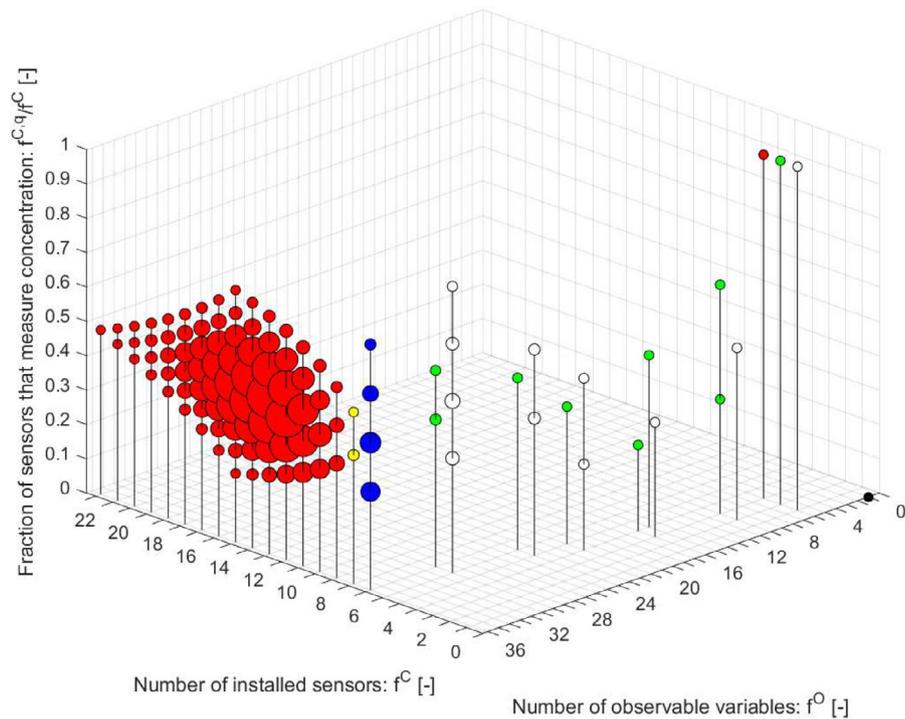


Fig. 3. Problem 6 – Placing sensors for flow rate and concentration of a conserved component in WWTP2. Alternative visualization of solutions on the Pareto given in Fig. 2, showing the number of sensors f^C , the number of observable variables f^O , and the fraction of installed sensors that are mass flow rate sensors $f^C q / f^C$. The size of the bubbles reflects the number of layouts in each point: (black) zero sensors; (white) incomplete observability, no redundancy; (green) incomplete observability, some or complete redundancy; (yellow) complete observability, some or complete redundancy; (red) surplus redundancy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

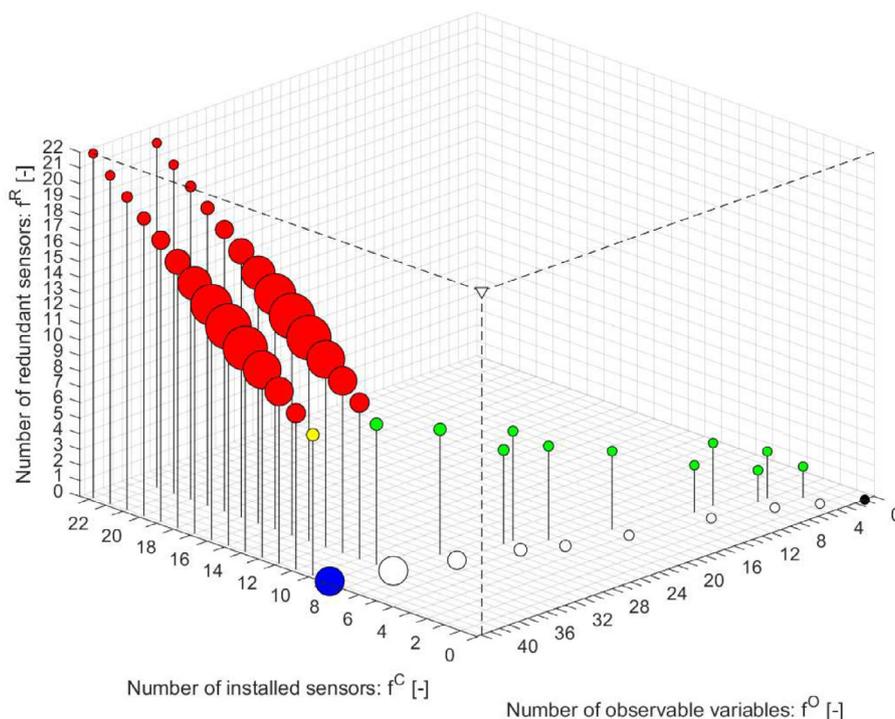


Fig. 4. Problem 7 – Placing sensors for flow rate and concentration of a reactive component (e.g., ammonia) in WWTP2. Visualization of the Pareto front. The ideal solution is indicated as a white triangle. The size of the bubbles reflects the number of layouts in each point on the Pareto front: (black) zero sensors; (white) incomplete observability, no redundancy; (green) incomplete observability, some or complete redundancy; (yellow) complete observability, some or complete redundancy; (red) surplus redundancy. Reaching observability for all variables and redundancy of all sensors requires a minimum of nine sensors, two more compared to Problem 6. Many other Pareto optimal choices are available, including layouts without observability and layouts with surplus redundancy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

write a balance equation involving the concentration in stream 7 which does not involve the reaction rate for stream 13, which is always an unknown variable that needs to be estimated. To see this, consider evaluating redundancy for the concentration sensor in stream 7, i.e. remove this sensor and check if this variable is still observable. When doing so, ρ_{13} and c_7 are two unknowns participating in a single independent balance equation only, meaning that these unknowns cannot be estimated, in turn indicating that the removed sensor is not redundant. The equivalent graph theoretical observation is that there is not a single cutset in the graph (Fig. 1f/g/h) including edge 7 but not including edge 13. More specifically, one cannot split the graph into two graphs so that the nodes connected by edge 7 are disconnected without removing edge 13. For this reason we consider this a structural identifiability issue. Importantly, this means it can only be resolved by modifying the definition of Problem 7. For example, one could consider (a) simultaneous consideration of balances for multiple components, thus requiring concentration sensors for more than one component, (b) by measuring all mass fractions in stream 7 (which is practically impossible for the water mass fraction), (c) allowing for installation of redundant sensors in the same stream (hardware redundancy), or (d) assuming that no reaction involving the component of interest occurs in the settler (thus removing stream 13 from the graph).

3.1.3. Problem 8: WWTP2 (MLE) with conserved solid component

Problem 8 concerns the case of a conserved solid component (e.g., total suspended solids), as reflected in Fig. 1h. The corresponding Pareto front is shown in Fig. 5. There are 33 bubbles representing 1,563,126 layouts (19% of all candidate layouts). The majority of these (1,478,757 layouts, 18% of all layouts) have surplus redundancy and are spread over 14 bubbles (red bubbles). The re-

maining layouts (84,369 layouts, 1.0% of all candidate layouts) do not exhibit surplus redundancy and are spread over 19 bubbles. These bubbles include (a) the trivial layout without any sensor (black), (b) 23,454 layouts with some but not all variables observable and one to seven non-redundant sensors (white), (c) 58,026 layouts with all variables observable and eight non-redundant sensors (blue), (d) 1,500 layouts with all variables observable and nine sensors, all of which are redundant (yellow), and (e) 1,388 layouts with some but not all variables observable and two to nine sensors, some or all of which are redundant (green).

In this case, reaching complete observability without redundancy requires one more sensor compared to Problem 6 and one less compared to Problem 7. This is explained as the consequence of adding the transport rate variable, which is introduced to account for component imbalances in the reactor and the settler and cannot be measured directly. The unmeasured rate participates as a single rate in two independent balance equations: for node *d* (stream 4, 5, and 13) and node *f* (stream 6, 7, 8, and 13). Consider measuring mass flow rates and concentrations in all physical streams. This makes all variables of interest observable. It also makes all sensors redundant as each mass flow rate and component concentration participates in at least one balance equation that excludes the unknown transport rate. Indeed, by considering the balances around nodes *a*, *b*, *c*, *e*, and *g* one obtains 10 independent equations (5 flow balances and 5 component balances) involving 24 measurements. Removing any single measurement, means that one obtains 10 equations with one unknown, so that any single measurement clearly is redundant. Following this path of reasoning, the Pareto front has to include at least one sensor layout for which all sensors are redundant while all variables of interest are observable.

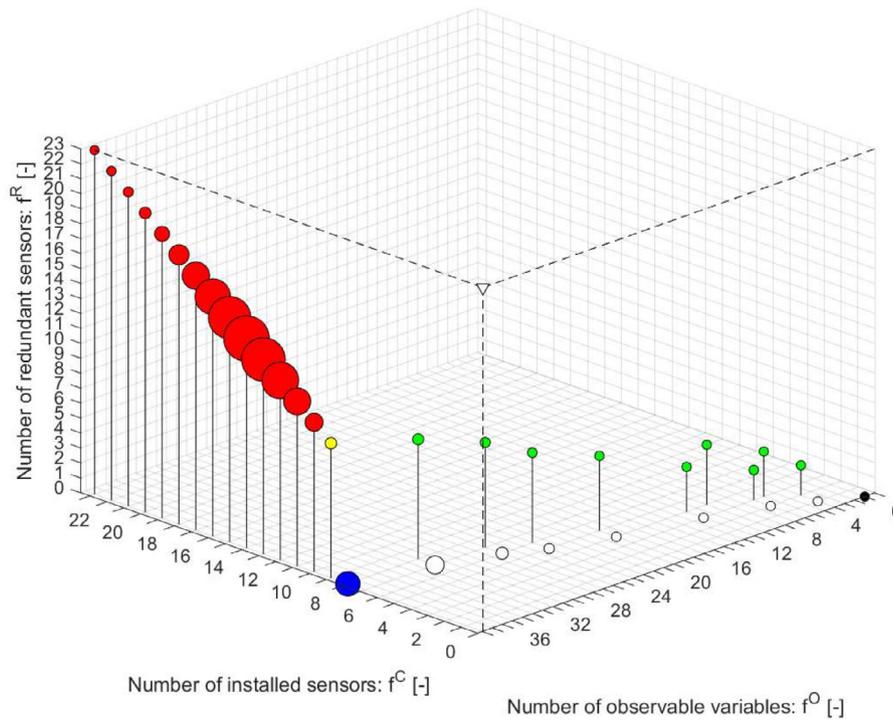


Fig. 5. Problem 8 – Placing sensors for flow rate and concentration of a solid component (e.g., total suspended solids) in WWTP2. Visualization of the Pareto front. The ideal solution is indicated as a white triangle. The size of the bubbles reflects the number of layouts in each point on the Pareto front: (black) zero sensors; (white) incomplete observability, no redundancy; (green) incomplete observability, some or complete redundancy; (yellow) complete observability, some or complete redundancy; (red) surplus redundancy. Reaching observability for all variables and redundancy of all sensors requires a minimum of eight sensors, one more compared to Problem 6 and one less compared to Problem 7. Many other Pareto optimal choices are available, including layouts without observability and layouts with surplus redundancy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.4. Problem 9: WWTP2 with conserved solid component, customized objectives

Problem 9 is a modification of Problem 8 by assuming different weights for the observability objective and knowledge about the flow rates in four streams. These streams are typically equipped with pumps. Such pumps often come with an associated flow rate measurement. Alternatively, the pumps' flow rate setpoints, or signals describing power consumption, heat loss, and pressure loss could act as a proxy for the flow rate. Consequently, there are now only 19 sensor locations available, resulting in 524,288 candidate sensor layouts. One can easily verify that the mass flow rate for stream 8 is observable without any additional sensor.

Fig. 6 shows the resulting Pareto front. This now includes 25 bubbles with 258,268 layouts (49% of all candidate layouts). The majority of these layouts have surplus redundancy (16 red bubbles, 256,957 layouts, 49% of all candidate layouts). The remaining 1,311 layouts (0.25% of all candidate layouts) are distributed over 9 bubbles. These bubbles include (a) the trivial layout without any sensor (black), (b) 181 layouts with some but not all variables observable and one to three non-redundant sensors (white), (c) 709 layouts with all variables observable and four non-redundant sensors (blue), (d) 162 layouts with all variables observable and five sensors, all of which are redundant (yellow), and (e) 258 layouts with some but not all variables observable and two to nine sensors, some or all of which are redundant (green).

3.1.5. Overview of all results

Table 3 gives an overview of the results obtained through solving all nine sensor layout optimization problems.

The flow rate sensor placement problems (Problem 1 and 5) are solved with v1.2 and v2.0 of the SOAR toolbox. For both WWTP1 and WWTP2 the resulting Pareto front is the same regardless of

the chosen software version, as expected. The major difference lies in the computational requirements. For WWTP1, the Pareto front is computed in less than 5 seconds with v1.2 and less than 3 seconds with v2.0. For WWTP2, the improvement is more impressive: a factor 10 improvement from 250 s to less than 25 s. Improvements are partially due to the implementation changes in the GENOBS and GENRED. These are small for GENOBS (reduction of less than 3 s for WWTP2) yet substantial for GENRED (reduction of 40 s for WWTP2). It is noteworthy that the average use of GENRED for evaluation of a single layout for WWTP2 takes only 0.05 ms with v2.0 while the same task requires 12 ms on average with v1.2, thus providing a factor 200 improvement. Despite this impressive relative improvement, the changes in GENOBS and GENRED are responsible for a reduction of the total computational load by 2 s only. The remaining reduction of 220 s for WWTP2 is therefore attributed to (a) an improved implementation of the multi-objective branch-and-bound search method, (b) deployment of the two-step search strategy discussed above. Most importantly, the computationally expensive GENOBS and GENRED procedures are only required in step A.

The seven sensor placement problems involving both flow rate and concentration sensors (Problem 2–4 and 6–9) are solved with v2.0 only. In this case, the computational effort is substantially larger, e.g. up to 51 h for Problem 8. Note however that multi-objective sensor placement problems of this complexity and scale could not be solved before (e.g., with v1.2). One can also see that the sensor layouts without surplus redundancy that are Pareto optimal represent a small fraction of the total set of feasible sensor layouts. For the WWTP2 configurations, this fraction is at most 1%. This has a two distinct advantages. The first advantage is that the computationally expensive GENOBS and GENRED procedures are only required to compute a small subset of the complete Pareto

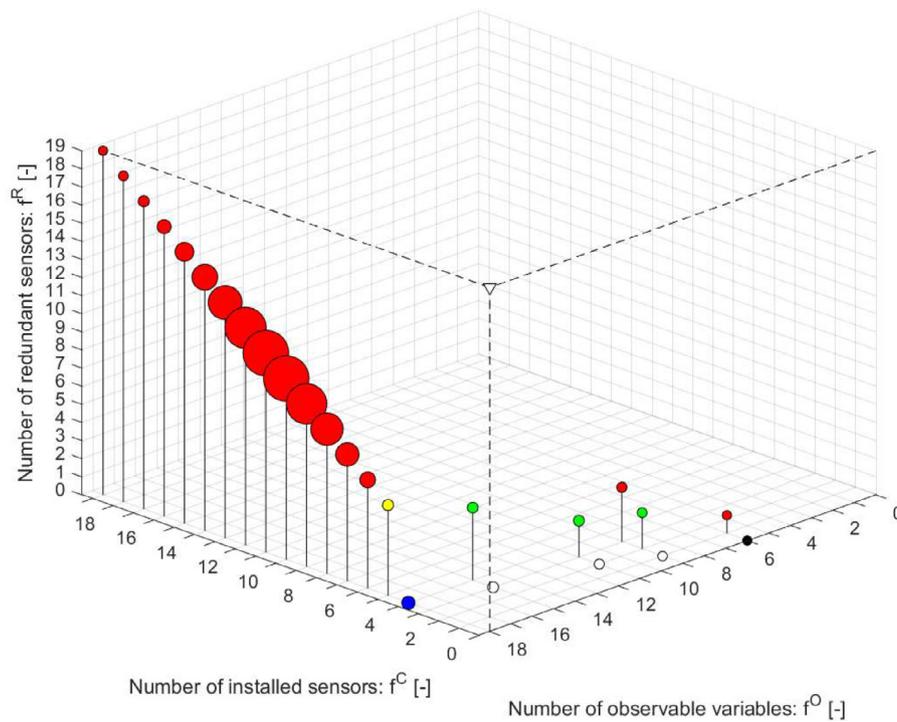


Fig. 6. Problem 9 – Placing sensors for flow rate and concentration of a conserved solid component (e.g., total suspended solids) in WWTP2 with non-uniform weights for the objective functions. Visualization of the Pareto front. The ideal solution is indicated as a white triangle. The size of the bubbles reflects the number of layouts in each point on the Pareto front: (black) zero sensors; (white) incomplete observability, no redundancy; (green) incomplete observability, some or complete redundancy; (yellow) complete observability, some or complete redundancy; (red) surplus redundancy. Reaching observability for all variables of interest and redundancy of all sensors requires a minimum of five sensors. Many other Pareto optimal choices are available, including layouts without observability and layouts with surplus redundancy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Results obtained with all sensor layout optimization problems.

Plant configuration	WWTP1					WWTP2							
	1	2	3	4	5	6	7	8	9				
Problem	1	2	3	4	5	6	7	8	9				
SOAR toolbox	v1.2	v2.0	v2.0	v2.0	v2.0	v1.2	v2.0	v2.0	v2.0				
# Sensor candidates													
Flow rate	[-]	7	7	7	7	12	12	[-]	12	12	12	8	
Concentration	[-]	0	0	7	7	7	0	0	[-]	11	11	11	11
# Sensor layouts													
Feasible	[-]	128	128	16,384	16,384	16,384	4096	4096	[-]	8,388,608	8,388,608	8,388,608	524,288
Evaluated total	[-]	128	128	15,425	12,627	14,574	3537	3964	[-]	7,186,549	6,089,723	7,174,826	497,555
Evaluated in step 1	[-]	-	128	14,569	12,200	13,970	-	3545	[-]	5,362,015	4,658,522	5,293,215	450,790
Evaluated in step 2	[-]	-	0	856	427	604	-	419	[-]	1,824,534	1,431,201	1,881,611	46,765
Pareto optimal	[-]	70	70	6,037	1,501	2,633	872	872	[-]	1,565,072	981,242	1,563,126	258,268
Pareto optimal (no surplus)	[-]	-	47	1,121	547	1,080	523	872	[-]	33,208	72,110	84,369	1,311
Evaluated total	[%]	100	100	94.15	77.07	88.95	86.35	96.78	[%]	85.67	72.60	85.53	94.90
Evaluated in step 1	[%]	-	100	88.92	74.46	85.27	-	86.55	[%]	63.92	55.53	63.10	85.98
Evaluated in step 2	[%]	-	0	5.22	2.61	3.69	-	10.23	[%]	21.75	17.06	22.43	8.92
Pareto optimal	[%]	54.69	54.69	36.85	9.16	16.07	21.29	21.29	[%]	18.66	11.70	18.63	49.26
Pareto optimal (no surplus)	[%]	-	36.72	6.84	3.34	6.59	-	12.77	[%]	0.40	0.86	1.01	0.25
Computational time													
Total	[s]	4.73	2.61	127.71	112.39	133.11	248.78	21.77	[h]	29.19	43.49	50.11	0.99
Time step 1	[s]	-	2.50	124.92	111.76	133.43	-	21.26	[h]	25.52	40.21	42.84	0.95
Time step 2	[s]	-	0.11	2.79	0.63	1.68	-	0.52	[h]	3.67	3.29	7.27	0.04
GENOBS	[s]	0.38	0.12	23.08	20.23	23.87	11.19	8.61	[h]	3.81	3.51	3.72	0.25
GENRED	[s]	1.21	0.05	38.82	34.23	39.69	41.69	0.20	[h]	10.47	11.78	11.72	0.24
GENOBS per layout	[ms]	2.98	0.90	1.58	1.66	1.71	3.16	2.43	[ms]	2.56	2.71	2.53	2.01
GENRED per layout	[ms]	9.45	0.36	2.37	2.09	2.42	11.79	0.05	[ms]	4.49	5.05	5.03	1.63

front (Step A). The second advantage is that structural criteria (for cost, observability, and redundancy) appear sufficient to reduce the number of candidate sensor layouts significantly.

One general phenomenon is that adding variables that cannot be measured directly (e.g., reaction rates, storage) decreases the number of layouts on the Pareto front. Thus, adding additional

degrees of freedom to the data reconciliation problem reduces the number of Pareto-optimal solutions. The table also shows that computing the sensor layouts with surplus redundancy on the basis of the sensor layouts without surplus redundancy can be executed fairly efficiently. Taking Problem 8 as an example, one can see that step 2 of the search strategy requires just under 15% of the

total computational time while this step is responsible to evaluate about 25% of the number of the evaluated sensor layouts.

4. Discussion

4.1. Optimal placement of flow rate and concentration sensors based on structural information

In this study, we search for Pareto optimal sensor layouts by finding sensor layouts which optimally trade off cost of sensor ownership, informativeness, and redundancy. This is achieved by expressing all optimization objectives as flexible functions that are monotonic in binary variables describing whether a sensor is installed, whether a variable is structurally observable, and whether an installed sensor is structurally redundant. The major benefits of this approach are demonstrated for the first time at a realistic scale and level of complexity through this study:

- The method is generally applicable and enables computation of the complete Pareto front within reasonable time windows.
- The Pareto front solutions are enumerated completely. This is achieved by deterministic optimization, which means the set of Pareto solutions is provably optimal and complete.
- Flow rate and concentration sensor layout optimization (i) can be executed for existing plants with sensors (brown-field, retrofitting) as well as for new plants prior to construction (green-field), (ii) does not rely on information of typical sensor signal quality, and (iii) does not require any measurements. In turn, such placement will impact the utility of real-world data by enabling effective data reconciliation at minimal cost, in turn transforming unchecked data produced in complex systems, such as WWTPs, into a valuable resource for machine learning, model identification, and monitoring of daily operations.
- Internal storage and release of components that are conserved at a plant-wide level are accounted for in a systematic manner, thanks to modification of the observability and redundancy labeling procedures of Kretsovalis and Mah (1988a,b). This modification is important as it leads to a Pareto front that is suitable for the total solids in wastewater plants. More specifically, it allows to account for temporary imbalances of the solids in individual subsystems of the plant (e.g., reactor, settler), e.g. due to rain events, while assuming that the solids content of the plant as a whole remains intact under normal operational conditions.
- Every Pareto front is found within 2.5 days of computing time on a single machine. This is in part due to the modified GENOBS and GENRED procedures for observability and redundancy labeling, which were implemented efficiently by enumerating all cycles and cutsets of the graphs a priori. Automated screening of candidate sensor layouts is therefore possible with conventional computing infrastructures.

4.2. Effects of objective functions, available sensor types and locations, and system topology on the resulting Pareto front

The relationship between the problem specification, i.e. the objective functions, the available sensor types and locations, and the structure of the system graph and its solution, i.e. the resulting Pareto front, is a highly nonlinear. Moreover, small changes in the problem specification can lead to large changes in its solution. Consequently, statements about this relationship that are true in general will have limited use in specific instances of the sensor placement problem. This is exactly why the ability to efficiently compute the Pareto front for general plant configuration is so valuable. Nevertheless, we mention a few general lessons that can be learned, either through graph theoretic considerations or from the results discussed above:

- Increasing the number of streams without increasing the number of cycles in the graph e.g., by representing a single pipe with two separate pipe sections - will lead to the same Pareto front as long as the number of sensor candidate sensors remains the same. This is because the added variables can be removed from any data reconciliation problem by simple application of the equality constraints describing the balances over the node connecting the two pipe sections. If instead additional sensor candidates are added for every added variable, then one should expect the Pareto front to be in the same location, while many points on the Pareto front will contain more sensor layouts (larger bubbles).
- Increasing the number of cycles without changing the number of streams, tends to move the Pareto front to solutions that require more measured flow rates and concentrations to achieve the same level of observability and redundancy. This follows from the idea that making a flow (component flow rate) for a particular stream is observable when every cycle with this stream must have at least one stream with a measured flow rate (component flow rate) or an independent estimate thereof.
- Adding a sensor to the set of candidate sensors for a variable that can be made observable already without this sensor will increase the number of layouts in the points on the Pareto front. Quite possibly, a better Pareto front may be obtained, e.g. if a concentration sensor is added for a concentration variable that otherwise requires a combination of flow sensors and concentration sensors to become observable.
- As observed in the results section, setting weights for distinct sensor types and location equal to each other, thus indicating equal preference for any candidate sensor leads to a large number of sensor layouts on the Pareto front. The solicitation of stronger preferences, e.g. a preference for flow rate sensors rather than concentration sensors, will significantly reduce the number of sensor layouts on the Pareto front.

4.3. Nature of sensor layout optimization in wastewater treatment plants

The following additional observations appear valid for typical WWTPs:

- Accounting for reactions (e.g., involving total ammonia, or the organic soluble fraction) and transport phenomena (e.g., for solids) adds additional variables without changing the number of balance equations. It is therefore not a surprise that the consideration of such phenomena requires additional sensors to achieve the same level of observability and redundancy, mainly to compensate for the additional degrees of freedom implied by unmeasurable rates (e.g., reaction rates, transport rates). In practical situations, one may choose to ignore certain reactions and transport phenomena. E.g., for data reconciliation over long periods of time, e.g. months to years, one may expect that short-term effects of reaction and internal transport can be ignored, thereby increasing the number of redundant sensors and the number of balance equations available for data reconciliation. In turn, this leads to improved precision of reconciled measurements and shorter times between the onset of sensor faults and their detection. In shorter time scales, e.g. days to weeks, it may very well be possible to ignore reactions still and only consider internal transport as a source of temporary imbalances within the plant. This would be the case of the total solids component in a plant subject to rain events. On the other hand, being able to assess reaction and transport rates can be very useful for process troubleshooting and optimization. Thus, it is important to define the sensor layout optimization objec-

tives according to the intended time scales for data reconciliation.

- The fraction of optimal layouts among all candidate layouts is fairly small (e.g., less than 20% for the most complex problems with the realistic WWTP2 configuration, 6–8). However, the total number of Pareto-optimal layouts tends to be large still (more than a million in some cases). We discuss a number of ways to reduce this number below. Based on current results, the resulting Pareto front is especially valuable to assist the end-user with further elucidation of sensor layout preferences.
- Restricting oneself to the most cost-effective sensor layouts, by selecting sensor layouts without surplus redundancy only, reduces the number of optimal sensor layouts to a very small fraction of the candidate sensor layouts (e.g., 1% or less for the most complex problems with the realistic WWTP2 configuration, 6–8).
- In all studied problems, the total number of sensors (flow rate and concentration) required to achieve complete observability and complete redundancy of all sensors is only one more than the total number of sensors (flow rate and concentration) to achieve complete observability without any redundancy. This shows that the additional investment necessary to transition from complete observability to complete observability and redundancy is likely small in many WWTPs, even when flow rates and concentrations need to be reconciled simultaneously.

4.4. Criticism

Despite the advances reported above, a few points of improvement must be considered for future work. These are:

- Equal weights were assumed for the cost of flow rate and concentration sensors. This leads to extreme flexibility. We illustrated this by showing how the fraction of sensors of one type can vary significantly within the sets of Pareto optimal sensor layouts delivering the same values for the optimization objectives. Several options are available to reduce this flexibility, should it be desired:
 1. Modify the weights of the cost function so that certain sensors, i.e. for a specific variable or in specific locations, cost more than others. For instance, sensor location with difficult access or sensors whose cost of ownership is expected to be high can be given a larger weight.
 2. Add an additional objective function in the multi-objective optimization problem to discriminate between sensor layouts with otherwise equal values for objective functions. E.g., one can add the number of concentration sensors as an additional objective to prefer sensor layouts with a larger fraction of flow rate sensors.
 3. Include practical objectives for observability (e.g., expected bias, expected precision) and redundancy (e.g., expected fault detection rates [Ali and Narasimhan, 1996](#), variance of reconstruction error [Valle et al., 1999](#)) as a way to discriminate between sensor layouts that are otherwise equivalent in performance.
- The system scale, while realistic for wastewater plants, is relatively small to more complex networks, such as sewer systems and drinking water networks. We expect relative ease to scale these methods to sewer systems, due to the tree-like structure of most sewer system graphs. Drinking water distribution networks are far more dense however and may therefore provide more challenging cases. Given that most of the computational effort is spent on the search for the Pareto front, rather than the GENOBS and GENRED procedures, we believe that a more efficient optimization scheme is desirable for such cases. This

could be achieved by designing algorithms for multi-objective deterministic optimization that exploit the nature of the sensor layout optimization problem better. Quite possibly, the use of parallel computing could also facilitate a faster delivery of the Pareto front. Alternatively, one can also abandon guaranteed optimality and completeness of Pareto front, thus allowing the use of stochastic optimization algorithms, such as genetic algorithms ([Sen et al., 1998](#); [Gerken and Heyen, 2008](#)).

- In this and prior work, we have only considered observability and redundancy objectives. However, redundancy only ensures that one can recognize the presence of sensor faults (detection), not necessarily identify them (isolation) (e.g., [Luong et al., 1997](#); [Commault and Dion, 2007](#); [Prez et al., 2011](#); [Sarrate et al., 2014](#); [Palleti et al., 2016](#)). To do so, the value of surplus redundancy should be studied in more detail. Adding an objective to evaluate structural isolability would be valuable for this (as in [Bhushan and Rengaswamy, 2000](#)). For instance, one could count the number of sensors for which a fault, when it occurs, can be uniquely attributed to this sensor. In our opinion, it might be even better to count the number of sensor pairs for which one can uniquely identify the faulty one, should one of the sensors in the pair be faulty. In addition, we believe that more refined sensor cost objectives and practical measures of observability, redundancy, and isolability are desirable objectives to include into our systematic approach for multi-objective sensor layout optimization.

4.5. Outlook

Our results show promise for research directions not considered yet:

- The evaluation of observability and redundancy in cases where linear balance equations, bilinear balance equations, and concentration equality in splitter nodes need to be considered can be executed very fast thanks to prior computation of all graph cutsets and graph cycles. The enumeration of the complete set of graph cutsets is equivalent to the enumeration of every linear flow balance equation or every bilinear balance equation for a single component. Thus, this means that one can also quickly evaluate whether a particular balance equation should be included to solve a particular data reconciliation problem. Selecting the linear total mass (bilinear component) flow balances that should be included is equivalent to selecting all cutsets among the cutsets which have streams with observable total mass flow rates (component flow rates) only and at least one redundant sensor. This means that it is likely that the automatic setup of data reconciliation problems could be executed in less than a second, which is three orders of magnitude faster than the result by [Le \(2019\)](#) through adoption of the method described in [Spindler \(2014\)](#). Fast formulation of data reconciliation problems could be essential when reconciled data are very valuable, especially in on-line control settings where the availability of measurement can change frequently (e.g., due to sensor or communication failures, or maintenance efforts). Note that a multi-parametric programming approach may solve this problem too ([Teles et al., 2012](#)).
- A quick evaluation of structural observability and redundancy would also be valuable in pressured networks, including sewer systems and drinking water supply networks. In this case, there are nonlinear equations describing the relationship between pressure drops and flow rates. To our understanding, there is no graph-theoretic approach for simultaneous labelling of flow rates, pressures, and concentrations (e.g., contaminants) in such networks. While not explored in detail yet, it is noteworthy that every cycle composed of edges representing physical streams

corresponds to a single pressure balance equation, i.e. the net pressure loss across the pipes in a loop must always be zero.

5. Conclusions

The optimal placement of flow rate and concentration sensors is addressed by solving a multi-objective integer sensor layout optimization problem considering weighted cost, weighted structural observability, and weighted structural redundancy. We solve this type of problems for the first time at a realistic scale and level of complexity and achieve this with a generally applicable method for physical systems described by linear and bilinear constraints involving total mass flow rates, heat flow rates, temperature, and mass flow rates and concentrations of any number of physico-chemical components. The applied graph-theoretic method is modified to account for internal storage and release of components that are otherwise conserved and special attention was given to an efficient implementation of this method. As a result, the provided method is efficient while it is applicable to any system described by linear and bilinear balance equations.

An interesting practical result is that the investments in sensors to move from complete observability to complete observability and redundancy requires only one more sensor (flow rate or concentration) in realistic plant configurations. This appears true despite effects of complexity, nonlinearity, and scale.

Finally, the applied optimization method is applicable for any plant configuration without data collection requirements or specification of sensor properties. This property of the method is attractive as a way to design data acquisition systems that produce informative and high-quality sensor signals. This is considered useful especially for tasks involving empirical models, for which reliable sensor signals are paramount.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRedit authorship contribution statement

Kris Villez: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Visualization, Writing - original draft. **Peter A. Vanrolleghem:** Conceptualization, Validation, Writing - review & editing. **Lluís Corominas:** Conceptualization, Validation, Writing - review & editing.

Acknowledgements

The authors thank Sergii Iglın for his graph theory toolbox and Eli Duenisch and Will Robertson for code facilitating visualisation and reporting of our results. 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. Lluís Corominas acknowledges the Ramon and Cajal grant RYC-2013-14595 and the I3 consolidation from the Spanish Ministry. ICRA is recognized as a consolidated research group by the Catalan Government with code 2017-SGR-1318.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compchemeng.2020.106880.

References

- Alferes, J., Tik, S., Copp, J., Vanrolleghem, P.A., 2013. Advanced monitoring of water systems using in situ measurement stations: data validation and fault detection. *Water Sci. Technol.* 68 (5), 1022–1030.
- Ali, Y., Narasimhan, S., 1996. Sensor network design for maximizing reliability of bilinear processes. *AIChE J.* 42 (9), 2563–2575.
- Bhushan, M., Rengaswamy, R., 2000. Design of sensor location based on various fault diagnostic observability and reliability criteria. *Comput. Chem. Eng.* 24 (2–7), 735–741.
- Chang, N.B., Prapinpongsonone, N., Ernest, A., 2012. Optimal sensor deployment in a large-scale complex drinking water network: comparisons between a rule-based decision support system and optimization models. *Comput. Chem. Eng.* 43, 191–199.
- Chmielewski, D.J., Palmer, T., Manousiouthakis, V., 2002. On the theory of optimal sensor placement. *AIChE J.* 48 (5), 1001–1012.
- Commaut, C., Dion, J.M., 2007. Sensor location for diagnosis in linear systems: structural analysis. *IEEE Trans. Automat. Control* 52 (2), 155–169.
- Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortes, U., Poch, M., 2018. Transforming data into knowledge for improved wastewater treatment operation: a critical review of techniques. *Environ. Modell. Softw.* 106, 89–103.
- 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.
- Crowe, C.M., 1989. Observability and redundancy of process data for steady state reconciliation. *Chem. Eng. Sci.* 44 (12), 2909–2917.
- Deo, N., 2004. *Graph Theory with Applications to Engineering and Computer Science*. PHI Learning Pvt. Ltd..
- de Winter, C., Palleti, V.R., Worm, D., Koopij, R., 2019. Optimal placement of imperfect water quality sensors in water distribution networks. *Comput. Chem. Eng.* 121, 200–211.
- Ehrgott, M., Gandibleux, X., 2002. *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*. Springer US, pp. 369–444.
- Gerkens, C., Heyen, G., 2008. Sensor placement for fault detection and localization. *Comput. Aided Chem. Eng.* 25, 355–360.
- Gertler, J., Singer, D., 1990. A new structural framework for parity equation-based failure detection and isolation. *Automatica* 26, 381–388.
- Hauduc, H., Gillot, S., Rieger, L., Ohtsuki, T., Shaw, A., Takács, I., Winkler, S., 2009. Activated sludge modelling in practice: an international survey. *Water Sci. Technol.* 60 (8), 1943–1951.
- International Water Association, 2019. Digital water - industry leaders chart the transformation journey. International Water Association. https://iwa-network.org/wp-content/uploads/2019/06/IWA_2019_Digital_Water_Report.pdf.
- Joshi, S., Boyd, S., 2008. Sensor selection via convex optimization. *IEEE Trans. Signal Process.* 57, 451–462.
- Kretsovalis, A., Mah, R.S.H., 1988. Observability and redundancy classification in generalized process networks - I. Theorems. *Comput. Chem. Eng.* 12 (7), 671–687.
- Kretsovalis, A., Mah, R.S.H., 1988. Observability and redundancy classification in generalized process networks - II. Algorithms. *Comput. Chem. Eng.* 12 (7), 689–703.
- Le, Q.H., 2019. Mass Balance-Based Experimental Design and Data Reconciliation for Wastewater Treatment Processes. Ghent University Ph.D. thesis.
- Le, Q.H., Verheijen, P.J., van Loosdrecht, M.C., Volcke, E.I., 2018. Experimental design for evaluating WWTP data by linear mass balances. *Water Res.* 142, 415–425.
- Luong, M., Maquin, D., Ragot, J., 1997. Sensor network design for failure detection and isolation. *IFAC Proc. Vol.* 30 (6), 823–828.
- Metcalf & Eddy (Ed.), 2003. *Wastewater Engineering: Treatment, Disposal and Reuse*. McGraw-Hill, New York.
- Mukherjee, R., Diwekar, U.M., Vaseashta, A., 2017. Optimal sensor placement with mitigation strategy for water network systems under uncertainty. *Comput. Chem. Eng.* 103, 91–102.
- Nahar, J., Liu, J., Shah, S.L., 2019. Parameter and state estimation of an agro-hydrological system based on system observability analysis. *Comput. Chem. Eng.* 121, 450–464.
- Nemhauser, G.L., Wolsey, L.A., 1988. *Integer and Combinatorial Optimization*. New York: Wiley.
- Newhart, K.B., Holloway, R.W., Hering, A.S., Cath, T.Y., 2019. Data-driven performance analyses of wastewater treatment plants: a review. *Water Res.* 157, 498–513.
- Ohmura, K., Thürlimann, C.M., Kipf, M., Carbajal, J.P., Villez, K., 2019. Characterizing long-term wear of ion-selective pH sensors. *Water Sci. Technol.* 80 (3), 541–550.
- Palleti, V.R., Narasimhan, S., Rengaswamy, R., Teja, R., Bhallamudi, S.M., 2016. Sensor network design for contaminant detection and identification in water distribution networks. *Comput. Chem. Eng.* 87, 246–256.
- Prez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., Peralta, A., 2011. Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Eng. Pract.* 19 (10), 1157–1167.
- Rico-Ramirez, V., Frausto-Hernandez, S., Diwekar, U.M., Hernandez-Castro, S., 2007. Water networks security: a two-stage mixed-integer stochastic program for sensor placement under uncertainty. *Comput. Chem. Eng.* 31, 565–573.
- Rieger, L., Langergraber, G., Siegrist, H., 2006. Uncertainties of spectral in situ measurements in wastewater using different calibration approaches. *Wat. Sci. Technol.* 53(12), 187–197.
- 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.
- Rieger, L., Thomann, M., Gujer, W., Siegrist, H., 2005. Quantifying the uncertainty of on-line sensors at WWTPs during field operation. *Water Res.* 39 (20), 5162–5174.
- Rosén, C., Olsson, G., 1998. Disturbance detection in wastewater treatment plants. *Water Sci. Technol.* 37 (12), 197–205.
- Rosén, C., Rieger, L., Jeppsson, U., Vanrolleghem, P.A., 2008. Adding realism to simulated sensors and actuators. *Water Sci. Technol.* 57 (3), 337–344.
- Sarrate, R., Blesa, J., Nejjari, F., Quevedo, J., 2014. Sensor placement for leak detection and location in water distribution networks. *Water Sci. Technol.* 14 (5), 795–803.
- Sen, S., Narasimhan, S., Deb, K., 1998. Sensor network design of linear processes using genetic algorithms. *Comput. Chem. Eng.* 22, 385–390.
- Serpas, M., Hackebeil, G., Laird, C., Hahn, J., 2013. Sensor location for nonlinear dynamic systems via observability analysis and MAX-DET optimization. *Comput. Chem. Eng.* 48, 105–112.
- Soldevila, A., Blesa, J., Tornil-Sin, S., Fernandez-Canti, R.M., Puig, V., 2018. Sensor placement for classifier-based leak localization in water distribution networks using hybrid feature selection. *Comput. Chem. Eng.* 108, 152–162.
- 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.
- Teles, J., Castro, P., Matos, H., 2012. Global optimization of water networks design using multiparametric disaggregation. *Comput. Chem. Eng.* 40, 132–147.
- Thomann, M., 2008. Quality evaluation methods for wastewater treatment plant data. *Wat. Sci. Technol.* 10, 1601–1609.
- Thomann, M., Rieger, L., Frommhold, S., Siegrist, H., Gujer, W., 2002. An efficient monitoring concept with control charts for on-line sensors. *Water Sci. Technol.* 46 (4–5), 107–116.
- Valle, S., Li, W., Qin, S.J., 1999. Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods. *Ind. Eng. Chem. Res.* 38 (11), 4389–4401.
- Venkatasubramanian, V., 2019. The promise of artificial intelligence in chemical engineering: is it here, finally? *AIChE J.* 65 (2), 466–478.
- Villez, K., 2019. GENOBS* and GENRED* Algorithms for Observability and Redundancy Labeling. Technical Report No. 8, v1.0. Technical Report. Eawag.
- Villez, K., Rosén, C., Anctil, F., Duchesne, C., Vanrolleghem, P.A., 2008. Qualitative representation of trends: an alternative approach to process diagnosis and control. *Water Sci. Technol.* 57 (10), 1525–1532.
- Villez, K., Vanrolleghem, P.A., Corominas, L., 2016. Optimal flow sensor placement on wastewater treatment plants. *Water Res.* 101, 75–83.
- 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.