State Variable Implications on Hydraulic State Estimation

Emilio Ruiz1 · Sarai Díaz1 · Javier González1,2

Received: 21 September 2021 / Accepted: 13 April 2022 / Published online: 5 May 2022
© The Author(s) 2022

Abstract
Hydraulic State Estimation allows to estimate the most likely state of a water supply network considering multiple sources of information and their associated uncertainty. It is set out as an optimization problem, often addressed according to the Weighted Least Squares criterion. It can be formulated differently depending on the selected set of state variables. This choice is not straightforward and leads to different problem dimensions, time complexities and convergence behavior. All possible essential approaches are gathered in this work and two are identified as the most suitable according to the resulting problem dimensions: a head-based approach and a demand-based approach. The particularities and the formulation according to both sets of state variables are discussed, seeking for efficient implementation. Their time complexity and convergence behavior are compared to draw conclusions that help to identify the most suitable approach for real practice applications.

Keywords Hydraulic state estimation · State variables · Demand-based · Head-based · Convergence behavior · Time complexity · Problem dimension

1 Introduction

Having a reliable knowledge of the hydraulic state of a water network is essential for its management, which seeks to meet user requirements. Traditionally, water network analysis has focused on models that solve the hydraulics of water systems (e.g. Cross 1936; Martin and Peters 1963; Epp and Fowler 1970; Hamam and Brameller 1971; Wood and Charles 1972; Todini and Pilati 1988). These methods are all projections of the same set of equations (Todini and Rossat 2013): the governing physical equations of pipe flow
(i.e. mass continuity and energy conservation) and the network topology. Their differences stem from the driving unknowns adopted to solve the problem. Note that the usual inputs required to solve the flow problem in a basic network are head levels at known-head nodes (reservoirs and tanks), nodal demands and pipe roughness values. Pipe flows and the rest of head levels constitute the problem unknowns (Elhay and Simpson 2011). Either head levels or flows can be considered the driving unknowns, leading to two different approaches to the problem: computing head levels and updating flows or computing flows and updating head levels. The Global Gradient Algorithm -GGA method- (e.g. Todini and Pilati 1988; Rossman 2000) and the Loop method (e.g. Epp and Fowler 1970; Alvarruiz et al. 2018) are good representatives of these head-based and flow-based approaches, respectively.

Regardless of the method adopted, the hydraulic model must be able to simulate the complex reality of the water system. In practice, not only traditional inputs (head levels at known-head nodes, nodal demands and roughness values) but any system variable may be measured or estimated based on historical records (i.e. pseudomeasured). Moreover, these measurements and/or estimates are always uncertain (Bragalli et al. 2016), as different sources of information may be subjected to different levels of uncertainty (Ruiz et al. 2022). Therefore, it can be argued that the traditional scheme of solving the flow network has two substantial drawbacks when coping with the complex reality of water systems: (1) the input set is closed, so other network measurements cannot be readily incorporated and (2) inputs do not systematically consider uncertainty in each model run. These two setbacks can be sorted out with simulation methods (Kumar et al. 2008), but they are computationally expensive because they need several model runs to characterize a single time.

Hydraulic State Estimation (HSE) can be considered an alternative to overcome these difficulties. HSE is posed as an optimization problem that respects the subjacent flow governing equations (Carpentier and Cohen 1991). All water system variables (i.e. head levels, nodal demands, pipe roughness values and pipe flows) are considered unknowns in this approach. They must be computed considering the set of measured variables and the set of state variables (Jung and Kim 2018). Measured variables come determined by the measurements and/or pseudomeasurements available at the network (Bargiela and Hainsworth 1989). This enables to consider any measurement/pseudomeasurement and their uncertainty as input. State variables are the group of independent variables required to describe the network hydraulic state, i.e. to compute the rest of the system variables (Sterling and Bargiela 1984). State variables are related to the rest of variables by the governing physical equations of pipe flow and the network topology (Bargiela 1984). Therefore, HSE solves the hydraulics of the water system, but it goes one step further: it provides the value of the state variables that minimizes the difference between computed measured variables and measurements/estimates (Díaz et al. 2017).

As it happened when solving the flow network problem, several approaches can be posed to address the HSE problem (e.g. Powell 1992; Andersen and Powell 2000; Díaz et al. 2017, 2018; Vrachimis et al. 2019; Wang et al. 2021). These approaches are characterized by the objective function criterion and the selected set of state variables. Weighted Least Squares (WLS) is one of the most popular criteria for the objective function in HSE literature due to its proved advantages (Tshehla et al. 2017). Regarding state variables, to the best of the authors knowledge, there is no thorough discussion about the implications of adopting a specific set of state variables. The choice is not trivial from a computational point of view: different sets lead to problems with different dimensions, convergence performance and time complexity. Hence, it is important in order to operationally implement HSE in large real systems.
The objective of this paper is to provide a conceptual framework to assess the computational implications of selecting a specific state variable set for HSE and to identify the most suitable approach in real practice. For this purpose, time complexity, execution time and convergence performance are here evaluated. An efficient computational implementation of the problem is proposed and formulae are developed for any possible measurement at a water system. HSE is formulated all along according to the WLS scheme because it is the most widely adopted criterion in the literature and it presents several advantages that will be later explained. A similar state variable analysis could be undertaken for other objective function criteria if needed, but it is out of the scope of this paper.

The rest of the paper is organized as follows. An overview of HSE under a WLS approach is first presented. Then, all possible essential sets of state variables are considered and two (a head-based and a demand-based approach) are identified as the most efficient, bearing resemblance with the two options that have been mainly adopted to solve the traditional flow network problem. These two sets are discussed in more detail, explaining their implications on the formulation and giving hints for efficient implementation. Their time complexity is generally analyzed and execution times and convergence performance are assessed for several case studies. Finally, conclusions are drawn in order to identify the most suitable approach.

2 Methodology

2.1 Overview of Hydraulic State Estimation

HSE is an optimization problem that provides the most likely state of a water supply network. This state is understood as the one which minimizes the difference between the computed measured variables and their corresponding measurements or pseudomeasurements. Such variables are calculated from the estimated state variables. Both state and measured variables are related non-linearly through the governing equations of water systems (Díaz et al. 2016b), i.e. mass continuity, energy conservation and network topology. This is expressed with the following model:

\[
\varepsilon = z - g_m(x)
\]  

where \( z \in \mathbb{R}^m \) is the vector of measurements, \( x \in \mathbb{R}^n \) is the vector of state variables, \( g_m(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is the function of non-linear relationships between measured and state variables and \( \varepsilon \in \mathbb{R}^m \) is the vector of differences between measured variables and their corresponding measurements, which is to be minimized. It is important to note that \( g_m(x) \) is a subspace of \( g(x) : \mathbb{R}^n \rightarrow \mathbb{R}^r \), which is the function of non-linear relationships between state variables and any other water system variable.

Several criteria might be chosen to minimize the model in Eq. (1), leading to different objective functions. State estimation was first adapted to water supply systems by Coulbeck (1977). He used a WLS approach for the objective function, although measurement weights were arbitrary. This criterion was chosen because the unweighted Least Squares yields the minimum variance unbiased solution when measurement uncertainties are solely Gaussian (Tshehla et al. 2017), but water system uncertainties are not and they register significant outliers (Bargiela 1984). Measurement weights are used to deal with high uncertainties while maintaining the virtues of the unweighted Least Squares. Bargiela (1984) defined measurement weights as the inverse of measurement uncertainty, establishing a clear relation between those two. He also compared WLS method with
the Weighted Least Absolute Values criterion. His studies drew that both approaches were comparable in terms of numerical stability and convergence, but WLS was a better option for large, hardly measured and noisy systems, such as water supply networks. These advantages explain why WLS has been widely adopted in the literature (and so in this work) for HSE. The objective function according to WLS criterion is posed as (Carpentier and Cohen 1991):

$$\min_x J(x) = \frac{1}{2} \epsilon^T C_z^{-1} \epsilon$$

subject to $$g_f(x) = 0$$

where $$J(x)$$ is the objective function to minimize, $$C_z \in \mathbb{R}^{m \times m}$$ is the variance-covariance matrix of measurements, which acts as the weight matrix, and $$g_f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^c$$ is the function of equality constraints. It is important to note that (1) $$C_z$$ is typically assumed diagonal, namely measurements are independent, and (2) $$g_f(x)$$ is also a subspace of $$g(x)$$. To specify the constraints, it is necessary to select the set of state variables.

The optimal solution of Eq. (2) and Eq. (3) is $$\hat{x} \in \mathbb{R}^n$$. As this solution minimizes the residuals between measurements and variable estimates, it represents the most likely hydraulic state of the water system under study. HSE allows to get distributed information from scattered measurements and/or estimates, as shown in Fig. 1. In this figure, several disjointed measurements and pseudomeasurements (i.e. inputs) are available, each one with a given uncertainty. From these inputs, distributed hydraulic state information is estimated, filling unmeasured gaps. Since inputs are uncertain, the most likely hydraulic state does not necessarily coincide with every input.

Several numerical methods are available to find $$\hat{x}$$. They consist on iteratively moving from one state to another through the objective function domain. This displacement might be accomplished in different ways and many of the methods achieve it by computing a decrease direction ($$d \in \mathbb{R}^n$$) per iteration (Bazaraa et al. 2006), as shown in Fig. 2. Newton’s method is widely used in literature since $$J(x)$$ is twice differentiable and $$g(x)$$ derivatives can be computed analytically:

$$x_{i+1} = x_i + \Delta x_i = x_i + d_{i+1} = x_i - (\nabla^2_x J(x_i))^{-1} \nabla_x J(x_i)$$

where $$i$$ is a counter for iterations, $$\Delta x$$ is the vector of total variation of each state variable per iteration, $$\nabla_x$$ is the gradient operator with respect to the state variables and $$\nabla^2_x$$ is the Hessian operator with respect to the state variables. It is a gradient method, i.e. the decrease direction is wisely computed using information about the direction of steepest
An advantage of Newton’s method is that its convergence is quadratic, unlike gradient methods that only use first derivative information, whose convergence rate is linear (Bazaraa et al. 2006).

As shown in Eq. (2) and Eq. (3), the problem is actually constrained. To take into account the constraints, Newton’s method is not only applied to \( J(x) \) but also to the Karush–Kuhn–Tucker first order optimality conditions. Thus, \( \Delta x_{i+1} \) computation takes the form of (Gómez-Quiles et al. 2013):

\[
\begin{bmatrix}
G_m^T C_z^{-1} G_m & G_f^T \\
G_f & 0
\end{bmatrix}
\begin{bmatrix}
\Delta x_{i+1} \\
\lambda_{i+1}
\end{bmatrix}
= \begin{bmatrix}
G_m^T C_z^{-1} [z - g_m(x_i)] \\
g_f(x_i)
\end{bmatrix}
\]

(5)

where \( G_m \in \mathbb{R}^{m \times n} \) and \( G_f \in \mathbb{R}^{c \times n} \) are the Jacobian matrices of \( g_m(x) \) and \( g_f(x) \) with respect to the state variables, respectively, and \( \lambda \in \mathbb{R}^c \) is the dual variable vector. \( G_m \) and \( G_f \) are sub-matrices of \( G \in \mathbb{R}^{v \times n} \), which is the Jacobian matrix of \( g(x) \) with respect to the state variables. The necessary and sufficient condition to be able to solve the problem is that \( \begin{bmatrix} G_m \\ G_f \end{bmatrix} \) has full rank, i.e. the water system is observable (Díaz et al. 2015). This can be achieved if there are enough measurements, estimations and constraints to estimate every state variable. Redundant measurements are recommended because that leads to better system matrix conditioning (e.g. Bargiela 1984).

In each iteration, once \( x \) is known, the rest of variables of the corresponding hydraulic state are computed through \( g(x) \). This last step is equivalent to solving the flow network according to the governing equations of pipe flow. HSE is not independent from solving the flow network: it embeds and broadens the plain resolution of the physics of the problem. The resulting linear system in Eq. (5) is solved iteratively starting from some initial state variable values \( (x_0) \) until the stopping criterion is met. This typically corresponds to \( ||\Delta x_{i+1}|| \) being lower than a chosen tolerance. To ensure constraint accomplishment, \( ||g_f(x_{i+1})|| \) being lower than a chosen tolerance might also be considered. The algorithm is summarized in Algorithm 1.
It is important to note that, in Eq. (5), the Hessian of $J(x)$ is approximated. It is assumed that the second derivative of $g(x)$ is null. Caro et al. (2011) state that this approximation is fair enough and holds whether $C_z$ is diagonal or not. However, in order to correctly apply Newton’s method, it is important to assume both the Hessian approximation and $C_z$ diagonal structure. This assures that the Hessian is symmetric positive definite. If the Hessian is not positive definite, it cannot be guaranteed that the computed decrease direction is an actual descent direction to the optimum in $J(x)$ domain (Castillo et al. 2002).

### 2.2 Possible Essential Sets of State Variables

Until now, the presented formulae are quite general and could be easily applied to any state estimation problem (i.e. not necessarily a water system). The key point to make the formulation hydraulic-specific is the function $g(x)$, and its subsequent mathematical objects: $g_m(x)$, $g_t(x)$, $G$, $G_m$ and $G_t$. They depend on the selected set of state variables. As stated in Sect. 1, this set is not unique and several configurations are possible. All essential possibilities for a basic network are gathered and compared in Table 1. $H \in \mathbb{R}^{r+j}$ represents head levels, $Q \in \mathbb{R}^p$ stands for pipe flows, $q \in \mathbb{R}^j$ is nodal demands and $R_{mat} \in \mathbb{R}^{mat}$ is used for pipe material roughness. Regarding dimensions, $r$, $j$, $p$ and $mat$ stand for known-head nodes, junctions (rest of nodes), pipes (links) and pipe materials, respectively. $H$ can then be split into $H_e \in \mathbb{R}^r$ for known-head nodes and $H_j \in \mathbb{R}^j$ for junctions. Pipe material roughness (i.e. roughness of each material) and pipe roughness (i.e. roughness of each pipe) are here distinguished to consider that several pipes can be associated with a same roughness value. Both are related through $C_m \in \mathbb{R}^{mat \times p}$, which is a matrix that links pipe material roughness with corresponding pipes. Its components are 1 if a pipe is made of a material and 0 otherwise. $R_p \in \mathbb{R}^p$ is the vector of pipe roughness, obtained from $R_{mat}$ and $C_m$. If $mat = p$, then $C_m = I$ and $R_p = R_{mat}$.

Table 1 shows that different sets of state variables lead to different problem dimensions. It is expected that the higher the problem dimension, the slower the computational time. In water
supply networks, \( j \leq p \) and \( mat \leq p \) necessarily (\( mat \) is usually much smaller than \( p \)) and \( r \ll j \) also holds in practice. It is important to note that, for clarity, the sets presented in Table 1 are the essential possibilities. They can be combined leading to plenty of secondary sets, but this does not provide any advantage in terms of dimension. For example, \( Q \) set has a higher dimension than \( q \) set and both can be combined resulting in a mixed set with flows and demands. Any nodal demand can be swapped by all flows entering and leaving the node and vice-versa. The dimension of the resulting set is always between the dimensions of \( Q \) and \( q \) sets. Then, the mixed set has no advantages according to dimensions because \( q \) set can be used instead and always has a lower dimension. This reasoning can be applied to all combinations.

From the state variable sets in Table 1, two are of special interest: \( H_r, H_j \) and \( R_{mat} \) (head-based), and \( H_r, q \) and \( R_{mat} \) (demand-based). These two sets are the ones associated with the lowest problem dimension, so they are the most computationally efficient. These two possible sets of state variables for HSE (head-based vs demand-based) bear resemblance with the choice between head levels or flows as driving unknowns when just solving the flow problem.

### 2.3 Head-Based State Estimation

In this case, state variables are \( H_r, H_j \) and \( R_{mat} \). With these state variables, \( g(x) \) can be defined in three blocks as:

\[
g(x) = \begin{cases} 
    h = \hat{h}(Q, R_{mat}) = -A_j^T H_j - A_r^T H_r \\
    Q = \hat{h}^{-1}(h, R_{mat}) \\
    q = A_j Q 
\end{cases}
\]

where \( h \in \mathbb{R}^p \) is the vector of pipe head-losses, \( \hat{h} : \mathbb{R}^{p+mat} \rightarrow \mathbb{R}^p \) is the Hazen-Williams head-loss formula and \( A \in \mathbb{R}^{(r+j) \times p} \) is the incidence matrix of the water system topology. \( A \) components are 0 if a node is not connected to a link, 1 if a node is the final end of a link and -1 if a node is the initial end of a link. \( A \) is split into \( A_r \in \mathbb{R}^{r \times p} \), only considering known-head nodes, and \( A_j \in \mathbb{R}^{j \times p} \) for junctions. The second block of Eq. (6) fully depends on the selected head-loss formula (Hazen-Williams in this work) and non-linearity stems from it.

It is important to note that, with this state variable set and \( g(x) \) defined as in Eq. (6), the rest of system variables (head-losses, flows and demands) are explicit by blocks in \( g(x) \), and these blocks can be sequentially and directly solved. Then, once state variables are available (i.e. computed in each HSE iteration), the rest of variables can be computed straight away. Traditional iterative numerical methods to solve non-linear systems of equations.
(typically used when solving pipe flow governing equations, e.g. in Loop or GGA methods) are not required to update the hydraulic state through \( g(x) \). This gives computational enhancement to the algorithm.

Junctions can be divided into transit nodes \((j_t)\), with nil demand, and demand nodes \((j_d)\). In this work, transit nodes are considered fixed constraints that impose null demand. This is the only type of constraint with this set of state variables \((q_t = 0; q_d \in \mathbb{R}^j)\). According to Eq. (6), these are non-linear constraints with respect to the chosen state variables. Since this division affects demands and not head levels, state variables \( H_j \) must still account for all junctions.

According to the state variables and Eq. (6), the structure of \( G \) is:

\[
G = \begin{bmatrix}
\frac{\partial H_t}{\partial H_t} & \frac{\partial H_t}{\partial H_j} & \frac{\partial H_t}{\partial R_{mat}} \\
\frac{\partial H_j}{\partial H_t} & \frac{\partial H_j}{\partial H_j} & \frac{\partial H_j}{\partial R_{mat}} \\
\frac{\partial R_{mat}}{\partial H_t} & \frac{\partial R_{mat}}{\partial H_j} & \frac{\partial R_{mat}}{\partial R_{mat}}
\end{bmatrix} = \begin{bmatrix}
I & 0 & 0 \\
0 & I & 0 \\
-D_0^{-1}A_r^T & -D_0^{-1}A_d^T & D_1C_m^T \\
-A_dD_0^{-1}A_r & -A_dD_0^{-1}A_d & A_dC_m^T \\
0 & 0 & I
\end{bmatrix}
\] (7)

where \( I \) is the corresponding dimension identity matrix and \( D \in \mathbb{R}^{p \times p} \) are diagonal matrices coming from the following derivatives: \( D_0 = \partial h / \partial Q \) and \( D_1 = \partial Q / \partial R_p \).

\( G_m \) and \( G_r \) are obtained from Eq. (7). Adequate matrix rows must be selected according to applicable constraints and available measurements. In case of full measurement redundancy (i.e. all water system variables are measured), all rows are selected either as constraints or measurements. In reality, financial resources are limited and not all possible variables are measured (Savic et al. 2009). Matrices are presented in this work for the full redundancy case for the sake of completeness.

2.4 Demand-Based State Estimation

In this case, state variables are \( H_t, q \) and \( R_{mat} \). This set poses a problem regarding \( G \) computation. Every derivative \( \partial / \partial q \) can be obtained through \( \partial Q / \partial q \) according to the chain rule. The issue is that the relation between \( Q \) and \( q \) \([q = q(Q)]\) is not always bijective. To see this, one might think of a looped water supply network. In that kind of network, several sets of \( Q \) are related to the same \( q \) image if energy is not considered. A necessary and sufficient condition to be able to invert a function is that it must be bijective (Devlin 2004). Then, \( q = q(Q) \) cannot be inverted to \( Q = Q(q) \) and \( \partial Q / \partial q \) does not exist in this case.

To solve this problem and be able to compute \( G \), it is necessary to resort to an equivalent water system model. The function \( q = q(Q) \) is bijective only when the water system is a tree network (in these networks, \( Q \) sets \( q \) unequivocally and vice-versa without considering energy). Working with the spanning tree of a looped network and considering chord link flows \((Q_L \in \mathbb{R}^l)\) as additional state variables is a valid solution to compute \( G \). In this case, \( g(x) \) blocks can be redefined as:
\[
g(x) = \begin{cases} 
Q = \begin{bmatrix} Q_T \\
Q_L \end{bmatrix} = M^T Q_L + \begin{bmatrix} A^{-1}_j & 0 \end{bmatrix} q \\
h = \hat{h}(Q, R_{\text{mat}}) \\
H_j = H_{rj} I - A^{-T}_j h 
\end{cases} \tag{8}
\]

where \(Q_T \in \mathbb{R}^T\) is the vector of tree pipe flows, \(L\) is the number of chord pipes or loops (\(L\) equals \(p - j\)), \(T\) is the number of tree pipes (\(T\) equals \(j\)), \(M \in \mathbb{R}^{L \times p}\) is the loop matrix, \(H_{rj}\) is the head level of the known-head node contained in the tree network, \(I \in \mathbb{R}^j\) is a vector whose components are 1 and \(A^{-T}_j \in \mathbb{R}^{j \times T}\) is the sub-matrix of \(A_j\) when only tree pipes are considered. For each chord pipe, it is necessary to find a related loop in the tree network. \(M\) associates loops and pipes. Its components are 0 if a pipe does not belong to a loop, 1 if a pipe belongs to a loop and both pipe and loop directions coincide and -1 if a pipe belongs to a loop but pipe and loop directions do not match. Again, the second block of Eq. (8) fully depends on the selected head-loss formula (Hazen-Williams in this work) and non-linearity comes from it.

Equation (6) could be alternatively used as \(g(x)\) with this set of state variables but, in this case, it would be necessary to resort to iterative numerical methods to solve the non-linear system of equations (e.g. as in Loop or GGA methods) in order to update the hydraulic state in each iteration. Redefining \(g(x)\) as Eq. (8) for the demand-based approach allows to maintain the advantages of solving the unknowns (flows of tree pipes, head-losses and heads of junctions) explicitly by blocks. Thus, traditional iterative numerical methods to solve non-linear systems of equations are neither necessary with these state variables. This gives computational enhancement to the algorithm.

There are two types of constraints in this demand-based approach. The first type is again null demands at transit nodes, which are here considered as constraints that impose a fix demand \((q_{j_h} = 0)\). The second type comes from the need for imposing the physical constraints of the original looped network, which would not be present otherwise in the tree network. Nodal head levels at chord pipe ends must agree with pipe head-losses \([h_L = \hat{h}(Q_L, R_{\text{mat}_L}) = -A^T_{j;L} H_j - A^T_{r;L} H_{rL}]\). Here, \(A_{j;L} \in \mathbb{R}^{j \times L}\) and \(A_{r;L} \in \mathbb{R}^{r \times L}\) are the sub-matrices of \(A_j\) and \(A_r\) when only chord pipes are considered, respectively. \(h_L \in \mathbb{R}^L\) is the vector of chord pipe head-losses and \(R_{\text{mat}_L} \in \mathbb{R}^{r \times m}\) is the vector of chord pipe material roughness.

The first type of constraint directly affects some state variables \((q_{j_h})\), so it can be imposed by straightforwardly considering those state variables as null. Thus, the amount of state variables to estimate reduces. The resulting state variables are \(H_r, q_{j_h} \in \mathbb{R}^{j_h}, Q_L\) and \(R_{\text{mat}}\). According to Eq. (8), the second type of constraint is non-linear with respect to the chosen state variables.

According to the state variables and Eq. (8), the resulting structure of \(G\) is:

\[
G = \begin{bmatrix} 
\partial H_r & \partial H_r & \partial H_r & \partial H_r & \partial H_r & \partial H_r & \partial H_r & \partial H_r \\
\partial H_r & \partial q_{j_h} & \partial Q_L & \partial H_j & \partial H_j & \partial H_j & \partial H_j & \partial H_j \\
\partial H_r & \partial q_{j_h} & \partial Q_L & \partial H_j & \partial H_j & \partial H_j & \partial H_j & \partial H_j \\
\partial H_r & \partial q_{j_h} & \partial Q_L & \partial H_j & \partial H_j & \partial H_j & \partial H_j & \partial H_j \\
\partial R_{\text{mat}} & \partial R_{\text{mat}} & \partial R_{\text{mat}} & \partial R_{\text{mat}} & \partial R_{\text{mat}} & \partial R_{\text{mat}} & \partial R_{\text{mat}} & \partial R_{\text{mat}} \\
\partial R_{\text{mat}} & \partial q_{j_h} & \partial Q_L & \partial q_{j_h} & \partial Q_L & \partial q_{j_h} & \partial q_{j_h} & \partial q_{j_h} \\
\partial R_{\text{mat}} & \partial q_{j_h} & \partial Q_L & \partial q_{j_h} & \partial Q_L & \partial q_{j_h} & \partial q_{j_h} & \partial q_{j_h} \\
\partial R_{\text{mat}} & \partial q_{j_h} & \partial Q_L & \partial q_{j_h} & \partial Q_L & \partial q_{j_h} & \partial q_{j_h} & \partial q_{j_h} \\
\partial R_{\text{mat}} & \partial q_{j_h} & \partial Q_L & \partial q_{j_h} & \partial Q_L & \partial q_{j_h} & \partial q_{j_h} & \partial q_{j_h} \\
\end{bmatrix} = \begin{bmatrix} I & 0 & 0 & 0 \\
0 & \Theta & \Lambda & \Xi \\
0 & 0 & I & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix} \tag{9}
\]
where $A_{j,d}^{-1} \in \mathbb{R}^{j \times j_d}$ is the sub-matrix of $A_{j,T}^{-1}$ related to demand junctions and the other matrices are:

$$
\Gamma = -(A_{j,T}A_{j,T}^T)^{-1}A_{j,T}A_{r,T}^T \tag{10}
$$

$$
\Theta = -(A_{j,T}A_{j,T}^T)^{-1}A_{j,T}D_0_\tau A_{j,d}^{-1} \tag{11}
$$

$$
\Lambda = -(A_{j,T}A_{j,T}^T)^{-1}A_{j,T}D_0_\tau M_{r,T} \tag{12}
$$

$$
\Xi = -(A_{j,T}A_{j,T}^T)^{-1}A_{j,T}D_2_r C_{m_T} \tag{13}
$$

where dimensions are: $\Gamma \in \mathbb{R}^{j \times r}, \Theta \in \mathbb{R}^{j \times j_d}, \Lambda \in \mathbb{R}^{j \times L}$ and $\Xi \in \mathbb{R}^{j \times \text{mat}}, A_{r,T} \in \mathbb{R}^{r \times T}, D_0 \in \mathbb{R}^{r \times T}, M_T \in \mathbb{R}^{L \times T}$ and $C_{m_T} \in \mathbb{R}^{\text{mat} \times T}$ are the sub-matrices of $A_{r}, D_0, M$ and $C_{m_T}$ when only tree pipes are considered. $D_2 \in \mathbb{R}^{p \times p}$ is a diagonal matrix coming from $\partial h/\partial R_p$ and $D_{2_T} \in \mathbb{R}^{T \times T}$ is its sub-matrix for tree pipes.

$G_m$ and $G_r$ are obtained from Eq. (9). Adequate matrix rows are selected and combined according to available measurements and energy conservation at chord pipes (second type of constraint). Like before, all rows would be selected in case of full measurement redundancy, which is unlikely to happen in real water supply networks.

### 2.5 Pre and Post-Processing

To perform HSE, pre-processing might be necessary. From Sect. 2.4, it can be guessed that demand-based HSE needs a mandatory topological pre-processing. A spanning tree of the network and the related set of chord pipes and loops must be previously identified. The two most common and broadly used algorithms to find a spanning tree are Kruskal and Prim’s algorithms (see Cormen et al. 2001). With respect to loops, a widely used method consists on obtaining a spanning tree of the network and adding loops connecting both ends of chord pipes through the tree. According to Alvarruiz et al. (2015), this method does not suit the typical Loop method to solve pipe flow equations because resulting loops have high overlapping. However, the method behooves HSE because it can benefit from the necessary spanning tree computation and chord links consideration as state variables and overlapping is not a critical issue. Alternative methods might be the ones proposed by Creaco and Franchini (2014) or Alvarruiz et al. (2015). When using head-based HSE, topological pre-processing is not mandatory, although it can be used to find looped and tree zones in complex networks.

Note that, in general, Eq. (9) components are more complex to obtain than those of Eq. (7). Part of this additional computational complexity comes from purely topological matrix operations. These operations can be performed as a pre-processing since their result is not going to change unless network topology changes (e.g. when there are control elements in the water system).

Post-processing can be an excellent add-on to make the most of HSE. The uncertainty of both the estimated state variables and the rest of computed water supply network variables can be assessed. First methods for uncertainty assessment were based on Monte Carlo and Latin Hypercube approaches (e.g. Pasha 2006; Kang 2009). These approaches require
many simulations. As one of the main advantages of HSE is that it estimates the most likely state in one simulation, First-Order Second Moment approaches (Caro et al. 2011; Díaz et al. 2016a) may be more convenient to assess HSE uncertainty. Post-processing is here mentioned for the purpose of motivation and completion, but it is out of the scope of this paper.

3 Results

Two possible sets of state variables have been identified as the most advantageous for HSE based on the problem dimension (i.e. number of state variables). Both of them can be applied to any water supply network, so it is important to assess their performance in order to choose the most appropriate in each situation. Two indicators are used to compare the behavior: time complexity and convergence performance. Time complexity is here expressed in Big-O notation using a uniform cost model, i.e. operations are assumed to have a constant temporal cost. Big-O notation gives an asymptotic upper bound to computational time increment when the input dimension increases (Cormen et al. 2001). Big-O notation is not strictly computational time, so execution times are also analyzed in this work for several case studies. Convergence performance refers to convergence speed (i.e. the number of iterations needed to converge). Convergence is here checked for the same case studies. Bad time results may be counteracted by a good convergence performance and vice-versa.

When assessing Big-O notation, worst-case scenarios and large input dimensions (i.e. close to infinity but computationally manageable) must be assumed to characterize the upper bound. Since Big-O notation describes asymptotic behavior, two approximations are usually taken into account when adding time complexities (Knuth 1997). First, if the addition involves time complexities of the same order, resultant numerical coefficients are neglected. Second, if the addition involves time complexities of different orders, the result is the time complexity with the highest order.

3.1 Time Complexity Analysis

3.1.1 Pre-Processing

Pre-processing is mandatory in demand-based HSE and optional in head-based HSE. Since it is the worst-case scenario for time complexity, it is here assumed that head-based HSE includes an equivalent topological analysis to that of demand-based HSE. As commented in Sect. 2.5, demand-based HSE pre-processing has two stages: topological analysis (tree and loops computation) and topology-related operations of \( \mathbf{G} \) (to speed calculations of matrices in Eq. 9). Pre-processing time complexities are shown in Table 2.

Regarding the spanning tree computation, both Kruskal and Prim’s algorithms are well known and their time complexity has been vastly studied. Unless Prim’s algorithm is combined with sophisticated heap queue data structures (whose practical advantage is disputed), both algorithms have the same time complexity (see Brodal 1996; Cormen et al. 2001; Brodal et al. 2012). In order to find loops, it is supposed that they are obtained from chord pipes and the previously computed spanning tree, as commented in Sect. 2.5. Regardless of the method, loops are built using shortest path algorithms. For unweighted graphs (such as water systems), a widely used and well known option is the Breadth-First
algorithm (Cormen et al. 2001), which is here adopted. Concerning \( G \) topology-related operations, their time complexity considers taking advantage of the highly sparse structure of the involved topological matrices.

As shown in Table 2, time complexity of demand-based HSE pre-processing is dominated by \( G \) topology-related operations and it tends to \( O((j_d + j_i)^2) \), i.e. it only depends on the amount of junctions. By contrast, time complexity of head-based HSE pre-processing is dominated by loop finding and tends to \( O(L \times (r + j_d + j_i)) \). It depends on the number of nodes and loops. While \( L < j \), which is usual in real water systems, time complexity of head-based approach tends to a linear rate and it is smaller than time complexity of demand-based approach, which tends to a quadratic rate. Otherwise, both time complexities are similar if \( j \approx L \) and demand-based time complexity is smaller than head-based time complexity if \( j < L \).

### 3.1.2 Vector Assembly

This time complexity comprises vectors \( x \) and \( z \) assembly and matrix \( C_z^{-1} \) computation. This time complexity is presented in Table 3 per iteration, i.e. expressions consider only one iteration of the while loop in Algorithm 1. The assembly of vectors \( x \) and \( z \) directly depends on the number of state variables and measurements, respectively. It is here assumed all along that measurements have full redundancy, i.e. all water system variables are measured, since it is the worst-case scenario for time complexity. With respect to \( C_z^{-1} \), a diagonal matrix structure is assumed, i.e. measurements are considered independent among each other, as discussed in Sect. 2.1.

Both demand-based and head-based HSE tend to the same linear rate, which is dominated by measurement-related operations in case of full-redundancy: \( O(r + j_d + j_i + \text{mat} + p) \). It depends on the total amount of water system variables.

### 3.1.3 \( G \) Computation

Time complexity of \( G \) computation is determined by those derivatives which (1) imply matrix operations and (2) cannot be completely pre-processed because they are not only topological. Results are again shown per iteration in Table 4. Note that time complexities take

---

**Table 2** Pre-processing time complexity

| Operation                  | Head-based                                      | Demand-based                                     |
|----------------------------|-------------------------------------------------|-------------------------------------------------|
| Computing spanning tree    | \( O(p \times \log(r + j_d + j_i)) \)           | \( O(p \times \log(r + j_d + j_i)) \)           |
| Finding loops              | \( O(L \times (r + j_d + j_i)) \)               | \( O(L \times (r + j_d + j_i)) \)               |
| Topology-related \( G \) operations | -                                              | \( O((j_d + j_i)^2) \)                          |
| Total                      | \( O(L \times (r + j_d + j_i)) \)               | \( O((j_d + j_i)^2) \)                          |

**Table 3** Vector assembly time complexity

| Operation | Head-based | Demand-based |
|-----------|------------|--------------|
| \( x \)   | \( O(r + j_d + j_i + \text{mat}) \)             | \( O(r + j_d + \text{mat} + L) \)               |
| \( z \)   | \( O(r + j_d + j_i + \text{mat} + p) \)          | \( O(r + j_d + j_i + \text{mat} + p) \)          |
| \( C_z^{-1} \) | \( O(r + j_d + j_i + \text{mat} + p) \)        | \( O(r + j_d + j_i + \text{mat} + p) \)        |
| Total     | \( O(r + j_d + j_i + \text{mat} + p) \)          | \( O(r + j_d + j_i + \text{mat} + p) \)          |
into account the structure of the involved matrices and which matrices are pre-processed, as well as the operation order. As it will be discussed later, operations according to the head-based approach involve highly sparse matrices, while matrices used in the demand-based approach are highly dense.

Time complexity of demand-based $G$ computation tends to $O((j_d + L + mat) \times (j_d + j_t)^2)$. It depends on the number of junctions and loops. Time complexity of head-based $G$ computation tends to $O((j_d + j_t)^2)$, i.e. it only depends on the amount of junctions. The time complexity of head-based approach is quadratic and smaller than that of the demand-based approach. Depending on the relationship among loops, transit nodes and demand nodes at a particular network, demand-based time complexity can be closer to a quadratic or a cubic rate.

### 3.1.4 Newton’s Method

Applying Newton’s method involves three major operations: assembling the matrix and right-hand vector of Eq. (5), solving the obtained linear system of equations and updating the state variables with Eq. (4). The time complexities of these operations are gathered per iteration in Table 5, taking into account the structure of the involved matrices. Regarding the head-based approach, both $G$ and the matrix of the system of equations have sparse structures. On the contrary, their structures are dense for the demand-based approach.

Under the assumptions in Sect. 2.1, the system matrix of Eq. (5) is symmetric positive definite. Then, plenty of direct methods are available to readily solve the system of equations. Usually, these methods have a cubic time complexity for dense matrices (e.g. see Golub and Van Loan 1996) but they can go quadratic for sparse matrices.

Both demand-based and head-based HSE tend to time complexities which depend on the number of state variables. Head-based time complexity tends to a quadratic rate and it is smaller than demand-based time complexity, which tends to a cubic rate.

### 3.1.5 Updating Hydraulic State

This time complexity implies computing the three blocks of Eqs. (6) or (8) once state variables have been updated. As mentioned in Sect. 2, Eqs. (6) and (8) can be solved explicitly by blocks. Resultant time complexities are shown in Table 6 per iteration.
Both demand-based and head-based HSE tend to a linear rate. Until now, numerical coefficients have been discarded when showing time complexity total results, as discussed previously and shown in Table 6. Nevertheless, time complexities of each block and between head-based and demand-based approaches are now very similar, i.e. neglecting numerical coefficients may lead to misleading comparisons. When numerical coefficients are taken into account, head-based time complexity is \(O((r+j_d+j_i+\text{mat})^2)\) and demand-based time complexity is \(O((r+j_d+L+\text{mat})^3)\). It can be seen that the head-based rate is slightly smaller than the demand-based rate. If asymptotic approximations are considered and numerical coefficients are neglected, as in Table 6, both head-based and demand-based time complexities are virtually equal. These approximations are still valid, but the issue is raised here only to highlight that the head-based rate is slightly more advantageous.

### 3.1.6 Execution Times for Several Case Studies

According to previous results, head-based HSE has a more advantageous time complexity. The benefits of head-based time complexity are significant in pre-processing, \(G\) computation and Newton’s method. These results are now tested comparing the execution times of 6 different case studies: Modena, KL, Sabsevar, Neyshabur, Torbat and Kashmar. They are close to real, extensive water systems that can be used to derive practical conclusions in a broad range of water networks. Important information about the topology of each case study is presented in Table 7. It can be observed that there are less loops than junctions in all seven case studies, as discussed when pre-processing time complexities were explained.

Modena (Bragalli et al. 2012) case study is available at the Centre for Water Systems of the University of Exeter. KL case study is a modified version of the network presented by Kang and Lansey (2012), available at ASCE Task Committee on Research Database for Water Distribution Systems. Modena and KL case studies are well known in water system analysis literature. Sabsevar, Neyshabur, Torbat and Kashmar case studies are generated upon Iranian cities using DynaVIBe-Web application (see Mair et al. 2014a, b) and they are available at its repository. Sabsevar, Neyshabur, Torbat and Kashmar case studies here presented have not been studied in literature before. Layouts of case studies are presented in Fig. 3.

### Table 5 Newton’s method time complexity

| Operation                     | Head-based                     | Demand-based                   |
|-------------------------------|--------------------------------|--------------------------------|
| Building system of equations  | \(O((r+j_d+j_i+\text{mat})^2)\) | \(O((r+j_d+L+\text{mat})^3)\) |
| Solving system of equations   | \(O((r+j_d+j_i+\text{mat})^2)\) | \(O((r+j_d+L+\text{mat})^3)\) |
| Updating state variables      | \(O(r+j_d+j_i+\text{mat})\)    | \(O(r+j_d+L+\text{mat})\)     |
| Total                         | \(O((r+j_d+j_i+\text{mat})^2)\) | \(O((r+j_d+L+\text{mat})^3)\) |

### Table 6 Updating hydraulic state time complexity

| Operation   | Head-based | Demand-based |
|-------------|------------|--------------|
| Equation block 1 | \(O(p+r)\) | \(O(p+j_d+j_i)\) |
| Equation block 2 | \(O(p)\)   | \(O(p)\)     |
| Equation block 3 | \(O(j_d+j_i)\) | \(O(j_d+j_i)\) |
| Total        | \(O(p)\)   | \(O(p)\)     |
Execution times are shown per iteration in Table 8 for both demand-based and head-based HSE. Measurements are generated from prior solutions from EPANET (Rossman 2000) for each case study. EPANET is a benchmark software in water system analysis and it implements the previously mentioned GGA method (i.e. head-based flow network resolution). In order to resemble reality, only some EPANET’s results are used as measurements, i.e. there is not full measurement redundancy, to simulate measurement scarcity. Since it is the usual case in real practice, it is supposed that either measurements or pseudomeasurements of water demands, roughness values, head levels at known-head nodes and flows leaving or entering known-head nodes are available. Only some of the remaining flow and head-level possible measurements are considered to be available. They are randomly taken per case study using a binary random number generator, i.e. generating 0 and 1 with equal probability. Noise is introduced to pipe flow and junction head level measurements to model uncertainty. It is assumed that possible values of noisy measurements follow a normal distribution with EPANET’s results as mean values ($\mu$) and given standard deviations ($\sigma$). Standard deviations of 1 m and 7.5 are here considered for head levels and pipe material roughness, respectively. A coefficient of variation ($\sigma/\mu$) of 1.1 is set for pipe flows and nodal demands. In these tests, noisy values might be within the 95% confidence interval ($\mu \pm 2\sigma$). The tests are carried out using MATLAB R2020a version on a computer with 16 GB of RAM and 1 Intel(R) Core(TM) i7-10510U processor at 1.80 GHz and 4 cores.

From Table 8, it can be observed that execution times support the obtained time complexities. Execution times are related to problem dimensions and head-based HSE is faster than demand-based HSE. One of the main reasons for the time difference is the distinctive structure of both $G$ and the matrix of the system of equations according to the two approaches. These structures are shown in Fig. 4 for Kashmar case study. In this figure, only non-zero matrix elements are colored in grey. It can be appreciated that matrices are sparser under the head-based approach. Moreover, it was stated in Sect. 2.5 that loop overlapping is not a critical issue in HSE. The part of demand-based $G$ that is influenced by loop overlapping is highlighted in black in Fig. 4. This portion is small compared with the rest of the matrix.

### 3.2 Convergence Performance

Convergence performance of both head-based and demand-based HSE is checked and compared in the same 6 case studies used before. Convergence performance refers to convergence speed (i.e. the number of iterations needed to converge) in this work. In these tests, each approach is tested under two different assumptions in order to check its robustness against measurement uncertainty: noisy values might be within the 95% confidence interval ($\mu \pm 2\sigma$) or within the 99% confidence interval ($\mu \pm 3\sigma$).
Tolerances for convergence adapt EPANET’s default tolerances (Rossman 2000). The convergence tolerance for the algorithms presented in this work is the norm of \( \Delta x \) being, at least, 3 orders of magnitude lower than the smallest component of \( x \) and smaller than \( 10^{-6} \). Furthermore, the admissible error when imposing the problem constraints is also set. The norm of the vector of errors associated with constraints must be smaller than \( 10^{-4} \) and, at least, 2 orders of magnitude lower than the smallest value of the variables involved in the constraints. Results are shown in Fig. 5.

From Table 7 and Fig. 5, it can be appreciated that, in contrast to time complexity and execution time, convergence performance does not appear to be related with problem dimension. Convergence seems to have a strong relationship with the particular

| Case study | Head-based HSE (s) | Demand-based HSE (s) |
|------------|--------------------|----------------------|
| Modena     | 0.012              | 0.026                |
| KL         | 0.0088             | 0.45                 |
| Sabsevar   | 0.062              | 103.18               |
| Neyshabur  | 0.061              | 86.69                |
| Torbat     | 0.070              | 84.21                |
| Kashmar    | 0.052              | 20.16                |

Fig. 3 Layouts of case studies
node interconnection of each network. Notwithstanding, there are two interesting appreciations. Firstly, demand-based HSE tends to converge with less iterations. This is observed in all case studies but Kashmar, where both approaches converge with the

![Fig. 4 Structures of $G$ and the matrix of the system of equations for both head-based and demand-based approaches for Kashmar case study. Overlapping sensitive part is highlighted in demand-based $G$.](image)

![Fig. 5 Convergence results](image)
same iterations. The convergence speed difference does not maintain a constant proportion among case studies. Note that, in the case studies where the demand-based approach converges in fewer iterations, the convergence speed difference is not high enough to counteract the slower execution time per iteration of this approach against the head-based approach. Secondly, demand-based HSE convergence appears to be more robust to high uncertainty. The number of iterations needed to converge in the demand-based approach are the same under both the $2\sigma$ and the $3\sigma$ assumptions. In the head-based approach, the iterations needed to converge under the $3\sigma$ assumption increase with respect to the ones needed under the $2\sigma$ assumption in 3 out of the 6 case studies.

4 Conclusions

HSE is a suitable method to estimate the most likely state of a water distribution network. It can overcome some disadvantages of typical flow network resolution models (e.g. GGA or Loop methods) regarding input redundancy and uncertainty without resorting to computationally expensive simulation methods (e.g. Monte Carlo simulations). HSE can be addressed with different optimization criteria, being WLS profusely used in the literature due to its benefits. WLS supports several formulation possibilities depending on the selected set of state variables. Choosing the set of state variables is not straightforward since different sets lead to problems with different dimensions, convergence performance and time complexity. In this work, all possible essential sets of state variables are gathered. Two of them are identified as the most efficient since they are related to the lowest possible dimension of the problem. One of the sets is based on head levels and the other one is based on water demands. The particularities of the so-called head-based and demand-based HSE are presented in this paper together with a computationally efficient formulation. The functioning of both sets is also tackled: their time complexity and convergence performance are addressed.

Both obtained time complexities and execution times clearly favor head-based HSE. Head-based HSE has a more advantageous time complexity, especially regarding preprocessing, $G$ computation and Newton’s method. The obtained execution times per iteration of the tested case studies are much faster in the head-based approach. This is greatly induced by the sparser structure of the matrices involved in the head-based approach. Since time results have been mainly assessed per iteration, the convergence performance (i.e. number of iterations) is also important. Convergence behavior appears to be case specific, although the demand-based approach seems to converge with less iterations, especially for great uncertainties. The lower number of iterations is not able to counteract the worse execution times per iteration.

Finally, it is important to highlight that this paper constitutes an effort to systematically assess the implications of the state variable choice for HSE. According to the results, head-based HSE seems to have a faster performance. It usually needs a few more iterations to converge than demand-based HSE, but its lower per-iteration execution times compensate this difference. Hence, Head-based HSE has been here identified as the most advantageous approach to undertake HSE according to a WLS criterion in real practice.
Author Contribution All authors have contributed to the paper. The first author wrote draft manuscripts and the two other authors reviewed and improved the manuscripts several times.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Grant D1N2019-010896 funded by MCIN/AEI /10.13039/ 501100011033. Grant PID2019-111506RB100 funded by MCIN/AEI /10.13039/ 501100011033. Grant SBPLY/19/180501/000162 funded by Junta de Comunidades de Castilla-La Mancha and ERDF A way of making Europe.

Data Availability Data used in this work can be found in the databases mentioned in Sect. 3.1.

Code Availability Custom code based on MATLAB R2020a has been used.

Declarations

Ethics Approval All work is in compliance with Ethical Standards.

Consent to Participate All authors confirm their co-authorship.

Consent to Publish The authors give their permission to publish.

Conflicts of Interest The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alvarruiz F, Martínez Alzamora F, Vidal AM (2015) Improving the efficiency of the loop method for the simulation of water distribution systems. J Water Resour Plan Manage 141(10):04015019. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000539

Alvarruiz F, Martínez Alzamora F, Vidal AM (2018) Efficient modeling of active control valves in water distribution systems using the loop method. J Water Resour Plan Manage 144(10):04018064. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000982

Andersen JH, Powell RS (2000) Implicit state estimation technique for water network monitoring. Urban Water 2(2):123–130. https://doi.org/10.1016/s1462-0758(00)00005-9

Bargiela A (1984) On-line monitoring of water distribution networks. Durham Theses, Durham University, Durham (UK)

Bargiela A, Hainsworth D (1989) Pressure and flow uncertainty in water systems. J Water Resour Plan Manage 115(2):212–229. https://doi.org/10.1061/(ASCE)0733-9496(1989)115:2(212)

Bazaraa MS, Sherali HD, Shetty CM (2006) Nonlinear programming. Theory and algorithms (3rd ed). John Wiley & Sons Inc, Hoboken (NJ)

Bragalli C, D’Ambrosio C, Lee J, Lodi A, Toth P (2012) On the optimal design of water distribution networks: a practical MINLP approach. Optim Eng 13(2):219–246. https://doi.org/10.1007/s11081-011-9141-7

Bragalli C, Fortini M, Todini E (2016) Enhancing knowledge in water distribution networks via data assimilation. Water Resour Manage 30(11):3689–3706. https://doi.org/10.1007/s11269-016-1372-0

Brodal GS (1996) Worst-case efficient priority queues. At: 7th annual ACM-SIAM symposium on discrete algorithms. Atlanta (GE)

Brodal GS, Lagogiannis G, Tarjan R (2012) Strict Fibonacci Heaps. At: 44th symposium on theory of computing. New York (NY). https://doi.org/10.1145/2213977.2214082
Caro E, Conejo AJ, Mínguez R (2011) A sensitivity analysis method to compute the residual covariance matrix. Electr Power Syst Res 81(5):1071–1078. https://doi.org/10.1016/j.epsr.2010.12.007
Carpentier P, Cohen G (1991) State estimation and leak detection in water distribution networks. Civ Eng Syst 8(4):247–257. https://doi.org/10.1080/02630259108970634
Castillo E, Conejo AJ, Pedregal P, García R, Alguacil N (2002) Formulación y Resolución de Modelos de Programación Matemática en Ingeniería y Ciencia. Universidad de Castilla-La Mancha, Ciudad Real (Spain)
Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introduction to algorithms, 2nd edn. MIT Press and McGraw-Hill, Boston Burr Ridge (IL)
Coulbeck B (1977) Optimisation and modelling techniques in dynamic control of water distributions systems. PhD thesis, University of Sheffield, Sheffield (UK)
Creaco E, Franchini M (2014) Comparison of Newton-Raphson global and loop algorithms for water distribution network resolution. J Hydraul Eng 140(3):313–321. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000825
Cross H (1936) Analysis of flow in networks of conduits or conductors. Engineering Experiment Station, University of Illinois, Urbana (IL)
Devlin K (2004) Sets, functions, and logic: an introduction to abstract mathematics, 3rd edn. Chapman & Hall/CRC Press, Boca Raton (FL)
Díaz S, González J, Mínguez R (2015) Observability analysis in water transport networks: Algebraic approach. J Water Resour Plan Manage 142(4):04015071. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000621
Díaz S, González J, Mínguez R (2016a) Uncertainty evaluation for constrained state estimation in water distribution systems. J Water Resour Plan Manage 142(12):06016004. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000718
Díaz S, Mínguez R, González J (2016b) Stochastic approach to observability analysis in water networks. Ingeniería del Agua 20(3):139–152. https://doi.org/10.4995/ia.2016.4625
Díaz S, Mínguez R, González J (2017) Calibration via Multi-period State Estimation in Water Distribution Systems. Water Resour Manage 31(15):4801–4819. https://doi.org/10.1007/s11269-017-1779-2
Díaz S, Mínguez R, González J (2018a) Topological State estimation in water distribution systems: Mixed-integer quadratic programming approach. J Water Resour Plan Manage 144(7):04018026. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000934
Elhay S, Simpson AR (2011) Dealing with zero flows in solving the nonlinear equations for water distribution networks. J Hydraul Eng 137(10):1216–1224. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000411
Epp R, Fowler AG (1970) Efficient code for steady-state flows in networks. J Hydraul Div 96(HY1):43–56
Golub GH, Van Loan CF (1996) Matrix computations, 3rd edn. Johns Hopkins University Press, Baltimore (MD)
Gómez-Quiles C, Gil HA, de la Villa Jaén A, Gómez-Expósito A (2013) Equality-constrained bilinear state estimation. IEEE Trans Power Syst 28(2):902–910. https://doi.org/10.1109/TPWRS.2012.2215058
Hamam Y, Brameller A (1971) Hybrid method for the solution of piping network. Proc IEE 118(11):1607–1612
Jung D, Kim JH (2018) State estimation network design for water distribution systems. J Water Resour Plan Manage 144(3):05017006. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000862
Kang DS (2009) Real-time demand estimation for water distribution systems. PhD thesis, University of Arizona, Tucson (AZ)
Kang DS, Lansey K (2012) Revisiting optimal water distribution system design: issues and a heuristic hierarchical approach. J Water Resour Plann Manage 138(3):208–217. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000165
Knuth DE (1997) The art of computer programming. Volume 1: Fundamental Algorithms (3rd ed). Addison Wesley Longman, Reading (MA)
Kumar SM, Narasimhan S, Bhallamudi SM (2008) State estimation in water distribution networks using graph-theoretic reduction strategy. J Water Resour Plan Manage 134(5):395–403. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:5(395)
Mair M, Mikovits C, Sengthaler M, Schöff M, Kinzel H, Urich C, Kleidorfer M, Sitzenfrei R, Rauch W (2014a) The application of Web-geographic information system for improving urban water cycle modelling. Water Sci Technol 70(11):1838–1846. https://doi.org/10.2166/wst.2014.327
Mair M, Rauch W, Sitzenfrei R (2014b) Spanning tree based algorithm for generating water distribution network sets by using street network data sets. At: World Environmental & Water Resources Congress. Portland (OR). https://doi.org/10.1061/9780784413548.050
Martin DW, Peters G (1963) The application of Newton’s method to network analysis by digital computer. J Inst Water Eng 17(2):115–129
Pasha FK (2006) Uncertainty analysis and calibration of water distribution quality models. PhD thesis, University of Arizona, Tucson (AZ)
Powell RS (1992) On-line monitoring for operational control of water distribution networks. Durham Theses, Durham University, Durham (UK)
Rossman LA (2000) Epanet 2 users manual. Water Supply and Water Resources Division, US Environment Protection Agency, Cincinnati (OH)
Ruiz E, Díaz S, González J (2022) Potential performance of hydraulic state estimation in water distribution networks. Water Resour Manage 36:745–762. https://doi.org/10.1007/s11269-021-03056-2
Savic DA, Kapelan ZS, Jonkergouw PMR (2009) Quo vadis water distribution model calibration? Urban Water J 6(1):3–22. https://doi.org/10.1080/15730620802613380
Sterling MJH, Bargiela A (1984) Minimum norm state estimation for computer control of water distribution systems. IEE Proc 131(2):57–63. https://doi.org/10.1049/ip-d.1984.0009
Todini E, Pilati S (1988) A gradient algorithm for the analysis of pipe networks. In: Coulbeck B, Orr CH (eds) Computer applications in water supply: Vol 1-systems analysis and simulation. Research Studies Press Ltd, Taunton (UK), pp 1–20
Todini E, Rossman LA (2013) Unified framework for deriving simultaneous equation algorithms for water distribution networks. J Hydraul Eng 139(5):511–526. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000703
Tshehla K, Hamam Y, Abu-Mahfouz A (2017) State estimation in water distribution network: a review. At: IEEE 15th International Conference on Industrial Informatics. Emden(Germany). https://doi.org/10.1109/INDIN.2017.8104953
Vrachimis SG, Timotheou S, Eliades DG, Polycarpou MM (2019) Iterative hydraulic interval state estimation for water distribution networks. J Water Resour Plan Manage 145(1):04018087. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001011
Wang S, Taha AF, Gatsis N, Sela L, Giacomoni M (2021) Probabilistic state estimation in water networks. IEEE Trans Control Syst Technol Early Access. https://doi.org/10.1109/TCST.2021.3066102
Wood DJ, Charles COA (1972) Hydraulic network analysis using linear theory. J Hydraul Div 98(HY7):1157–1170

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.