Topological state estimation in water distribution systems: Mixed integer quadratic programming approach

Sarai Díaz[1], Roberto Mínguez[2], and Javier González[3]

[1]Dr. Eng, Dept. of Civil Eng., Univ. of Castilla-La Mancha, Av. Camilo José Cela s/n, 13071 Ciudad Real (Spain). E-mail: Sarai.Diaz@uclm.es.

[2]Dr. Eng, HIDRALAB INGENIERÍA Y DESARROLLOS, S.L., Spin-Off UCLM, Hydraulics Laboratory Univ. of Castilla-La Mancha, Av. Pedriza, Camino Moledores s/n, 13071 Ciudad Real (Spain). E-mail: roberto.minguez@hidralab.com.

[3]Dr. Eng, Dept. of Civil Eng., Univ. of Castilla-La Mancha, Av. Camilo José Cela s/n, 13071 Ciudad Real (Spain). // HIDRALAB INGENIERÍA Y DESARROLLOS, S.L., Spin-Off UCLM, Hydraulics Laboratory Univ. of Castilla-La Mancha, Av. Pedriza, Camino Moledores s/n, 13071 Ciudad Real (Spain). E-mail: Javier.Gonzalez@uclm.es.

The final publication is available at

https://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000934

ABSTRACT

State estimation (SE) techniques can be applied to compute the most likely hydraulic state of a water distribution system from the available measurements at a given time. Different approaches exist in the technical literature to undertake such an analysis, but in all of them it is assumed that pump and valve statuses are known beforehand. Such consideration may lead to unrealistic results if real-time unnotified changes in the operation of the network take place, thus limiting the usefulness of the information provided by telemetry systems. This work drops the known-status assumption and presents the concept of *topological state estimation* (TSE), which permits not only to compute the

hydraulic state of the system, but also the current pump and valve status according to the existing measurements. More specifically, a novel methodology for TSE is set out in this paper. The proposed method is derived from the original mixed integer non-linear programming formulation of the problem, which is transformed in an iterative mixed integer quadratic programming problem by linearizing some hydraulic constraints. The potential of the methodology is presented by means of an illustrative example and a large case study, where pumps, gate valves and check valves exist. Results show that TSE would successfully contribute to make the most of available telemetry systems, hence expanding the online monitoring possibilities of water distribution networks.

Keywords: weighted least squares, network topology, monitoring, observability, reliability

INTRODUCTION

Nowadays, telemetry systems have become essential for the online monitoring of large water distribution networks. Such systems collect real-time data of the metering devices distributed throughout the network, which can eventually be converted into information about its hydraulic state. In this regard, *state estimation* (SE) has been considered as an efficient technique to process available measurements in both power supply (Schweppe and Wildes 1970) and water systems (Coulbeck 1977) for many years. A SE algorithm provides the most likely hydraulic state of a network by minimising the differences between the available measurements and the estimated variables (i.e. pressure, flow and demand) at a given time, which are related to each other by the flow governing equations (Díaz et al. 2016b).

SE has been traditionally posed as a weighted least-squares (WLS) problem in the water domain (Bargiela 1984; Brdys and Ulanicki 2002). Nevertheless, alternative approaches have been presented over the years to enhance the performance of the state estimator. On the one hand, Sterling and Bargiela (1984) adapted the weighted least-absolute-values (WLAV) estimation from the power field (Irving et al. 1978) in order to better deal with gross errors, and Powell et al. (1988) even developed a method combining the advantages of WLS and WLAV in one algorithm. On the other hand, Bargiela and Hainsworth (1989) highlighted the importance of evaluating state estimation uncertainty, based on which the so called set-bounded state estimation (SBSE) problem was formulated (Brdys and Chen 1993; Gabrys and Bargiela 1996), and several other uncertainty evaluation strategies were presented (Nagar and Powell 2000; Díaz et al. 2016a). Concurrently, variations of the traditional WLS have been developed with different purposes, such as dealing with low measurement redundancy (Andersen and Powell 2000), introducing measurement bounds in the objective function (Andersen et al. 2001), or using graph theory to reduce the complexity of the problem (Carpentier and Cohen 1991; Kumar et al. 2008). In all of these approaches, the state estimator was applied to a measurement configuration assuming a given network topology, i.e. considering a given pump and valve status configuration. In this paper we denote network topology as the pump and valve status configuration of the system, i.e. its connectivity. This is common practice in water distribution system management in general (Giustolisi et al. 2008), and state estimation applications in particular (Díaz et al. 2017b).

However, the status of pumps and valves (i.e. network topology) changes with frequency in water distribution systems to adapt the network to the population needs. Some of these changes are considered normal operation procedures that are required in order to provide high-quality supply at different times. Such is the case of the intentional starting or stopping of a pump, or the deliberate closure of a given valve. Also, the status of pumps and valves in the system can change due to the failure of one of the elements in the network (e.g. pipe burst, pump out of service) or as a response to such failures in order to enhance reliability (Wright et al. 2015), i.e. the ability of the system to perform well in abnormal operation conditions (Goulter 1995). Therefore, even though water utilities are normally aware of some usual operating scenarios, it is difficult to keep record of all the changes that take place in the network topology in real-time applications, i.e. it is complicated to keep an up-to-date hydraulic model of the system. This constitutes a major limitation for the online implementation of SE techniques, as the aforementioned methodologies assumed a given topology that may not exactly simulate the current situation of the water distribution network. It is worth mentioning that if the assumed topology is not consistent with the actual system, SE results may be unrealistic, as measurements provided by telemetry systems correspond to a reality that is different from the hydraulic simulation model.

The importance of incorporating changes in the network topology has been discussed before in the context of solving (Giustolisi et al. 2008) and calibrating (Lansey and Basnet 1991; Liberatore and Sechi 2009; Laucelli et al. 2011; Sophocleous et al. 2016) water systems. Also, Díaz et al. (2017b) delved in the possibilities of topological observability analysis in water distribution systems. This approach adapted the traditional observability analysis, which is a prior necessary step to any SE algorithm (Carpentier and Cohen 1991; Díaz et al. 2016), to the possibility of now knowing the pump and valve status beforehand. This method permits to identify not only the hydraulic variables that could be computed in a subsequent SE process, but also the status of which pumps and valves could be inferred from the available measurements. Nevertheless, this analysis only permits to asses if there are sufficient algebraic relationships to infer the pump or valve status from the available measurements, but as measurements are prone to errors, it does not guarantee that the estimated status is correct. When measurement uncertainty is high, sometimes there may not be enough statistical significance to infer the pump or valve status, even if they are observable. To take into account noisy measurements and properly estimate the status of controlling devices, a method for *topological state estimation* (TSE) must be developed.

TSE drops the assumption of the network topology being known at each time, enabling to consider that pumps and valves can be either open or closed, and the correct status must be inferred from the available measurements. Therefore, such methodology permits not only to compute the hydraulic state of the system, but also the current network topology by including as many binary variables as the number of pumps and valves that exist in the network. The inclusion of binary variables leads to a mixed integer non-linear programming (MINLP) problem that is difficult to solve, especially when a large number of controlling devices exist in the system. The aim of this work is to present a novel methodology for TSE that transforms such a challenging formulation into an iterative mixed integer quadratic programming (MIQP) problem. The proposed algorithm iteratively linearizes the hydraulic constraints that introduce non-linearity to the problem (i.e. headloss equations and characteristic curves of the pumps in the system) and solves the TSE optimisation problem, which is formed by a quadratic objective function and a set of linear hydraulic constraints. Expressions for the consistent simulation of pumps, gate valves and check valves are here provided. Moreover, the proposed method permits to take into account that either the characteristic curve of a pump is known or completely unknown, being possible to estimate the most likely pumping head even in the latter case. In this study, it is assumed that the number and location of meters have been previously assessed via topological observability analysis, and the measurements from the meters were synthetically generated. Also, the hydraulic network model used in this study has been previously calibrated.

The paper is organised as follows: firstly, the formulation of TSE as a MINLP problem is explained, including a description of the objective function, the general hydraulic constraints and the specific topological hydraulic constraints required to simulate the presence of pumps and valves. Then, the adopted solution strategy, which converts the MINLP problem in an iterative MIQP problem, is provided. The novel methodology is subsequently applied to a small example, which permits to illustrate the potential of TSE. Afterwards, the C-Town case study is analysed to prove that the method is robust for larger systems, and finally, relevant conclusions are drawn.

TOPOLOGICAL STATE ESTIMATION

Problem formulation

Objective function

The TSE problem can be posed as a non-linear WLS mathematical programming problem with objective function:

$$\underset{h_i;\forall i\in\mathcal{V}}{\text{Minimize}} \left[\sum_{\forall i\in\mathcal{V}_\text{m}} \left(\frac{\tilde{h}_i - h_i}{\sigma_i^h} \right)^2 + \sum_{\forall i,j\in\mathcal{L}_\text{m}^\text{PI}} \left(\frac{\tilde{Q}_{i,j} - Q_{i,j}}{\sigma_{i,j}^Q} \right)^2 + \sum_{\forall i\in\mathcal{V}_\text{m}^\text{Q}} \left(\frac{\tilde{q}_i - q_i}{\sigma_i^q} \right)^2 \right], \tag{1}$$

where the squared difference between existing measurements and estimated variables is to be minimised. In terms of variables, $h_i; \forall i \in \mathcal{V}$ refers to head levels at all the nodes in the system ($\mathcal{V}$), $Q_{i,j}; \forall i, j \in \mathcal{L}^\text{PI}$ is the water flow through the pipe elements in the system ($\mathcal{L}^\text{PI}$), and $q_i; \forall i \in \mathcal{V}^\text{Q}$ is the water consumption at demand nodes ($\mathcal{V}^\text{Q}$). On the other hand, variables overlined with a tilde represent the value of the measurement in the subset of nodes and pipes where variables are measured. Therefore, $\mathcal{V}_\mathrm{m}$, $\mathcal{L}_\mathrm{m}^\mathrm{PI}$ and $\mathcal{V}_\mathrm{m}^\mathrm{Q}$ represent the subset of nodes where head levels are measured, the subset of pipes where flows are metered, and the subset of nodes where water demands are measured, respectively. Consequently, $\sigma_i^h; \forall i \in \mathcal{V}_\mathrm{m}$, $\sigma_{i,j}^Q; \forall i,j \in \mathcal{L}_\mathrm{m}^\mathrm{PI}$, and $\sigma_i^q; \forall i \in \mathcal{V}_\mathrm{m}^\mathrm{Q}$ are the standard deviations for measurements of head, flow and demand, respectively. Variables and measurements are independent of time t all along this work because a pseudo-static approach is considered for TSE, i.e. flow is steady and each estimation can be understood as an instantaneous snapshot of the system. Also, it must be highlighted that only measurement uncertainty is taken into account in Eq. (1). This is the traditional scope for state estimation methodologies, which normally assume that the hydraulic model (e.g. pipe infrastructure information, tank dimensions) has been previously calibrated (Díaz et al. 2016).

At this point, it is important to identify the state variables of the problem, which are the minimum set of variables that enable to characterise the hydraulic state of the system. Note that if the status of pumps and valves is known, nodal heads can be considered the state variables of the system, as any combination of head levels leads to a certain and credible flow solution (Díaz et al. 2016). However, if some of the statuses of pumps and valves are unknown, the state variable set must also include the unknown binary variables, without which the state of the system cannot be fully defined. As it will be presented later on, binary variables enable to simulate that pumps and valves can either be open or closed, and the TSE algorithm must determine their status based on the existing measurements. Once state variables are computed, the rest of variables (i.e. flows and demands) can be inferred thanks to the rest of hydraulic constraints, which are now presented.

Hydraulic constraints

In this work, the Hazen-Williams headloss equation is assumed all along:

$$h_i - h_j = K_{i,j} Q_{i,j} |Q_{i,j}|^{c-1}; \forall i,j \in \mathcal{L}^\mathrm{PI}, \tag{2}$$

where $K_{i,j}$ is the flow resistance pipe coefficient, $c = 1.852$ is the Hazen-Williams exponent and h_i and h_j refer to the head levels at the initial and final nodes of the pipeline. Note that adopting the Darcy-Weisbach equation instead of the Hazen-Williams approach would require calculation of the friction factor from an implicit function. Therefore, and even though the Hazen-Williams formula for pipe head loss applies over a limited range of flows, it is a common choice in the field because its explicit formulation is easy to compute (Eck and Mevissen 2012). In this work, flow is taken as positive when water moves from the lower to the higher numbering node. For each node, two subsets can be defined. Ω_i^{O} contains the water outflows from node i to the rest of nodes $j > i$ connected to i with a pipe. On the other hand, Ω_i^{I} corresponds to the inflows to node i from the rest of nodes $j < i$ connected to i. Note that the head level can be estimated from the node elevation and pressure. However, as pumps and valves must be included in the hydraulic model for TSE purposes, computation of the head level at each node must also take into account the energy provided by pump elements ($\forall i, j \in \mathcal{L}^{\mathrm{P}}$, being $\mathcal{L}^{\mathrm{P}}$ the set of all pumps in the network):

$$h_i = e_i + p_i + \sum_{k,l \in \mathcal{L}^{\mathrm{P}}} \delta_{i,k,l} h_{k,l}^{P}; \forall i \in \mathcal{V}, \tag{3}$$

where $e_i; \forall i \in \mathcal{V}$ is the elevation at the node (considered a constant parameter), and $p_i; \forall i \in \mathcal{V}$ is its pressure, being $\mathcal{V}$ the set of all nodes in the network. The following summand refers to the pumping head $h_{k,l}^{P}; \forall k, l \in \mathcal{L}^{\mathrm{P}}$ that any pump can inject at a specific node. If node i is the final node of a given pump $k, l \in \mathcal{L}^{\mathrm{P}}$, then Kronecker delta $\delta_{i,k,l}$ is equal to one, and the pumping head must be added. The open/closed status of the pump is later on introduced in the formulation through topological hydraulic constraints. Gate valves and check valves are also considered in this paper, but they are not included in Eq. (3) because we assume that there is no headloss through them, as it will be presented afterwards.

The continuity equation can be written as:

$$q_i = -\sum_{\forall j \in \Omega_i^{\mathrm{I}}} Q_{i,j} + \sum_{\forall j \in \Omega_i^{\mathrm{O}}} Q_{i,j};\ \forall i \in (\mathcal{V}^{\mathrm{Q}} \cup \mathcal{V}^{\mathrm{T}}), \tag{4}$$

where q_i is negative at demand nodes ($\forall i \in \mathcal{V}^{\mathrm{Q}}$, being $\mathcal{V}^{\mathrm{Q}}$ the subset of demand nodes) and null at the so called transit nodes ($\forall i \in \mathcal{V}^{\mathrm{T}}$, being $\mathcal{V}^{\mathrm{T}}$ the subset of transit nodes). Note that Eq. (4) is only applied at demand and transit nodes, because in a pseudo-static estimation tanks and reservoirs can be represented by only their head level. An additional condition is in fact that water demand at transit nodes is zero:

$$q_i = 0; \forall i \in \mathcal{V}^{\mathrm{T}}. \tag{5}$$

Physical limits must also be defined for pipe and node elements as follows:

$$Q_{i,j} \le Q_{i,j}^{max}; \forall i, j \in \mathcal{L}^{\mathrm{PI}} \tag{6}$$

$$-Q_{i,j} \le Q_{i,j}^{max}; \forall i, j \in \mathcal{L}^{\mathrm{PI}} \tag{7}$$

$$h_i \le h_i^{max}; \forall i \in \mathcal{V} \tag{8}$$

$$h_i \ge h_i^{min}; \forall i \in \mathcal{V}, \tag{9}$$

where Eqs. (6)-(7) impose a maximum $Q_{i,j}^{max}$ for the absolute value of the water flow through pipes, and Eqs. (8)-(9) establish a maximum (h_i^{max}) and minimum (h_i^{min}) value for head levels, respectively.

Topological hydraulic constraints: pumps

Specific expressions must also be derived for pump elements ($\forall i, j \in \mathcal{L}^{\mathrm{P}}$), which are here treated as link elements of zero length whose end nodes are at the same elevation. These elements can be either open or closed, and the on/off (i.e. 1/0) status of each pump is taken into account by introducing a binary variable for each existing pump $b_{i,j}^{P}; \forall i, j \in \mathcal{L}^{\mathrm{P}}$. To derive the hydraulic constraints of such elements, two situations must be distinguished beforehand. Pumps only admit a predefined unidirectional flow (i.e. there is no possibility of having reverse flow through a pump), but this unidirectional flow can either be of positive or negative sign for a specific pump according to the aforementioned sign criterion. Therefore, we need to differentiate two subsets $\forall i, j \in \mathcal{L}_{+}^{\mathrm{P}}$ and $\forall i, j \in \mathcal{L}_{-}^{\mathrm{P}}$ in order to correctly simulate that a particular device is pumping water in the positive or negative direction of flow, respectively. Figure 1 shows that there is a $h^P_{i,j}$ increase of pressure at node j when the device is located to pump water in the positive direction of flow (from i to j), but this additional energy is applied to the initial node i in devices that pump the water in the negative direction (from j to i). In addition to this classification, two types of pumps are differentiated in this work attending to the available information about the pump: pumps with known characteristic curve ($\forall i, j \in \mathcal{L}^{\mathrm{P}}_{\mathrm{K}}$, being $\mathcal{L}^{\mathrm{P}}_{\mathrm{K}}$ the subset of pump links with known characteristic curve), and pumps with unknown characteristic curve ($\forall i, j \in \mathcal{L}^{\mathrm{P}}_{\mathrm{U}}$, being $\mathcal{L}^{\mathrm{P}}_{\mathrm{U}}$ the subset of pump links with unknown characteristic curve). This distinction is made because even though manufacturers normally provide the characteristic curve for each pump at the time of sale, some network operators know nothing about the characteristic curve of some pumps, for example if they are working on a longstanding and/or undocumented water system. As presented by Díaz et al. (2017b), when the characteristic curve of the pump is available, a known relationship exists between the pumping head and the circulating flow, thus the curve can be introduced as a constraint and no additional information is required to characterise the system. On the contrary, if the characteristic curve of the pump is unknown, an additional unknown is added to the problem, and thus more measurements are needed to guarantee the system observability.

According to this classification, two independent pairs of subsets exist for the general pump set $\mathcal{L}^{\mathrm{P}}$: subsets $\mathcal{L}^{\mathrm{P}}_{+}$ and $\mathcal{L}^{\mathrm{P}}_{-}$ account for the flow direction through the pump, i.e. $\mathcal{L}^{\mathrm{P}} \equiv \mathcal{L}^{\mathrm{P}}_{+} \cup \mathcal{L}^{\mathrm{P}}_{-}$ and $\mathcal{L}^{\mathrm{P}}_{+} \cap \mathcal{L}^{\mathrm{P}}_{-} \equiv \emptyset$, whereas subsets $\mathcal{L}^{\mathrm{P}}_{\mathrm{K}}$ and $\mathcal{L}^{\mathrm{P}}_{\mathrm{U}}$ indicate the available information about the characteristic curve of the element, i.e. $\mathcal{L}^{\mathrm{P}} \equiv \mathcal{L}^{\mathrm{P}}_{\mathrm{K}} \cup \mathcal{L}^{\mathrm{P}}_{\mathrm{U}}$ and $\mathcal{L}^{\mathrm{P}}_{\mathrm{K}} \cap \mathcal{L}^{\mathrm{P}}_{\mathrm{U}} \equiv \emptyset$. Therefore, the hydraulic constraints related to pump elements can be defined as:

$$h^P_{i,j} \geq 0; \forall i, j \in \mathcal{L}^{\mathrm{P}} \tag{10}$$

$$-M(1 - b^P_{i,j}) \leq p_i - p_j \leq M(1 - b^P_{i,j}); \forall i, j \in \mathcal{L}^{\mathrm{P}} \tag{11}$$

$$Q_{i,j} \leq Q^{max}_{i,j} b^P_{i,j}; \forall i, j \in \mathcal{L}^{\mathrm{P}}_{+} \tag{12}$$

$$Q_{i,j} \geq Q^{min}_{i,j} b^P_{i,j}; \forall i, j \in \mathcal{L}^{\mathrm{P}}_{+} \tag{13}$$

$$- Q_{i,j} \leq Q_{i,j}^{max} b_{i,j}^{P}; \forall i,j \in \mathcal{L}_{-}^{\mathrm{P}} \tag{14}$$

$$- Q_{i,j} \geq Q_{i,j}^{min} b_{i,j}^{P}; \forall i,j \in \mathcal{L}_{-}^{\mathrm{P}} \tag{15}$$

$$h_{i,j}^{P} = \left(A_{i,j} (Q_{i,j})^2 + B_{i,j} |Q_{i,j}| + C_{i,j} \right) b_{i,j}^{P}; \forall i,j \in \mathcal{L}_{\mathrm{K}}^{\mathrm{P}} \tag{16}$$

$$h_{i,j}^{P} \leq h_{i,j}^{P,max} b_{i,j}^{P}; \forall i,j \in \mathcal{L}_{\mathrm{U}}^{\mathrm{P}}, \tag{17}$$

where $A_{i,j}, B_{i,j}, C_{i,j}; \forall i,j \in \mathcal{L}_{\mathrm{K}}^{\mathrm{P}}$ are the parameters of the characteristic curve of each pump, M is a large enough positive constant, and $b_{i,j}^{P}$ is the binary variable at pump link from node i to node j. Eq. (10) forces the pumping head to be positive no matter the scenario, because a pump always supplies energy to the system regardless of the injection point (see Figure 1). Constraint (11) forces the pressure at each side of the pump to be the same when the pump is open ($b_{i,j}^{P} = 1$), but it permits a hydraulic disconnection between nodes i and j when the pump is closed ($b_{i,j}^{P} = 0$), i.e. p_i and p_j can be significantly different when there is no flow through the pump. Hence, constant M must be large enough to model the difference in pressure head between nodes. Eqs. (12)-(13) determine the maximum ($Q_{i,j}^{max}$) and minimum ($Q_{i,j}^{min}$) flow through pumps in which the water moves from node i to node j, and Eqs. (14)-(15) are analogous for elements with negative flow. Constraint (16) provides the characteristic curves when they are available, forcing the pumping head to be null when the pump is closed. Similarly, Eq. (17) enables the pumping head to reach a pre-set maximum value $h_{i,j}^{P,max}; \forall i,j \in \mathcal{L}_{\mathrm{U}}^{\mathrm{P}}$ when information about the characteristic curve is not available, and fixes the pumping head to zero when the pump is not working.

Topological hydraulic constraints: gate valves

Gate valves enable isolation of different segments and/or areas within a given water distribution system. These controlling devices only have two positions, which are represented in this work by binary variables $b_{i,j}^{GV}; \forall i,j \in \mathcal{L}^{\mathrm{GV}}$, i.e. $b_{i,j}^{GV} = 1$ when they are open, and $b_{i,j}^{GV} = 0$ when they are closed, being $\mathcal{L}^{\mathrm{GV}}$ the subset of gate valve links. As assumed with the pumps, valves are introduced in the hydraulic model as elements of zero length whose nodes keep the same elevation. In this particular case, we assume that there is no headloss through such valves when they are fully open.

Hence, the only constraints are:

$$-M(1-b_{i,j}^{GV}) \leq h_i - h_j \leq M(1-b_{i,j}^{GV}); \forall i,j \in \mathcal{L}^{\mathrm{GV}} \tag{18}$$

$$Q_{i,j} \leq Q_{i,j}^{max} b_{i,j}^{GV}; \forall i,j \in \mathcal{L}^{\mathrm{GV}} \tag{19}$$

$$-Q_{i,j} \leq Q_{i,j}^{max} b_{i,j}^{GV}; \forall i,j \in \mathcal{L}^{\mathrm{GV}}. \tag{20}$$

Eq. (18) ensures that the energy gradient is zero when the valve is open (i.e. no energy losses), and enables h_i and h_j to differ when the valve is closed. This expression is analogous to Eq. (11) for pumps, because as the elevation at end nodes is the same in gate valves and no losses are assumed $h_i - h_j = p_i - p_j$. Analogously to the pump case, constant M must be large enough to model the difference in pressure head between nodes. Eqs. (19)-(20) establish that the absolute value of flow through the valve must be lower than the maximum allowed in the network. Hence, if the valve is closed, $b_{i,j}^{GV} = 0$ and the flow through the pipe is zero.

Topological hydraulic constraints: check valves

Check valves are placed in water distribution systems with the aim of avoiding inverse flows (Deuerlein et al. 2009). Consequently, they are directional devices that require the allowed direction of flow to be defined beforehand. As with pumps, the general set for check valves ($\forall i,j \in \mathcal{L}^{\mathrm{CV}}$) must then be divided in two subsets: $\mathcal{L}_{+}^{\mathrm{CV}}$ when the water is only allowed to move from the lower to the higher numbering node, and $\mathcal{L}_{-}^{\mathrm{CV}}$ if only negative flow is allowed, with $\mathcal{L}^{\mathrm{CV}} \equiv \mathcal{L}_{+}^{\mathrm{CV}} \cup \mathcal{L}_{-}^{\mathrm{CV}}$ and $\mathcal{L}_{+}^{\mathrm{CV}} \cap \mathcal{L}_{-}^{\mathrm{CV}} \equiv \oslash$. The following constraints are required to simulate the status $b_{i,j}^{CV}; \forall i,j \in \mathcal{L}^{\mathrm{CV}}$ of such controlling devices:

$$-M(1-b_{i,j}^{CV}) \leq h_i - h_j \leq M(1-b_{i,j}^{CV}); \forall i,j \in \mathcal{L}^{\mathrm{CV}} \tag{21}$$

$$Q_{i,j} \leq Q_{i,j}^{max} b_{i,j}^{CV}; \forall i,j \in \mathcal{L}_{+}^{\mathrm{CV}} \tag{22}$$

$$-Q_{i,j} \leq Q_{i,j}^{max} b_{i,j}^{CV}; \forall i,j \in \mathcal{L}_{-}^{\mathrm{CV}} \tag{23}$$

$$Q_{i,j} \geq 0; \forall i, j \in \mathcal{L}_{+}^{\mathrm{CV}} \tag{24}$$

$$-Q_{i,j} \geq 0; \forall i, j \in \mathcal{L}_{-}^{\mathrm{CV}}. \tag{25}$$

Eqs. (21)-(23) are equivalent to (18)-(20) for gate valves, i.e. it is assumed that no energy loss occurs when the valve is open and maximum values of flow are established. Eqs. (24)-(25) force the water to move in the positive and negative direction, respectively, and impose null flow otherwise.

Solution methodology

Eqs. (1)-(25) define the MINLP problem required to implement TSE in any water distribution system. More specifically, Eqs. (1)-(9) set out the traditional non-linear programming (NLP) problem for SE, which is then complemented with hydraulic constraints that simulate the presence of pumps, gate valves and check valves. Therefore, if the status of pumps and valves is known beforehand thanks to a good monitoring or the installation of position sensors in controlling devices, binary variables can be considered fixed, and problem (1)-(25) can be treated as a NLP problem. However, as the network topology is not likely to be fully known in large systems, and the objective of TSE is to infer both the hydraulic state and the pump/valve status of the system from a set of available measurements, TSE must be generally addressed as a MINLP problem.

Developing robust solution methodologies for MINLP problems is a challenging issue at present, and their computational implementation is expected to experience a significant rise in the years to come (Berthold 2014). Even though various solvers are commercially available to solve MINLP problems, their suitability highly depends on each particular problem, especially if the problem is nonconvex (Bussieck and Vigerske 2011), as it is the case of TSE. As selecting a reliable and robust solver that works well in different networks is not straightforward, the original MINLP problem is transformed in this work into an iterative MIQP problem. The change from MINLP to MIQP requires to linearize the non-linear equations of problem (1)-(25), i.e. the headloss equation and the characteristic curves associated with pumps. Note that linearizing the Hazen-Williams equation is not a novel idea, and a literature review can be found in Eck and Mevissen (2012). For example, a piecewise linearization for Eqs. (2) and (16) could be adopted. However, this strategy implies a significant increase in the number of binary variables in order to accurately represent the curvature of such non-linear functions (Morsi et al. 2012). Alternatively, a successive linearization strategy is here adopted. It is true that linearization may lead to instabilities along the iterative process, but as the algorithm is conceived for on-line state estimation, it is likely to run with a pre-specified frequency. Therefore, major simultaneous changes are not expected from one state estimation to the next. This enables the state estimation problem to be initialised from the previous result, i.e. the linearization occurs around a point which is reasonably close to the estimated flow scenario. Once Eqs. (2) and (16) are linearized, the objective function remains as the only non-linearity of problem (1)-(25), thus the optimisation problem can be solved with a MIQP approach. Current state-of-the-art MIQP solvers, such as *CPLEX* (www-01.ibm.com/software/commerce/optimization/cplex-optimizer) or *Gurobi* (www.gurobi.com), can solve this type of problems very efficiently. In any case, the problem linearization comes at the cost of an iterative algorithm, which will be presented once the linear versions of the aforementioned equations are posed.

Eq. (2) can be linearized around an initial flow $Q_{0_{i,j}}; \forall i, j \in \mathcal{L}^{\mathrm{PI}}$ as follows:

$$h_i - h_j = (1-c) K_{i,j} |Q_{0_{i,j}}|^{c-1} Q_{0_{i,j}} + K_{i,j} c |Q_{0_{i,j}}|^{c-1} Q_{i,j}; \forall i, j \in \mathcal{L}^{\mathrm{PI}}. \tag{26}$$

Similarly, the characteristic curve of any pump (Eq. (16)) can be linearized with respect to $Q_{0_{i,j}}; \forall i, j \in \mathcal{L}^{\mathrm{P}}_{\mathrm{K}}$ as:

$$h^P_{i,j} = \left[C_{i,j} - A_{i,j} (Q_{0_{i,j}})^2 + (2A_{i,j}|Q_{0_{i,j}}| + B_{i,j}) \operatorname{sign}(Q_{0_{i,j}}) Q_{i,j} \right] b^P_{i,j}; \forall i, j \in \mathcal{L}^{\mathrm{P}}_{\mathrm{K}}, \tag{27}$$

however, as the characteristic curve is affected by the binary variable $b^P_{i,j}; \forall i, j \in \mathcal{L}^{\mathrm{P}}_{\mathrm{K}}$, the previous expression is still non-linear. In order to provide a linear alternative to Eq. (16), Eq. (27) can be linearized as follows:

$$h^{P,min}_{i,j} b^P_{i,j} \leq h^P_{i,j} \leq h^{P,max}_{i,j} b^P_{i,j}; \forall i, j \in \mathcal{L}^{\mathrm{P}}_{\mathrm{K}} \tag{28}$$

$$h_{i,j}^{P,min}(1-b_{i,j}^{P}) \leq C_{i,j}-A_{i,j}(Q_{0_{i,j}})^2+(2A_{i,j}|Q_{0_{i,j}}|+B_{i,j})\ \mathrm{sign}(Q_{0_{i,j}})Q_{i,j}-h_{i,j}^{P} \leq h_{i,j}^{P,max}(1-b_{i,j}^{P}); \forall i,j \in \mathcal{L}_{\mathrm{K}}^{\mathrm{P}}, \tag{29}$$

where $h_{i,j}^{P,min}; \forall i,j \in \mathcal{L}_{\mathrm{K}}^{\mathrm{P}}$ and $h_{i,j}^{P,max}; \forall i,j \in \mathcal{L}_{\mathrm{K}}^{\mathrm{P}}$ have been introduced as parameters, and represent the minimum and maximum pumping head at pump link from node i to node j. Hence, when the pump is closed ($b_{i,j}^{P} = 0$), Eq. (28) fixes the pumping head to zero, and when the pump is open ($b_{i,j}^{P} = 1$), Eq. (29) forces $h_{i,j}^{P}$ to the linearized characteristic curve, i.e. Eqs. (28)-(29) are equivalent to Eq. (27).

Therefore, the MINLP problem (1)-(25) can be formulated as the MIQP problem (1), (3)-(15), (17)-(26), (28)-(29), which incorporates the linearized Hazen-Williams headloss equations and characteristic curves (if known). In this paper, the *CPLEX* solver is used to solve such problem for a given initialisation. As state estimation techniques are normally fed with online telemetry data, the initial value for $Q_{0_{i,j}}; \forall i,j \in (\mathcal{L}^{\mathrm{PI}} \cup \mathcal{L}_{\mathrm{K}}^{\mathrm{P}})$ can be obtained from the previous TSE, or from any other flow scenario in the system. Then, the MIQP solution must be used as initialisation for the following iteration, enabling a linearization that is progressively closer to the optimal solution. In order to speed convergence and enhance stability, a sub-relaxation method is here adopted when the iteration number k exceeds a selected value k_{lim}:

$$\begin{cases} Q_{0_{i,j}}^{(k+1)} = Q_{i,j}^{(k)} & \text{if} \quad k \leq k_{lim} \\ Q_{0_{i,j}}^{(k+1)} = Q_{i,j}^{(k)}\lambda + (1-\lambda)Q_{0_{i,j}}^{(k)} & \text{if} \quad k_{lim} < k < k_{max}, \end{cases} \tag{30}$$

with $k_{lim} = 10$. The maximum number of iterations and the subrelaxation factor are set out as $k_{max} = 100$ and $\lambda = 0.3$, respectively, all along this work. The iterative algorithm (summarised in Figure 2) finishes when the relative error of the objective function between successive iterations becomes lower than a specified tolerance, which is considered 10^{-6} in this paper. Then, the algorithm can be formally written as follows:

Algorithm 1 *MIQP approach for TSE*

Input: *Measurements and standard deviation of measurements, model parameters, available infor-*

mation of pumps, flow direction in pumps and check valves, and $Q_{0_{i,j}}; \forall i,j \in (\mathcal{L}^{\mathrm{PI}} \cup \mathcal{L}^{\mathrm{P}}_{\mathrm{K}})$*, which can be obtained from a previous TSE result or from any other flow scenario in the system.*

Step 1: Solve MIQP problem. *The linearized MIQP problem (1), (3)-(15), (17)-(26), (28)-(29) is solved for* $Q_{0_{i,j}}; \forall i,j \in (\mathcal{L}^{\mathrm{PI}} \cup \mathcal{L}^{\mathrm{P}}_{\mathrm{K}})$.

Step 2: Check tolerance. *If the relative error of the objective function is lower than* 10^{-6} *continue with **step 3**, otherwise, update the iteration counter* $k \rightarrow k+1$*, update* $Q_{0_{i,j}}; \forall i,j \in (\mathcal{L}^{\mathrm{PI}} \cup \mathcal{L}^{\mathrm{P}}_{\mathrm{K}})$ *according to Eq. (30), and continue with **step 1**.*

Step 3: Output. *Estimated head levels at nodes* $\hat{h}_i; \forall i \in \mathcal{V}$*, water flows through link elements* $\hat{Q}_{i,j}; \forall i,j \in \mathcal{L}$ *with* $\mathcal{L} = \{\mathcal{L}^{\mathrm{PI}}, \mathcal{L}^{\mathrm{P}}, \mathcal{L}^{\mathrm{GV}}, \mathcal{L}^{\mathrm{CV}}\}$*, demands at junction nodes* $\hat{q}_i; \forall i \in \mathcal{V}^{\mathrm{Q}}$*, and binary variables for pumps* $\hat{b}^{P}_{i,j}; \forall i,j \in \mathcal{L}^{\mathrm{P}}$*, gate valves* $\hat{b}^{GV}_{i,j}; \forall i,j \in \mathcal{L}^{\mathrm{GV}}$ *and check valves* $\hat{b}^{CV}_{i,j}; \forall i,j \in \mathcal{L}^{\mathrm{CV}}$.

This algorithm at present is heuristic, since it cannot be assured that a global solution of the original MINLP problem has been achieved. Nevertheless, the algorithm proposed in the paper presents good convergence properties and, as it will be shown in the following examples, it is computationally efficient. As opposed to other approximate methods lacking rigorous grounding, such as genetic algorithms or simulated annealing, the proposed solution approach is a mathematically sound heuristic based on the application of a well-known optimization technique.

ILLUSTRATIVE EXAMPLE

The small water system presented by Díaz et al. (2017b) is used as an illustration in this paper. Figure 3 provides the network layout, in which 6 junctions and 2 reservoirs are connected to each other through 7 pipes, a gate valve and a pump. TSE is here undertaken considering two different scenarios in what regards the available information about the pump: (1) the pump has a characteristic curve defined by $A_{6,8} = -2.2204 \cdot 10^{-16}\ \mathrm{h^2/m^5}$, $B_{6,8} = -0.3126\ \mathrm{h/m^2}$ and $C_{6,8} = 125.2806$ m, and (2) there is no information whatsoever about the characteristic curve of the pump. Also, we assume that water levels at both reservoirs are metered, water demands at all nodes are pseudomeasured, and a number of flow meters exist. As shown in Figure 3, two settings of flow meters are considered in this example for the sake of comparison: a first set of three flow meters located in pipes 1-2, 3-7 and 2-5, and a second set of four flow meters located in pipes 1-2, 3-7, 2-5 and 4-8. It must be highlighted that all measurement configurations have been previously assessed with a topological observability analysis (Díaz et al. 2017b), which is a prior necessary step to TSE. Such an analysis enables to guarantee that there are enough algebraic relationships to infer the hydraulic variables and the pump and valve statuses from the available measurements.

Since the method is heuristic, we have to check its performance under a controlled setting, and for this reason, measurements at each of such locations are synthetically generated in this work, i.e. sampling experiments are undertaken. With this purpose, a known pump/valve status is assumed (e.g. both the valve and the pump are considered opened), and the flow network solution is computed. Then, 1000 measured values are synthetically generated for each of the available measurements by considering a random variable distribution with mean equal to the corresponding flow, head or demand value obtained from solving the flow network, and a standard deviation. It must be highlighted that even though we assume a network topology to generate the measurements, pump/valve binary variables are freed before running the TSE algorithm in order to test the ability of the method to correctly determine the network topology. The standard deviations assumed for each type of measurement are $\sigma_i^h = 0.01$ m; $\forall i \in \mathcal{V}_\mathrm{m}$, $\sigma_i^q = 20\%\,\overline{q}_i; \forall i \in \mathcal{V}_\mathrm{m}^\mathrm{Q}$ and $\sigma_{i,j}^Q = \sqrt{\frac{1}{6} + (2\%\overline{Q}_{i,j})^2}; \forall i,j \in \mathcal{L}_\mathrm{m}^\mathrm{PI}$, where $\overline{q}_i$ and $\overline{Q}_{i,j}$ represent the value of demand and flow variables in the flow network solution. It is worth mentioning that a greater uncertainty is introduced in each demand because we assume that water consumptions are pseudomeasured, i.e. estimated from historic data, as it may be the case in a real water system (Bargiela and Hainsworth 1989). On the other hand, $\sigma_{i,j}^Q$ is constituted by a fixed uncertainty of $\frac{1}{\sqrt{6}}$ that comes from considering that flow is metered as volume in m^3 and the difference in volume over time provides a measurement of flow, and a variable term that depends on the water flow itself, as it normally occurs in real flow meters. Once measurements have been computed, Monte Carlo simulations are undertaken: the TSE problem is solved considering each of the 1000 artificially generated measurements/pseudomeasurements. The process is repeated for all different combinations of pump and valve status: open valve-open pump, open valve-closed pump, closed valve-open pump, and closed valve-closed pump. In all of these cases the algorithm for TSE described in Figure 2 is initialised $Q_{0_{i,j}}; \forall i, j \in (\mathcal{L}^{\mathrm{PI}} \cup \mathcal{L}^{\mathrm{P}}_{\mathrm{K}})$ from the flow network solution associated with an open valve and pump, which is here subjected to a 5% noise to avoid straightforward solutions when open valve-open pump is the real network topology. This permits to evaluate to what extent the methodology presented in this paper permits to detect changes in the network topology. Note that synthetic measurements are here used to test the algorithm, or otherwise you need a real network with a huge set of metering devices in order to use some of them for validation.

Now, TSE results considering the characteristic curve of the pump known and unknown are presented. We assume $Q^{max}_{i,j} = 10$ m/s $\cdot S_{i,j}; \forall i, j \in \mathcal{L}$, $h^{max}_i = \max(h^{max}_i) + 200$ m; $\forall i \in \mathcal{V}$, $h^{min}_i = -h^{max}_i; \forall i \in \mathcal{V}$, $M = 300$, $Q^{min}_{i,j} = 0.01$ m/s $\cdot S_{i,j}; \forall i, j \in \mathcal{L}^{\mathrm{P}}$, $h^{P,max}_{i,j} = 200$ m; $\forall i, j \in \mathcal{L}^{\mathrm{P}}$ and $h^{P,min}_{i,j} = 0$ m; $\forall i, j \in \mathcal{L}^{\mathrm{P}}_{\mathrm{K}}$ all along this paper, where $S_{i,j}$ is the pipe cross-sectional area.

Known characteristic curve

Table 1 provides the TSE results of Monte Carlo simulations when the characteristic curve of the pump is known. For each of the four network topologies assumed, the percentage of success in estimating the pump status (S_P), the percentage of success in estimating the gate valve status (S_{GV}), the average number of iterations required for the TSE algorithm to converge ($\overline{k}$), and the maximum number of iterations required to achieve convergence ($\max(k)$) are provided for the two measurement settings in which 3 and 4 flow meters exist. Additionally, this table gives the average of the mean squared error in terms of flows ($\overline{MSE}$) for the 1000 simulations, and the time needed by the Monte Carlo method to converge in an Intel(R) Core(TM) i7-6700 CPU 3.40 GHz 16GB RAM desktop computer. The mean squared error for each of the measurement configurations considered for the analysis can be computed as:

$$MSE = \frac{1}{n_L} \sum_{\forall i,j \in \mathcal{L}} (\hat{Q}_{i,j} - \overline{Q}_{i,j})^2, \qquad (31)$$

with $n_L = 9$ link elements in the illustrative network. Eq. (31) provides the mean squared difference between the estimated flows ($\hat{Q}_{i,j}; \forall i, j \in \mathcal{L}$) and the real values in the system, which correspond to their values in the corresponding flow network solution $\overline{Q}_{i,j}; \forall i, j \in \mathcal{L}$. Therefore, its average over 1000 simulations provides an insight of how much TSE results differ from the real network state.

Beginning with the results obtained for the 3 flow meters measurement configuration, the success in estimating the gate valve status (S_{GV}) is 100% no matter the real network state. This is fully expected because of the fact that all inflows and outflows to the valve are metered. Nevertheless, S_P ranges between 61.8 and 92.0% depending on the network topology. This can be explained by the fact that only water level meters and demand pseudomeasurements are available near the pump location, and this does not seem to be enough to correctly estimate the pump status in all cases. As presented by Díaz et al. (2017b), measurement noise may lead to an incorrect status determination during TSE even if the variable has been identified as observable in a prior observability analysis. This is due to the fact that this analysis only considers the number and location of meters, but measurement uncertainty can only be taken into account through TSE itself. We want to highlight that this is not a limitation of the TSE algorithm, as it provides the best possible estimation by taking into account the available noisy measurements. It has been verified that if measurements were exact, a 100% success would be obtained for both the gate valve and the pump in all cases. However, real water systems are subjected to noise, and this affects the capability of the algorithm to detect changes in the pump/valve status. This reality can be counteracted with the addition of metering devices. Table 1 shows that S_P increases to 100% in all scenarios when an additional flow meter is located in pipe 4-8, i.e. this measurement configuration would permit to guarantee that the pump and valve status could be correctly detected regardless of the high uncertainty of the measurements and pseudomeasurements, where uncertainty is particularly important. Consequently, $\overline{MSE}$ is reduced by one order of magnitude when a fourth flow meter is added.

In what regards the number of iterations, $\overline{k}$ varies between 2 and 7 depending on the original network state. Note that when both devices are closed, the method converges in few iterations, as it is easy for the algorithm to detect that the pump is closed if the characteristic pump is available. Finally, it must be highlighted that the mean convergence time required for each simulation to converge is between 2 and 3 seconds in all cases.

Unknown characteristic curve

Table 2 provides the same results when the characteristic curve of the pump is unknown. As it happened before, $S_{GV} = 100\%$ in all cases no matter the number of flow meters installed. On the contrary, S_P changes depending on the real network state. As information about the characteristic curve is no longer available, it is expected that the success in the estimation of the correct pump status reduces with respect to the previous scenario. Such is the case of open valve-closed pump and closed valve-closed pump cases when either 3 or 4 flow meters exist. Nevertheless, S_P values in Table 2 are improved with respect to those in Table 1 in the open valve-open pump and closed valve-open pump cases when three flow meters are adopted. This result is not intuitive, but can be explained: by removing the information about the characteristic curve of the pump, we allow the TSE algorithm to estimate the head gain and flow through the pump freely, i.e. no specific relationship is required between $h^P_{6,8}$ and $Q_{6,8}$ and they are both independently adjusted to minimise the objective function. Therefore, the method is correctly identifying the pump/valve status at the cost of failing to accurately estimate the rest of the hydraulic variables in the system, which are no longer subjected to a $h-Q$ constraint in the proximity of the pump. Consequently, $\overline{MSE}$ values are increased from Table 1 to Table 2, as the estimated flows and the real water flows differ considerably regardless of the improvement in the network topology estimation.

In what regards the number of iterations, the loss of information about the characteristic curve of the pump negatively affects the speed of the method in this example. Table 2 shows that even $\max(k) = k_{max} = 100$ is reached in some cases, which considerably burdens $\overline{k}$. More specifically, 100 iterations have been achieved in 3% of the measurement configurations of the Monte Carlo simulation in the open valve-open pump scenario, and this percentage changes to 3.3% and 1.5% in the open valve-closed pump and closed valve-closed pump cases, respectively. In most of these scenarios, the reason why the algorithm achieves k_{max} is that there is an oscillation between $b^P_{6,8} = 0$ and $b^P_{6,8} = 1$ along the iterative process. The relative error of the objective function between these two possibilities is in all cases lower than 10%, i.e. the noise of the measurements is such that there is a negligible difference between considering the pump open or closed. This phenomenon disappears when a fourth flow meter is added, thus it can be concluded that even though the system is observable when there is no information about the characteristic curve of the pump, the algorithm may find it difficult to converge if there are few metering devices subjected to a significant noise. Such limitation comes from the nature of the measurements rather than from the method itself. Regardless of this situation, the mean time for each TSE remains between 2 and 3.5 s.

C-TOWN CASE STUDY

In this section, the algorithm for TSE is tested in the well-known C-Town case study, firstly presented in the so called "Battle of Background Leakage Assessment for water networks" (Giustolisi et al. 2014). This system presents 1 reservoir, 7 tanks, 388 nodes, 432 pipes, 11 pumps (grouped in 5 pumping stations S1, S2, S3, S4 and S5), 1 gate valve, and 1 check valve. In this work, we consider the steady state of the network, neglecting the demand patterns and controls provided when the problem was first posed for leakage detection purposes.

This water system is here tested under two different assumptions for TSE purposes: (1) information about the characteristic curves of the 11 existing pumps is known, and (2) there is no information about the characteristic curve of any of them. In each of them, three measurement settings are considered. Firstly, we assume that only water levels are metered and demands at junction nodes are pseudomeasured. Secondly, flow meters are added at four of the eight pipes that come out of the tanks and the reservoir in the system, more specifically, at the ones associated with the highest flows. Finally, four additional flow meters (eight in total) are included in the rest of entrances to the network. For each of these measurement settings, we synthetically generate measurements as in the illustrative example, assuming the same standard deviations. In this case, only two network topologies are considered for the artificial generation of measurements in each of the settings: firstly all pumps are working, the gate valve is open and the check valve is closed to avoid inverse flow, and secondly pumps PU1, PU4, PU6 and PU10 are closed, i.e. one pump is closed in pumping stations S1, S2, S3 and S5. As before, the TSE algorithm is initialised from the flow network solution associated with everything opened subjected to a 5% noise. Moreover, a Monte Carlo simulation of 1000 measurement configurations is applied to each of such topologies and measurement settings. We assume the same model parameters as the illustrative example.

Known characteristic curves

Table 3 shows TSE results for the network topologies and measurement settings considered when the characteristic curve of all pumps is known. This table provides the percentage of success in determining the number of pumps working in each pumping station ($S_{P_{Si}}; \forall i = 1,2,3,4,5$), the percentage of success in determining the status of the gate valve (S_{GV}) and the check valve (S_{CV}), the mean ($\overline{k}$) and maximum (max(k)) number of iterations required for the algorithm to converge, the mean squared error of flows over the 1000 simulations ($\overline{MSE}$), and the time required to undertake the sampling experiment in the aforementioned desktop computer. The success criterion for pumps is now to correctly determine the number of pumps working per pumping station rather than correctly estimating the status of each individual pump, because all the pumps in each pumping station present the same characteristic curve.

This table shows that the status of the pumps and valves in the system cannot be determined with certainty when only water levels at reservoirs and water consumptions are measured. Even though the system is observable under such configuration, measurement inaccuracy does not enable a good TSE. Status estimation is specially bad for $S_{P_{Si}}; \forall i = 1,2,3,4,5$, whereas S_{GV} and S_{CV} remain above 80% no matter the real status of the system. S_{GV} and S_{CV} present the same values when all pumps are working and when some of them are not, because the same water level and demand measurements apply. The situation considerably improves when 4 or 8 flow meters are added. In both cases, the number of working pumps and the status of the valves are inferred correctly in all Monte Carlo simulations. As before, $\overline{MSE}$ is considerably reduced with the addition of measurements, i.e. a better estimation can be obtained when metering devices are added. This improvement is higher when four flow meters are added, and lower for the next four extra flow meters. This fact highlights the importance of having redundant measurements, i.e. more measurements than the strictly required to make the system observable, to counteract measurement inaccuracy.

On the other hand, it is worth mentioning that $\overline{k}$ and $\max(k)$ are not significantly increased with respect to the illustrative example network, and the average time for each TSE remains 3-4 s, i.e.

the algorithm has potential for real time implementation. In this regard, we want to highlight that the aim of this work is to present a methodology that enables to infer both the network topology and hydraulic state of the system from the available measurements, but additional issues must be explored with view to its on-line implementation. For example, hydraulic uncertainty is unavoidable in real-time scenario, thus the hydraulic model needs to be periodically updated in order for TSE to be reliable. Recently, Díaz et al. (2017a) have proposed a calibration method to adjust model parameters based on multi-period state estimation results. Also, demand estimation in real-time is challenging since most large networks have limited data that can be used to estimate demands. This is the main reason why state estimation techniques have not been systematically applied to water systems on an operational level yet, and the only real applications are all related to water transport networks (González et al. 2017; Vrachimis et al. 2016). Water transport networks are pipeline systems that provide water to distinct District Metered Areas (DMA), where incoming flows are normally monitored. Hence, in this type of systems demand uncertainty can be considerably reduced. Further research should apply the methodology here proposed to a real system, as there are always aspects that the model has not reflected and that may introduce errors in the determination of the states of the elements. Additionally, sensitivity analysis, leak detection strategies and uncertainty evaluation should be progressively incorporated to TSE in order to address the on-line monitoring issue in all its complexity (Díaz 2017). In this regard, explicit expressions for state estimation sensitivity analysis have been recently proposed (Díaz et al. 2017c) to further the understanding of state estimation solutions.

Unknown characteristic curves

Table 4 provides the same results when the characteristic curves of the pumps in the system are considered unknown. It must be highlighted that such assumption leads to an unobservable system for all scenarios, and this implies that state estimation results cannot be trusted (Díaz et al. 2017b). It is worth mentioning that some high percentages of S_P, S_{GV} and S_{CV} are obtained in some of the configurations, but they must not be relied upon, as they are mere coincidence. This justifies the inconsistencies between Tables 3 and 4 in terms of $\overline{MSE}$. These results highlight the importance of observability analysis techniques, without which we would just believe the results of the TSE algorithm gathered in Table 4. For this reason, topological observability analysis is a prior necessary step that enables to pre-assess the suitability of measurement configurations.

To finish with, we want to highlight that these last scenarios in which no information about the characteristic curves of pumps is available, few meters exist, and all pumps and valves statuses are to be determined, are not consistent with the reality of current water systems. For example, system outflows are metered rather than pseudomeasured in actual water transport networks that provide water to DMAs. Moreover, the characteristic curve of most pumps is normally known, and even many of the statuses of the pumps and valves in the network are known thanks to the existence of position sensors in many controlling devices. This implies that their binary variables can be fixed and only some unknown statuses need to be inferred, thus reducing the complexity of the original MINLP problem. The aim of these examples is to show that the methodology proposed in this paper is robust and can be used in conjunction with observability analysis techniques to assess the behaviour of the system in real-time even when changes in the network topology take place.

CONCLUSIONS

In this work, the importance of undertaking TSE rather than traditional SE is highlighted, and a novel methodology for TSE is presented. Implementing TSE rather than traditional SE permits to drop the assumption that the network topology is known beforehand, enabling to infer both the pump and valve status and the hydraulic state of the system from available measurements. This is of utmost importance at present, as changes in the network topology may take place in order to improve the quality of the supply service and the system reliability. TSE complements available techniques for topological observability analysis, as it permits to take into account measurement noise when estimating pump/valve status rather than only analysing if sufficient algebraic relationships exist.

The method for TSE has been set out as a MINLP problem, where the presence of pumps and valves is simulated through binary variables. Due to the complexity of developing a robust method for solving such a challenging problem, it has been here transformed into an iterative MIQP problem by linearizing Hazen-Williams headloss equations and the characteristic curves of pumps.

Only pumps, gate valves and check valves have been introduced in this work as controlling devices, but additional hydraulic constraints could be similarly added to the optimisation problem in order to simulate the presence of other types of valves.

The proposed methodology has proven to work robustly in conjunction with topological observability analysis in an illustrative example and a larger case study. Different hypotheses have been tested about the information available in the network and the number of metering devices, showing that the approach is versatile and can track changes in the topology even when there is no information about the characteristic curve of an individual pump or group of pumps. The computational cost of the new algorithm proves that it has potential for on-line implementation, although several related issues must be addressed before its real-time application (e.g. hydraulic uncertainty, sensitivity analysis). Additionally, it could constitute a powerful tool when used to train network operators on how to respond to incidences. Furthermore, it could be used as a basis on which other methods for optimal meter placement could be based, not only assessing the best location for additional devices, but also taking into account the possibility of installing position sensors, which would fix the binary variables thus reducing the complexity of the problem. These are subjects for further research.

REFERENCES

Andersen, J. H. and Powell, R. S. (2000). "Implicit state-estimation technique for water network monitoring." *Urban Water*, 2(2), 123–130.

Andersen, J. H., Powell, R. S., and Marsh, J. F. (2001). "Constrained state estimation with applications in water distribution network monitoring." *Int. J. Syst. Sci.*, 32(6), 807–816.

Bargiela, A. (1984). "On-line monitoring of water distribution networks." Ph.D. thesis, Univ. of Durham, UK.

Bargiela, A. and Hainsworth, G. D. (1989). "Pressure and flow uncertainty in water systems." *J. Water Resour. Plann. Manage.*, 115(2), 212–229.

Berthold, T. (2014). "Heuristic algorithms in global MINLP solvers." Ph.D. thesis, Technical Univ. of Berlin, Germany.

Brdys, M. A. and Chen, K. (1993). "Joint state and parameter estimation of dynamic water supply systems with unknown but bounded uncertainty." *Integrated Computer Applications in Water Supply*, B. Coulbeck, ed., Vol. 1, Research Studies Press, Baldock, UK, 335–355.

Brdys, M. A. and Ulanicki, B. (2002). *Operational control of water systems: Structures, algorithms and applications*. Prentice Hall, London, UK.

Bussieck, M. R. and Vigerske, S. (2011). "MINLP solver software." *Wiley Encyclopedia of Operations Research and Management Science*, John Wiley & Sons, Published online.

Carpentier, P. and Cohen, G. (1991). "State estimation and leak detection in water distribution networks." *Civ. Eng. Syst.*, 8(4), 247–257.

Coulbeck, B. (1977). "Optimisation and modelling techniques in dynamic control of water distribution systems." Ph.D. thesis, Univ. of Sheffield, UK.

Deuerlein, J. W., Simpson, A. R., and Dempe, S. (2009). "Modeling the behavior of flow regulating devices in water distribution systems using constrained nonlinear programming." *J. Hydraul. Eng.*, 135(11), 970–982.

Díaz, S. (2017). "Comprehensive approach for on-line monitoring water distribution systems via state estimation related techniques." Ph.D. thesis, Univ. of Castilla-La Mancha, Spain.

Díaz, S., González, J., and Mínguez, R. (2016). "Observability analysis in water transport networks: Algebraic approach." *J. Water Resour. Plann. Manage.*, 142(4), 04015071.

Díaz, S., González, J., and Mínguez, R. (2016a). "Uncertainty evaluation for constrained state estimation in water distribution systems." *J. Water Resour. Plann. Manage.*, 142(12), 06016004.

Díaz, S., Mínguez, R., and González, J. (2016b). "Stochastic approach to observability analysis in water networks." *Ingeniería del Agua*, 20(3), 139–152.

Díaz, S., Mínguez, R., and González, J. (2017a). "Calibration via multi-period state estimation in water distribution systems." *Water Resour. Manage.*, doi: 10.1007/s11269-017-1779-2.

Díaz, S., Mínguez, R., and González, J. (2017b). "Topological observability analysis in water distribution systems." *J. Water Resour. Plann. Manage.*, 143(5), 06017001.

Díaz, S., Mínguez, R., González, J. and Savic, D. (2017c). "Explicit expressions for state estimation sensitivity analysis in water systems." *J. Water Resour. Plann. Manage.*, doi: 10.1061/(ASCE)WR.1943-5452.0000914.

Eck, B. J. and Mevissen, M. (2012). "Valve placement in water networks: Mixed-integer non-linear optimization with quadratic pipe friction." *IBM Research Report RC25307, Dublin, Ireland*.

Gabrys, B. and Bargiela, A. (1996). "Integrated neural based system for state estimation and confidence limit analysis in water networks." *Proceedings of the European Simulation Symposium, Genoa, Italy*, 2, 398–402.

Giustolisi, O., Berardi, L., Laucelli, D., Savic, D., Walski, T., and Brunone, B. (2014). "Battle of background leakage assessment for water networks (BBLAWN) at WDSA Conference 2014." *Procedia Eng.*, 89, 4–12.

Giustolisi, O., Kapelan, Z., and Savic, D. (2008). "Algorithm for automatic detection of topological changes in water distribution networks." *J. Hydraul. Eng.*, 134(4), 435–446.

González, J., Mínguez, R., and Díaz, S. (2017). *Observability study for hydraulic state estimation of the sectorised water supply network. Booklet Research, Development and Innovation 23.* In edition. Canal de Isabel II Gestión, S.A., Madrid, Spain.

Goulter, I. (1995). "Analytical and simulation models for reliability analysis in water distribution systems." *Improving efficiency and reliability in water distribution systems*, E. Cabrera and A. F. Vela, eds., Kluwer Academics, London, UK, 235–266.

Irving, M. R., Owen, R. C., and Sterling, M. J. H. (1978). "Power-system state estimation using linear programming." *Proceedings of the Institution of Electrical Engineers*, 125(9), 879–885.

Kumar, S. M., Narasimhan, S., and Bhallamudi, S. M. (2008). "State estimation in water distribution networks using graph-theoretic reduction strategy." *J. Water Resour. Plann. Manage.*, 134(5), 395–403.

Lansey, K. E. and Basnet, C. (1991). "Parameter estimation for water distribution networks." *J. Water Resour. Plann. Manage.*, 117(1), 126–144.

Laucelli, D., Berardi, L., Giustolisi, O., Vamvakeridou-Lyroudia, L. S., Kapelan, Z., Savic, D., and Barbaroand, G. (2011). "Calibration of water distribution systems using topological analysis."

12th Annual Conference on Water Distribution Systems Analysis (WDSA), Tucson, USA, 1664–1681.

Liberatore, S. and Sechi, G. M. (2009). "Location and calibration of valves in water distribution networks using a scatter-search meta-heuristic approach." *Water Resour. Manage.*, 23(8), 1479–1495.

Morsi, A., Geissler, B., and Martin, A. (2012). "Mixed integer optimization of water supply networks." *Mathematical optimization of water networks*, A. Martin, K. Klamroth, J. Lang, G. Leugering, A. Morsi, M. Oberlack, M. Ostrowski, and R. Rosen, eds., Springer Basel, Basel, Switzerland, 35–54.

Nagar, A. K. and Powell, R. S. (2000). "Observability analysis of water distribution systems under parametric and measurement uncertainty." *Build. Partnerships*, 1–10.

Powell, R. S., Irving, M. R., and Sterling, M. J. H. (1988). "A comparison of three real-time state estimation methods for on-line monitoring of water distribution systems." *Computer Applications in Water Supply*, B. Coulbeck, ed., Vol. 1, Research Studies Press, Taunton, UK, 333–348.

Schweppe, F. C. and Wildes, J. (1970). "Power system static-state estimation, Part I: Exact model." *IEEE Trans. Power Appar. Syst.*, PAS-89(1), 120–125.

Sophocleous, S., Savic, D. A., Kapelan, Z., and Giustolisi, O. (2016). "A two-stage calibration for detection of leakage hotspots in a real water distribution network." *18th Water Distribution Systems Analysis, Cartagena, Colombia*.

Sterling, M. J. H. and Bargiela, A. (1984). "Minimum norm state estimation for computer control of water distribution systems." *IEE Proc.*, 131(2), 57–63.

Vrachimis, S. G., Eliades, D. G., and Polycarpou, M. M. (2016). "Real-time hydraulic interval state estimation for water transport networks: A case study." *14th International Computing and Control for the Water Industry (CCWI) Conference, Amsterdam, The Netherlands*.

Wright, R., Abraham, E., Parpas, P., and Stoianov, I. (2015). "Control of water distribution networks with dynamic DMA topology using strictly feasible sequential convex programming." *Water Resour. Res.*, 51(12), 9925–9941.

List of Tables

1 TSE results for 1000 Monte Carlo sampling experiments in the illustrative example network: known characteristic curve 29
2 TSE results for 1000 Monte Carlo sampling experiments in the illustrative example network: unknown characteristic curve 30
3 TSE results for 1000 Monte Carlo sampling experiments in C-Town case study: known characteristic curve 31
4 TSE results for 1000 Monte Carlo sampling experiments in C-Town case study: unknown characteristic curve 32

TABLE 1. TSE results for 1000 Monte Carlo sampling experiments in the illustrative example network: known characteristic curve

Case	Parameters	Measurement configuration: 3 flow meters	Measurement configuration: 4 flow meters
Open valve Open pump	S_P (%)	61.8	100
	S_{GV} (%)	100	100
	$\overline{k}$	6.0890	6.1530
	max(k)	26	44
	$\overline{MSE}$ (m^6/h^2)	1.1776e3	1.8999e2
	Time for 1000 simulations (s)	3.0442e3	2.6467e3
Open valve Closed pump	S_P (%)	73.9	100
	S_{GV} (%)	100	100
	$\overline{k}$	5.5500	5.2980
	max(k)	28	26
	$\overline{MSE}$ (m^6/h^2)	1.2199e3	1.9406e2
	Time for 1000 simulations (s)	2.5596e3	2.5725e3
Closed valve Open pump	S_P (%)	79.9	100
	S_{GV} (%)	100	100
	$\overline{k}$	3.6750	5.1060
	max(k)	6	7
	$\overline{MSE}$ (m^6/h^2)	2.1496e3	2.0453e2
	Time for 1000 simulations (s)	2.4007e3	2.5176e3
Closed valve Closed pump	S_P (%)	92.0	100
	S_{GV} (%)	100	100
	$\overline{k}$	2.2650	2.0130
	max(k)	5	5
	$\overline{MSE}$ (m^6/h^2)	2.0227e3	3.2767e2
	Time for 1000 simulations (s)	2.2927e3	2.2792e3

TABLE 2. TSE results for 1000 Monte Carlo sampling experiments in the illustrative example network: unknown characteristic curve

Case	Parameters	Measurement configuration: 3 flow meters	Measurement configuration: 4 flow meters
Open valve Open pump	S_P (%)	68.5	100
	S_{GV} (%)	100	100
	$\overline{k}$	13.6280	5.8170
	max(k)	100	28
	$\overline{MSE}$ (m^6/h^2)	2.3341e3	2.0352e2
	Time for 1000 simulations (s)	3.3303e3	2.6164e3
Open valve Closed pump	S_P (%)	64.9	98.8
	S_{GV} (%)	100	100
	$\overline{k}$	10.7860	5.3150
	max(k)	100	26
	$\overline{MSE}$ (m^6/h^2)	1.9212e3	1.9410e2
	Time for 1000 simulations (s)	3.1276e3	2.6472e3
Closed valve Open pump	S_P (%)	95.3	100
	S_{GV} (%)	100	100
	$\overline{k}$	4.8240	2.9730
	max(k)	9	5
	$\overline{MSE}$ (m^6/h^2)	2.6955e3	3.0500e2
	Time for 1000 simulations (s)	2.6066e3	2.3710e3
Closed valve Closed pump	S_P (%)	53.7	98.8
	S_{GV} (%)	100	100
	$\overline{k}$	6.5030	2.0420
	max(k)	100	5
	$\overline{MSE}$ (m^6/h^2)	2.0565e3	3.2768e2
	Time for 1000 simulations (s)	2.7773e3	2.3212e3

TABLE 3. TSE results for 1000 Monte Carlo sampling experiments in C-Town case study: known characteristic curve

Case	Parameters	Measurement configuration: no flow meters	Measurement configuration: 4 flow meters	Measurement configuration: 8 flow meters
Open GV Closed CV Open pumps	$S_{P_{S1}}$ (%)	0.0	100	100
	$S_{P_{S2}}$ (%)	0.4	100	100
	$S_{P_{S3}}$ (%)	0.0	100	100
	$S_{P_{S4}}$ (%)	0.0	100	100
	$S_{P_{S5}}$ (%)	0.1	100	100
	S_{GV} (%)	96.3	100	100
	S_{CV} (%)	83.0	100	100
	$\overline{k}$	2.9160	4.0830	3.9990
	max(k)	13	6	5
	$\overline{MSE}$ (m^6/h^2)	5.0407e4	2.2044	1.5199
	Time for 1000 simulations (s)	3.0164e3	3.2103e3	4.2936e3
Open GV Closed CV Some closed pumps	$S_{P_{S1}}$ (%)	0.3	100	100
	$S_{P_{S2}}$ (%)	7.8	100	100
	$S_{P_{S3}}$ (%)	0.1	100	100
	$S_{P_{S4}}$ (%)	0.0	100	100
	$S_{P_{S5}}$ (%)	3.1	100	100
	S_{GV} (%)	96.3	100	100
	S_{CV} (%)	83.0	100	100
	$\overline{k}$	2.9160	5.7030	5.0480
	max(k)	13	10	6
	$\overline{MSE}$ (m^6/h^2)	3.3290e4	2.3889	1.5028
	Time for 1000 simulations (s)	3.0233e3	3.5202e3	4.4427e3

TABLE 4. TSE results for 1000 Monte Carlo sampling experiments in C-Town case study: unknown characteristic curve

Case	Parameters	Measurement configuration: no flow meters	Measurement configuration: 4 flow meters	Measurement configuration: 8 flow meters
	$S_{P_{S1}}$ (%)	0.0	0.0	0.0
	$S_{P_{S2}}$ (%)	0.0	0.0	0.0
	$S_{P_{S3}}$ (%)	0.0	0.0	0.0
	$S_{P_{S4}}$ (%)	0.0	0.0	0.0
Open GV	$S_{P_{S5}}$ (%)	0.0	0.0	0.0
Closed CV	S_{GV} (%)	98.0	100	100
Open pumps	S_{CV} (%)	83.1	100	100
	$\overline{k}$	2.8220	12.8430	4.1860
	max(k)	14	100	15
	$\overline{MSE}$ (m^6/h^2)	4.8366e4	1.1997e3	6.4974
	Time for 1000 simulations (s)	2.9525e3	5.8960e3	3.4802e3
	$S_{P_{S1}}$ (%)	0.8	81.3	99.9
	$S_{P_{S2}}$ (%)	1.0	100	100
	$S_{P_{S3}}$ (%)	0.0	100	100
	$S_{P_{S4}}$ (%)	0.0	0.0	0.0
Open GV	$S_{P_{S5}}$ (%)	0.0	4.6	100
Closed CV	S_{GV} (%)	98.0	100	100
Some closed pumps	S_{CV} (%)	83.1	100	100
	$\overline{k}$	2.8220	2.8600	5.5760
	max(k)	14	14	17
	$\overline{MSE}$ (m^6/h^2)	3.1292e4	5.1755e2	6.3411
	Time for 1000 simulations (s)	2.9742e3	3.2190e3	3.8039e3

List of Figures

1 Scenarios within a pump element: positive and negative flow 34
2 Flow chart for TSE: Iterative linearization approach 35
3 Available flow meters at different measurement configurations 36

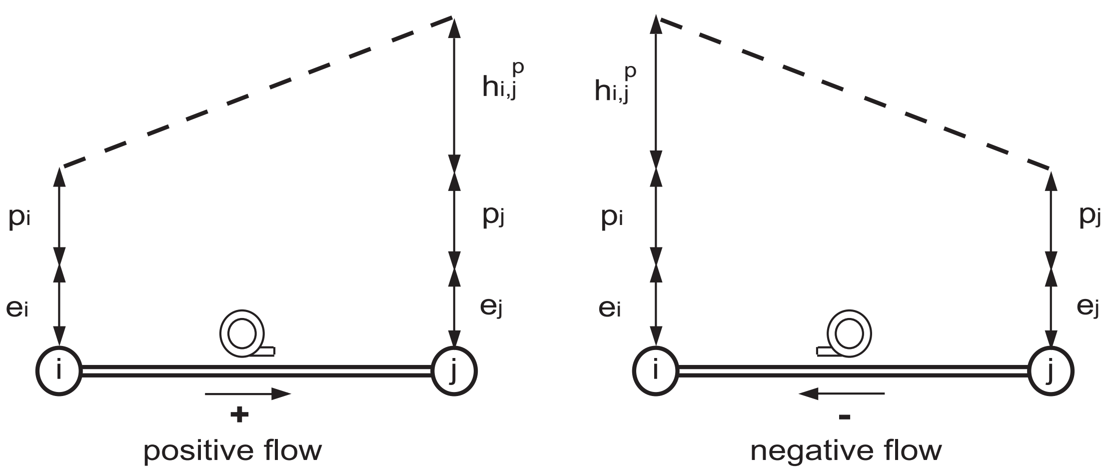

Fig. 1. Scenarios within a pump element: positive and negative flow

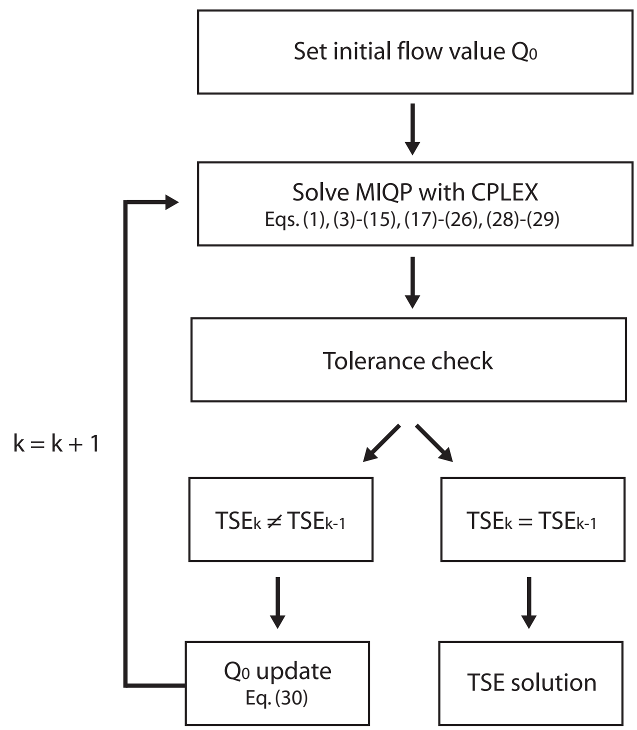

Fig. 2. Flow chart for TSE: Iterative linearization approach

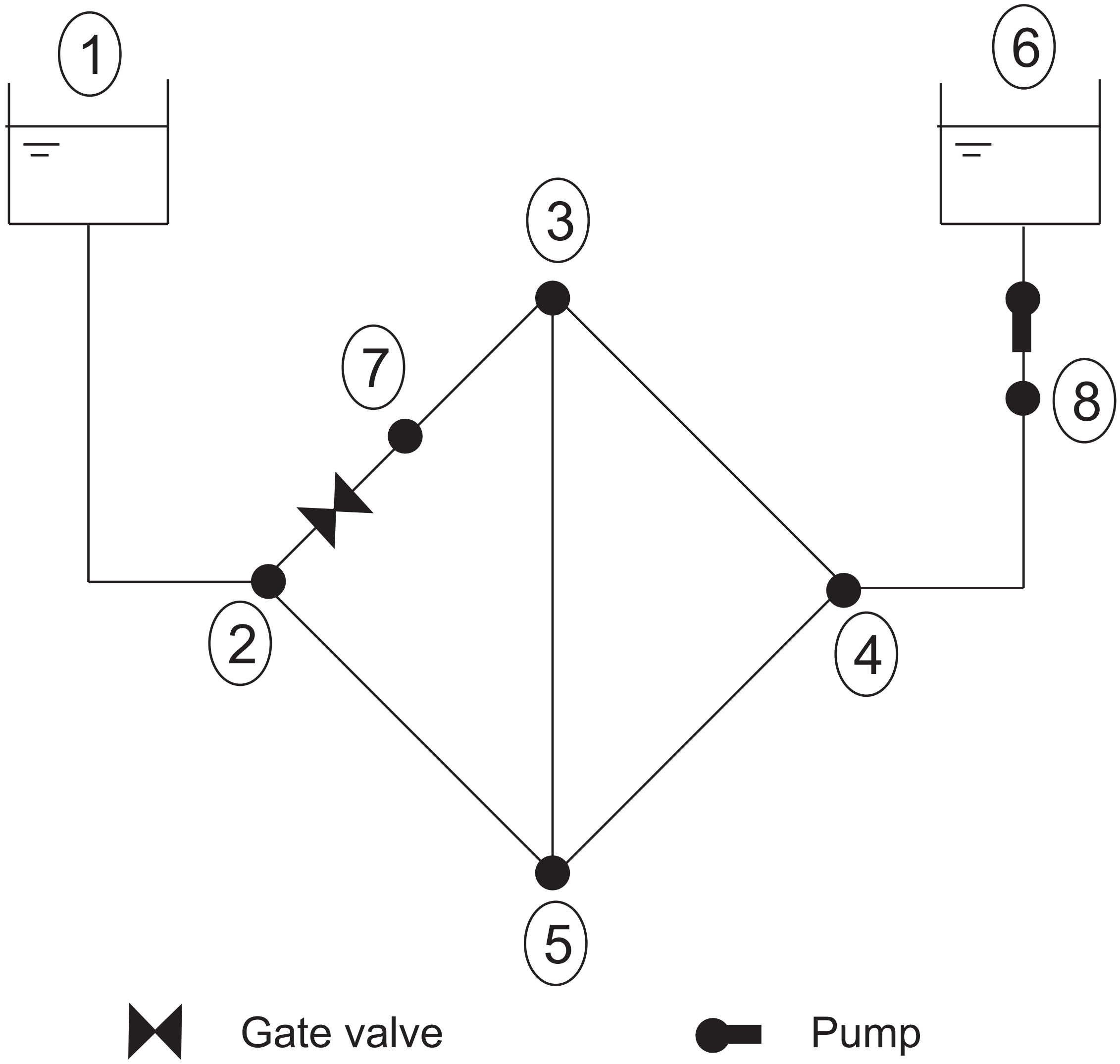

Fig. 3. Available flow meters at different measurement configurations