Iterative hydraulic interval state estimation for water distribution networks

Stelios G. Vrachimis[1], Stelios Timotheou[2], Demetrios G. Eliades[3], and Marios M. Polycarpou[4]

[1]Ph.D. Candidate, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus. Email: vrachimis.stelios@ucy.ac.cy

[2]Research Associate, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus. Email: timotheou.stelios@ucy.ac.cy

[3]Research Associate, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus. Email: eldemet@ucy.ac.cy

[4]Professor, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus. Email: mpolycar@ucy.ac.cy

ABSTRACT

State estimation of hydraulics (i.e. pressure and flows) in water distribution networks is an important tool for efficient and resilient operation. However, hydraulic state estimation is a challenging task in practice due to the scarcity of measurements and the presence of several modeling uncertainties. Standard state estimation techniques may produce unreliable estimates with no information of the estimation error magnitude, especially when historical data are used in place of missing measurements. In this paper, we propose a comprehensive methodology for generating hydraulic state bounding estimates by considering both measurement and parametric uncertainties. The methodology is based on solving the nonlinear interval hydraulic equations using bounding linearization, a technique that restricts the nonlinearities within a convex set, thus converting the problem in a form which is solvable using linear optimization. An iterative procedure improves the bounding linearization, converging to the tightest possible bounds. Simulation results demonstrate that the proposed methodology produces tight state bounds that can replace Monte-Carlo simulations.

Keywords: interval, bounds, optimization, state estimation, water distribution networks

INTRODUCTION

The water industry is being modernized with the installation of sensors for monitoring Water Distribution Networks (WDN) and computer systems to process these data. Integrated platforms are already being developed that algorithmically combine real-time sensor measurements from Supervisory Control and Data Acquisition (SCADA) systems, geographic information systems (GIS), and hydraulic models to provide useful information to the operators. A state estimation algorithm infers the complete system state, such as water flows in pipes, consumer water demands, pressures at nodes and tank levels, using the available measurement set and network equations. A complete view of the network state supports the decision-making process and enables the efficient operation of these systems, improves customer service and enables the early detection of emergency events, thus minimizing their impact. Examples of the use of state estimation in real systems include the use for online burst detection (Okeya et al. 2014) as well as for online modelling (Machell et al. 2010) and control of WDN (Rao and Salomons 2007).

Standard state estimation techniques require a measurement set that makes the system observable, i.e. the sensor number and locale ensure that the system state can be calculated (Bargiela 1985; Nagar and Powell 2000; Díaz et al. 2016). Additionally, the statistical characterization of sensor measurement error is needed to give more weight to measurements originating from more accurate sensors. Then, using a mathematical model of the network, a state estimation algorithm can infer the system state. Many approaches have been proposed to solve the state estimation problem for water systems, such as Kalman Filtering and Weighted Least Squares (WLS), with the latter being the most widely used and varied. The above methods produce a point in state-space and are referred to as *point state estimation* (Powell et al. 1988; Andersen et al. 2001; Kang and Lansey 2009).

State estimation in WDN is a challenging task, mainly due to the scarcity of measurements. In contrast to other large scale engineering systems (e.g. power systems) where the number of measurements guarantees observability, in WDN observability is almost never achievable with the available measurements. Some parts of the WDN may be widely monitored, such as the transport network, however even a single sensor failure could make these parts unobservable (Vrachimis et al. 2016). The large area covered by WDN and the large number of system states is one of the main reasons that an impractical number of sensors must be installed to guarantee observability. A common practice to reduce the complexity, is to skeletonize the network by treating a group of consumers as a single demand point. It is then possible to use *pseudo-measurements*, which are demand estimates determined from population densities and historical data, to complement the missing measurements (Hutton et al. 2014). Recent advances in water demands research have also made possible the higher resolution modeling of water demands, thus reducing the need for skeletonization of networks and increasing the accuracy of state estimation (Avni et al. 2015).

The use of pseudo-measurements may introduce new problems to the state estimation process, as they are highly uncertain and the resulting estimates may deviate significantly from the real system state. This in turn could affect other algorithms which rely on state-estimation, such as feedback control or fault-diagnosis. Efforts have been made to characterize the uncertainty of pseudo-measurements (Bargiela and Hainsworth 1989), but it is improbable that a statistical characterization will be available. Thus, standard state estimation techniques such as WLS may not be capable of producing a reliable measure of the estimation error. Consequently, researchers have tried to combine online estimation of demands with state estimation, in order for the latter to be more accurate (Preis et al. 2011).

Another significant source of uncertainty which complicates WDN state estimation is modelling and parameter uncertainty. Recent works provide explicit expressions for the sensitivity of the state estimation problem to these uncertainties (Díaz et al. 2018). Typically, the network topology is assumed known, especially after the process of skeletonization which simplifies the network graph. However, even when the topology is known, pipe parameters such as length and diameter are rarely known accurately and estimates are used in place. This is especially true for pipe roughness coefficients, which along with length and diameter, are used to calculate the headloss across pipes. This is why, even with an observable sensor configuration, model calibration is required *a priori* or during state estimation for the latter to produce feasible solutions (Gao 2017). Model calibration can be considered as the inverse problem of state estimation, during which estimates of the unknown model parameters are calculated based on measurements (Kapelan et al. 2003; Savic et al. 2009). But even after calibration, it is possible that the calculated parameter set satisfies the constraints imposed by measurements, but deviates from the true parameter set.

Considering the many unknowns and uncertainties in WDN state-estimation, it is evident that accurate state-estimates are difficult to be generated without some kind of trade-off. A practical approach for state estimation in the presence of demand and modeling uncertainty, is *interval state estimation* (Bargiela et al. 2003; Łangowski and Brdys 2007). This approach models the uncertainties on input data as intervals, defined by lower and upper bounds. Then, considering this bounded uncertainty, interval state estimation provides lower and upper bounds on the state estimates, in contrast to point state estimation methods which only provide a single point. Providing a range of values for each state, is often more useful to an operator than providing point estimates which give no indication on their proximity to the true state value. Additionally, having reliable interval state estimation is essential in many methodologies related to event and fault detection such as leakage detection (Pérez et al. 2009), water contamination detection (Eliades et al. 2015) and sensor fault detection

(Vrachimis et al. 2015).

The use of bounds for the representation of measurement uncertainty and the subsequent calculation of state estimate bounds was introduced in (Bargiela and Hainsworth 1989). This idea was further developed in (Brdys and Chen 1994) as the *set-bounded* state estimation problem. The process of calculating bounds for state estimates caused by measurement uncertainty is also referred to as Confidence Limit Analysis which can be solved using different approaches, including Neural Networks (Gabrys and Bargiela 1997), the Error Maximization method (Arsene et al. 2011), the Ellipsoid method and Linear Programming (Bargiela et al. 2003). All these approaches assume a known network model which can be linearized in order to solve the non-linear equations that characterize WDN and provide state bounds based on measurement uncertainty. Few methodologies can guarantee the inclusion of the true state in the bounds based on given uncertainty, while the effect of modeling uncertainty is not considered. Another approach that could incorporate modeling uncertainty is the use of Monte-Carlo Simulations (MCS), where state bound estimates are obtained by randomly generating and evaluating a large number of model parameter sets or realizations (Pasha and Lansey 2010). This approach requires a sufficiently large number of simulations, and even then some cases may not be covered, leading to underestimation of the range of the true state bounds.

In this work we propose a new interval hydraulic state-estimation approach for WDN that considers the combined effect of bounded measurement and modeling uncertainties. The proposed methodology calculates the bounds on state estimates using the nonlinear form of the network equations, by also modeling pressure-dependent demands and background leakages. The nonlinear modeling guarantees that when accurate uncertainty bounds are provided, the bounds on state estimates will include the true system state. This is achieved using bounding linearization, a technique which restricts the nonlinearities within a convex set, thus converting the hydraulic equations in a form where the minimum and maximum of each state can be found using linear optimization. Then, an iterative procedure is followed to minimize the distance between upper and lower state bounds, by improving the bounding linearization at each step and converging to the tightest possible bounds. The contributions of this work are:

- The consideration of both modeling uncertainty, in the form of uncertain parameters, as well as measurement uncertainty in the interval state estimation problem for WDN.
- The development of a novel algorithm that calculates tight hydraulic bounding estimates based on the considered uncertainties.
- The use of the nonlinear form of the hydraulic equations which also considers pressure-dependent demands and background leakages, in order to ensure that the bounding estimates guarantee the inclusion of the true system state if the uncertainties have been accurately represented.

This paper is organized as follows: Section "Problem formulation" formulates the problem of hydraulic state estimation in WDN where the uncertainty on model parameters and measurements is represented by intervals. Section "Iterative Hydraulic Interval State Estimation" presents a methodology to solve this problem based on the Iterative Hydraulic Interval State Estimation (IHISE) algorithm. In Section "Case Studies" this methodology is applied on different benchmark water networks and its performance is assessed.

PROBLEM FORMULATION

The topology of a WDN is modeled by a directed graph denoted as $\mathcal{G} = (\mathcal{N}, \mathcal{L})$. Let $\mathcal{N} = \{1, \cdots, n_n\}$ be the set of all nodes, where $|\mathcal{N}| = n_n$ is the total number of nodes. These represent junctions of pipes, consumer water demand locations, reservoirs and tanks. The unknown quantity associated with nodes is the *hydraulic head*, indicated by h_i. Reservoirs and water tanks that have level sensors installed, can be considered as nodes with known head. We define the set of nodes with unknown head $\mathcal{N}_u = \{1, \cdots, n_u\}$, where $|\mathcal{N}_u| = n_u$ is the number of nodes with unknown head. The set of nodes with known head is defined as $\mathcal{N}_h = \{n_u + 1, \cdots, n_n\}$, where $|\mathcal{N}_h| = n_h$ is the number of nodes with known head and

$\mathcal{N} = \mathcal{N}_u \cup \mathcal{N}_h$. Each node j with unknown head is associated with a water consumer demand at the node location, denoted by $q_{ext,j}$.

Let $\mathcal{L} = \{1, \cdots, n_l\}$ be the set of links, where $|\mathcal{L}| = n_l$ is the total number of links. These represent network pipes, water pumps and pipe valves, with the last two being the main hydraulic control elements in a water network. We define the set of links that represent pipes as $\mathcal{L}_p = \{1, \cdots, n_p\}$, where $|\mathcal{L}_p| = n_p$ is the total number of pipes. We also define the set of links that represent pumps as $\mathcal{L}_{pu} = \{n_p + 1, \cdots, n_l\}$, where $|\mathcal{L}_{pu}| = n_{pu}$ is the total number of pumps. The unknown quantity associated with a link i is the *water flow*, indicated by q_i.

Formulation of hydraulic equations

It is a common practice in WDN to receive sensor measurements of flows, pressures or tank water levels at constant time intervals, which typically range from five minutes to one hour. These sensors may also give an average measurement for the elapsed time interval, thus fast changing dynamics (e.g. pressure transients) are neglected. As a result, standard state estimation in WDN is carried out at steady state, with the system state being recalculated when measurements arrive.

In this work we assume that only lower and upper bounds on measurements are available. The bounds can be derived from real sensor measurements, or from population densities and historical data (pseudo-measurements). The measurement bounds are available at every discrete time step k for all nodal demand outflows and for all tank and reservoir levels. This sensor configuration guarantees the topological observability of the network. Other sensor configurations are also possible, given that they satisfy the topological observability condition, which can be checked using the algorithm in (Díaz et al. 2017). The vector of measured external water demand outflow is indicated by $\bar{q}_{ext}(k) \in \mathbb{R}^{n_u}$. The known head vector, which results from tank and reservoir level measurements, is indicated by $\bar{h}_{ext}(k) \in \mathbb{R}^{n_l}$. The unknown state of the network are the water flows in pipes $\bar{q}(k) \in \mathbb{R}^{n_l}$ and the unknown hydraulic heads at nodes $\bar{h}(k) \in \mathbb{R}^{n_u}$.

The state is calculated using a hydraulic model of a WDN, which is a set of equations derived from the laws of: *1)* conservation of energy and *2)* conservation of mass in the network. In this work we use the *pipe formulation* of these equations as used by (Todini and Pilati 1987), which has been shown to be robust in computer simulations (Rossman 2000). The only dynamic component of these equations are the changing tank levels (Boulos et al. 2006). Because tank levels are assumed to be measured, the resulting hydraulic equations are not dynamic, thus the discrete time notation k is omitted. The formulation of these equations follows.

Conservation of energy equations

Energy in WDN is associated with the head at nodes and when water flows through a network link i which connects two nodes, a flow dependent head function $f_i(q_i)$ describes the change in head. In the case of pipes, energy is dissipated due to friction of water flowing through the pipe, resulting in head-loss between two connected nodes. Head-loss depends on the water flow through the pipe but also on pipe parameters. Each pipe $i \in \mathcal{L}_p$ is characterized by pipe length l_i, pipe diameter d_i and pipe roughness coefficient c_i. All these quantities are used in the empirical Hazen-Williams (H-W) formula (Boulos et al. 2006) to calculate head-loss. The effect of pipe parameters in this formula is aggregated in the H-W resistance coefficient r_i of each pipe, which is a function $f^r_{HW} : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}^+$ of pipe parameters, defined as: $r_i = f^r_{HW}(c_i, d_i, l_i)$. The head-loss across pipe $i \in \mathcal{L}_p$ is then calculated using the H-W formula as follows:

$$f_i(q_i) = r_i \left|q_i\right|^{(\nu-1)} q_i \ , \tag{1}$$

where ν is a constant exponent associated with the H-W formula and q_i is the water flow in pipe i.

Another example of network element are pumps $i \in \mathcal{L}_{pu}$ which are characterized by a head-flow curve, which is used to relate the flow through the pump to the head-gain across the pump, according to each pump specifications. This is given by:

$$f_i(q_i) = -(w_1 - w_2 q_i^{w_3}) \ , \tag{2}$$

where $w_1, w_2, w_3 \in \mathbb{R}$ are coefficients of the pump head-flow curve, while the minus sign indicates that in the case of pumps there is head-gain instead of head-loss.

The energy equations for all the network links, considering elements modeled by (1) and (2), can be written as follows:

$$f_i(q_i) + \sum_{j \in \mathcal{N}_u} (B_{ij}\ h_j) = h_{ext,i}, \quad i \in \mathcal{L}, \tag{3}$$

where:

- h_j is the unknown head of node $j \in \mathcal{N}_u$.
- $B \in \mathbb{R}^{n_l \times n_u}$ is the incidence flow matrix, indicating the connectivity of nodes with links. Element $B_{ij} = +1$ if the direction of link i enters node j; element $B_{ij} = -1$ if the direction of link i leaves from node j; otherwise $B_{ij} = 0$. Nodes with known head are excluded from this matrix.
- $h_{ext,i}$ is the sum of known (measured) heads that appear in each equation $i \in \mathcal{L}$. In vector notation, the known head vector is given by $\bar{h}_{ext} \in \mathbb{R}^{n_l}$.

Conservation of mass equations

The conservation of mass law for a node $j \in \mathcal{N}_u$ is similar to Kirchhoff's current law in electric circuit analysis and can be summarized as follows: the sum of branch water flows from pipes incident to a node j must be equal to the node's external water demand $q_{ext,j}$.

A demand-driven modeling approach assumes that the demand at each node is independent of the pressure at that node. However, this analysis is not valid when power outages, fire fighting, pipe breaks or temporarily closed portions of a WDN lead to pressure-deficit conditions. In those cases, the consumers do not receive the requested demand, thus the modeling of demand is no longer valid and a pressure-dependent demand modeling is recommended. The pressure-demand relationship can be modeled by multiplying the user requested demand $q_{ext,j}$ at node j by the pressure depended function $f_{ext,j}(h_j)$, which is given by (Wagner et al. 1988; Giustolisi and Laucelli 2011; Klise et al. 2017):

$$f_{ext,j}\left(h_j\right) = \begin{cases} 0 & h_j \leqslant H_{\min,j} \\ \left(\frac{h_j - H_{\min,j}}{H_{req,j} - H_{\min,j}}\right)^{0.5} & H_{\min,j} \leqslant h_j \leqslant H_{req,j} \\ 1 & h_j \geqslant H_{req,j} \end{cases} . \tag{4}$$

In (4), $H_{req,j}$ is the head above which the consumer can receive the requested demand $q_{ext,j}$ (depends on node elevation), $H_{min,j}$ is the minimum desired head at node j (depends on node elevation) below which the consumer does not receive any water.

Background leakage flows are also present in real WDN and are modeled as an added demand component at nodes. Leakage flows are pressure-depended and are modeled similarly to pressure-driven demands as follows (Giustolisi et al. 2008):

$$q_{leak,j}(h_j) = \begin{cases} \beta_j (h_j - Z_j)^{\gamma_j}, & h_j - Z_j > 0 \\ 0 & h_j - Z_j \leqslant 0 \end{cases} \tag{5}$$

where Z_j is the elevation of node j, and β_j and γ_j are leakage parameters depending on pipe deterioration and material.

The conservation of mass equations, considering all the nodes of the network, can be written using the incidence flow matrix as follows:

$$\sum_{i \in \mathcal{L}} \left(B_{ij}^{\top}\, q_i\right) = q_{ext,j}\, f_{ext,j}(h_j) + q_{leak,j}(h_j), \quad j \in \mathcal{N}_u. \tag{6}$$

In vector notation, the requested external water demands for all nodes are given by $\bar{q}_{ext} \in \mathbb{R}^{n_u}$ and the leakage flow at nodes by $\bar{q}_{leak} \in \mathbb{R}^{n_u}$. Equations (3) and (6) define the network state, which are the water flows in pipes and hydraulic heads at nodes, indicated by $\bar{x} = [\bar{q}^{\top}\ \bar{h}^{\top}]^{\top}$.

Measurement and parameter uncertainty

As mentioned in the "Introduction" section, we consider sensor measurements, or pseudo-measurements that are uncertain, with a bounded measurement error. We assume that these are available for all nodal demand outflows, as well as for all tank and reservoir levels, guaranteeing an observable sensor configuration. The network topology is available in the hydraulic model, but pipe parameters are only approximately known.

The uncertainties are modeled in this work using intervals, which is equivalent to a uniform probability distribution. For notational convenience, we adopt the convention of denoting intervals in bold. Let $\bar{\boldsymbol{v}} = [\bar{v}^l, \bar{v}^u]$ be a closed interval vector, where $\bar{v}^l$ is the lower bound vector and $\bar{v}^u$ is the upper bound vector, such that: $\bar{\boldsymbol{v}} = \{\bar{v} \in \mathbb{R}^n : v_i^l \leq v_i \leq v_i^u, \forall\, i = \{1, .., n\}\}$, and n is the size of the vector. Note that calculations performed in equations containing intervals require the use of interval arithmetic (Daumas et al. 2009; Kearfott 1996; Moore et al. 2009).

The uncertain requested water demands and the reservoir/tank levels are given by the interval vectors $\bar{\mathbf{q}}_{ext} = \left[\bar{q}_{ext}^{\,l},\ \bar{q}_{ext}^{\,u}\right]$ and $\bar{\mathbf{h}}_{ext} = \left[\bar{h}_{ext}^{\,l},\ \bar{h}_{ext}^{\,u}\right]$ respectively. Note that, measurement bounds that are derived from nodes with actual sensors, do not require to be accompanied by a pressure-dependent function, so $f_{ext,j}\left(h_j\right) = 1$.

We also consider the uncertainty on the head function $f_i\left(q_i\right)$. When this function contains uncertain parameters, these will be modelled as intervals defined by a lower and upper bound, and the head function will be indicated in bold as $\boldsymbol{f}_i\left(q_i\right)$. Uncertainty in pipe parameters is included in the uncertain H-W coefficients r_i. These are calculated using uncertain pipe parameters, which are the roughness coefficients c_i, diameter d_i and length l_i. For a certain pipe i, the uncertain H-W coefficient is given by: $\mathbf{r_i} = \left[r_i^{\,l},\ r_i^{\,u}\right]$. An interval H-W coefficient transforms the pipe headloss function given by (1), into an interval function given by:

$$\boldsymbol{f}_i\left(q_i\right) = \mathbf{r}_i\ |q_i|^{\nu-1} q_i, \quad i \in \mathcal{L}_p \tag{7}$$

Similarly, for i corresponding to a pump with an uncertain pump curve, (2) becomes:

$$\boldsymbol{f}_i(q_i) = -(\mathbf{w}_1 - \mathbf{w}_2 q_i^{\mathbf{w}_3}), \quad i \in \mathcal{L}_{pu}. \tag{8}$$

The uncertainty is considered in the case of leakage modeling, is on the leakage parameters, which can be represented by the intervals $\boldsymbol{\beta}_j = \left[\beta_j^l,\, \beta_j^u\right]$ and $\boldsymbol{\gamma}_j = \left[\gamma_j^l,\, \gamma_j^u\right]$. This will result in an interval leakage function $\boldsymbol{q}_{leak,j}(h_j)$.

As a practical note, a calibration pre-step of network topology parameters is recommended, as it will reduce the uncertainty of these parameters in the sense that their boundaries will be more restrictive. As a result, the bounded state estimates calculated by the IHISE algorithm will be less conservative. In this work the parameter boundaries are assumed constant for all time steps, because these parameters vary very slowly over time. The parameters can be updated whenever a calibration procedure takes place for the network.

Problem definition

The problem of solving the hydraulic equations of a WDN when these contain uncertainty in the form of intervals, is reduced to finding the set of all point solutions for the state $\bar{x} = [\bar{q}^\top \; \bar{h}^\top]^\top$, that satisfy the following systems of equations:

$$\boldsymbol{f}_i\left(q_i\right) + \sum_{j \in \mathcal{N}_u} \left(B_{ij} \; h_j\right) = \boldsymbol{h}_{ext,i}, \qquad i \in \mathcal{L} \tag{9a}$$

$$\sum_{i \in \mathcal{L}} \left(B_{ij}^\top \; q_i\right) = \boldsymbol{q}_{ext,j} \, f_{ext,j}(h_j) + \boldsymbol{q}_{leak,j}(h_j), , \qquad j \in \mathcal{N}_u \tag{9b}$$

Problem (9) is a Nonlinear Interval System of Equations (NISE) and the set of solutions for $\bar{x}$ that satisfy (9) may have a rather complex form that needs to be described with non-linear functions. This is why, in the literature, ‘interval solutions’ are most often considered, with the aim of finding the Interval Hull (IH) solution, i.e. the smallest interval vector $\bar{\mathbf{x}}$ containing all solutions for the NISE. Finding the IH solution to general NISE is a very challenging problem; even for the general Linear Interval System of Equations (LISE), finding the IH is an NP-hard problem. For this reason there are several solutions proposed in the literature that approximate the IH, either with an Outer Interval (OI) solution, which is any interval vector enclosing the IH solution, or with an INner Interval (INI) solution, which is any interval vector that is a subset of the IH solution. Most approaches deal with the problem of finding an OI solution, while the INI solution is used as a measure of the overestimation of the solution (Neumaier and Pownuk 2007; Kolev 2004a). An example of a method for finding the INI solution to a NISE are Monte-Carlo Simulations, where solutions are calculated by randomly generating and evaluating a large number of non-interval equations with parameters within the defined intervals (Eliades et al. 2015). The set of solutions is always a subset of the IH solution.

The literature on finding an OI solution to NISE is limited, but some approaches have been proposed, such as (Kolev 2004b), which uses affine arithmetic to represent the equations and interval linearization to deal with the nonlinearities. However, this approach does not consider interval multiplicative terms in the nonlinear functions, thus cannot be applied to (9). The solution to NISE can also be approached using optimization, as in (Jiang et al. 2008) where the task of solving nonlinear interval number programming problems was investigated. This method, however, does not ensure that the solution is an OI, thus it is not suitable for use with methodologies such as fault detection which require outer bounds on states.

Good approaches in the literature that provide tight OI solutions exist for Linear Interval Systems of Equations (LISE) and are mainly divided in two categories. The first uses interval arithmetic (Daumas et al. 2009; Moore et al. 2009) to find the solution. Due to the fact that when using interval arithmetic to solve LISE, the solution interval is inherently an overestimation, iterative methods are used to approximate the IH solution, such as the Gauss Elimination method, LU decomposition method and the iterative Jacobi method (Zieniuk et al. 2015). The second category formulates LISE as an Interval Linear Programming problem, where intervals can exist both in the objective function and in the constraints (Chinneck and Ramadan 2000; Huang and Cao 2011). This approach is promising since the formulation of the equations as a Linear Program provides the opportunity to add and manipulate constraints to the problem. Additionally, powerful software that solve Linear Programs efficiently exist, which reduce computation time.

ITERATIVE HYDRAULIC INTERVAL STATE ESTIMATION

In this work we propose an iterative method for finding an OI solution of the NISE in (9), named *Iterative Hydraulic Interval State Estimation* (IHISE), which closely approximates the IH solution. This method deals with the nonlinearities in (9) using bounding linearization, which encloses the interval nonlinearities in a convex set and converts (9) into a system of linear inequalities. The resulting linear inequalities are then appropriately formulated into a Linear Program and new bounds on the state variables are calculated. An iterative procedure then approximates the IH solution of (9) by using the new bounds on the states to improve the bounding linearization. Initial bounds on state variables can be defined from physical properties of WDN.

The IHISE algorithm is comprised of five main steps:

1. Find initial bounds on the state variables using physical constraints of the system.
2. Use bounding linearization to bound the nonlinearities in a convex set.
3. Formulate Linear Programs (LPs) for each state using the resulting linear inequalities.
4. Solve one maximization and one minimization LP for each state to produce new upper and lower bounds.
5. Iteratively improve the OI solution of (9) by repeating steps 2 to 4 until convergence of bounds.

Next, the five steps of IHISE are described in detail.

Step 1: Initial bounds on state vector

The first step of the IHISE algorithm is to impose initial bounds on the state vector $\bar{x} = [\bar{q}^\top \; \bar{h}^\top]^\top$ which should be an OI solution of (9). The initial bounds on the unknown head vector $\bar{h}$ are given by the interval vector $\bar{\mathbf{h}}^{(0)} = \left[\bar{h}^{l\,(0)},\ \bar{h}^{u\,(0)}\right]$ and are chosen by considering physical properties of the network. The minimum head vector $\bar{h}^{l\,(0)}$ is set equal to the elevation of each node and the maximum head vector $\bar{h}^{u\,(0)}$ is set equal to the sum of reservoir and pump heads, which is the maximum head that any node in the network can have.

The special structure (9a), in which each equation contains only one flow state q_i, allows us to use the initial bounds on heads $\mathbf{h}^{(0)}$ to find the initial bounds on the flows. We rewrite (9a) as follows:

$$\boldsymbol{f}_i\left(q_i\right) = -\sum_{j \in \mathcal{N}_u} \left(B_{ij}\ \mathbf{h}_j^{(0)}\right) + \mathbf{h}_{ext,i} = \mathbf{y}_i, \tag{10}$$

where $\mathbf{y}_i = \left[y_i^l, y_i^u\right]$ is a known interval derived from the known terms in (10) using interval analysis. The function $\boldsymbol{f}_i\left(q_i\right)$, when $i \in \mathcal{L}_p$, is given by (7). This function is *inclusion isotonic* (Moore et al. 2009) meaning that if $\mathbf{q}_i^1 \subseteq \mathbf{q}_i^2$ then $\boldsymbol{f}(q_i^1) \subseteq \boldsymbol{f}(q_i^2)$. This property enables the derivation of analytical bounds on the unknown pipe flows, by rearranging (10) with respect to q_i, $i \in \mathcal{L}_p$.

In the case of pumps, $\boldsymbol{f}_i(q_i)$ is given by (8). This function is not inclusion isotonic in its whole range, but this property holds in the special case when $q_i > 0$ or $q_i < 0$. This implies that the flow direction in pump links needs to be known *a priori*, which is a valid assumption, since pumps are physically restricted to move water in one direction. Assuming that the flow in pump links is always positive, the bounds on flows q_i, $i \in \mathcal{L}_{pu}$ can be found by rearranging (10) with respect to q_i.

The initial bounds on the flow state vector are denoted by $\bar{\mathbf{q}}^{(0)} = \left[\bar{q}^{l\,(0)}, \bar{q}^{u\,(0)}\right]$. An analytical derivation of these bounds for pipes and pumps is given in Appendix S1 of the Supplemental Data.

Step 2: Bounding linearization of interval nonlinear terms

This step aims at enclosing in a convex set $\mathcal{S}$ the nonlinear uncertain functions that exist in Problem (9). This will allow the formulation of a Linear Program. Problem (9) contains three nonlinear uncertain functions: $\boldsymbol{f}_i(q_i)$, $\boldsymbol{q}_{ext,j}\, f_{ext,j}(h_j)$ and $\boldsymbol{q}_{leak,j}(h_j)$, which are all functions of one bounded variable. The bounds on these variables have been calculated in Step 1, such that for flow variables $q_i \in \left[q_i^l, q_i^u\right]$ and for head variables $h_j \in \left[h_j^l, h_j^u\right]$. The goal is to construct convex sets $\mathcal{S}$ that include all the points of the uncertain functions in the given range, e.g. $\boldsymbol{f}_i(q_i) \in \mathcal{S},\ \forall q_i \in \left[q_i^l, q_i^u\right],\ i \in \mathcal{L}$.

In this work we design the convex sets $\mathcal{S}$ using bounding linearization (Kolev 2004b). This procedure can be used on any uncertain nonlinear function of one bounded variable and it encloses the function between two lines. For example, given the nonlinear uncertain function $\boldsymbol{f}_i(q_i)$ for an interval $q_i \in \left[q_i^l, q_i^u\right]$, a lower line $f_{L,i}^l(q_i) = \lambda_i^l q_i + \beta_i^l$ and an upper line $f_{L,i}^u(q_i) = \lambda_i^u q_i + \beta_i^u$ can be designed such that:

$$f_{L,i}^l(q_i) \leq \boldsymbol{f}_i(q_i) \leq f_{L,i}^u(q_i),\ \forall q_i \in \left[q_i^l, q_i^u\right],\ i \in \mathcal{L} \tag{11}$$

The goal of the bounding linearization procedure is to define the line parameters to minimize the area of the resulting convex set $\mathcal{S}$. A detailed description on how to obtain linearization bounds for each of the nonlinear functions considered can be found in Appendix S2 of the Supplemental Data.

Step 3: Formulation of Linear Program

The bounding linearization of Step 2, yields linear inequality constraints for the interval functions $\boldsymbol{f}_i(q_i)$, $\boldsymbol{q}_{ext,j}\, f_{ext,j}(h_j)$ and $\boldsymbol{q}_{leak,j}(h_j)$. These inequalities can replace these functions in (9) with new variables $\zeta_{e,i}$, $\zeta_{p,j}$ and $\zeta_{l,j}$ respectively, thus transforming these equations into linear inequalities. Bound inequalities also arise on state variables q_i and h_j through Step 1.

The Linear Program will then have the following constraints:

$$h^l_{ext,i} \leq \zeta_{e,i} + \sum_{j \in \mathcal{N}_u} (B_{ij}\ h_j) \leq h^u_{ext,i}, \qquad i \in \mathcal{L} \tag{12a}$$

$$\sum_{i \in \mathcal{L}} \left(B_{ij}^\top\ q_i\right) = \zeta_{p,j} + \zeta_{l,j}, \qquad j \in \mathcal{N}_u \tag{12b}$$

$$\lambda^l_{e,i} q_i + \beta^l_{e,i} \leq \zeta_{e,i} \leq \lambda^u_{e,i} q_i + \beta^u_{e,i}, \qquad i \in \mathcal{L} \tag{12c}$$

$$\lambda^l_{p,j} h_j + \beta^l_{p,j} \leqslant \zeta_{p,j} \leqslant \lambda^u_{p,j} h_j + \beta^u_{p,j}, \qquad j \in \mathcal{N}_u. \tag{12d}$$

$$\lambda^l_{l,j} h_j + \beta^l_{l,j} \leqslant \zeta_{l,j} \leqslant \lambda^u_{l,j} h_j + \beta^u_{l,j}, \qquad j \in \mathcal{N}_u. \tag{12e}$$

$$q^l_i \leq q_i \leq q^u_i \qquad i \in \mathcal{L} \tag{12f}$$

$$h^l_j \leq h_j \leq h^u_j \qquad j \in \mathcal{N}_u \tag{12g}$$

Note that the interval parameters in (9a) are eliminated through the use of their upper and lower bounds to convert them into the inequalities (12a) and (12b). The LP decision variables vector is defined as $\bar{z} = [\bar{x}^\top, \bar{\zeta}^\top]^\top \in \mathbb{R}^{(2n_l + 3n_u)}$ where $\bar{x} = \left[\bar{q}^\top, \bar{h}^\top\right]^\top$ is the state vector and $\bar{\zeta} = [\bar{\zeta_e}^\top, \bar{\zeta_l}^\top, \bar{\zeta_l}^\top]^\top$ is the auxiliary variable vector. Using the constraints (12a) - (12g), two LP problems can be formulated for obtaining lower (LPmin) and upper (LPmax) bounds on each state z_i of the vector $\bar{z}$:

LPmin:

$$\left\{ \begin{array}{ll} \min\limits_{\{\bar{x},\bar{\zeta}\}} & x_i \\ s.t. & (12a) - (12g) \end{array} \right\}$$

LPmax:

$$\left\{ \begin{array}{ll} \max\limits_{\{\bar{x},\bar{\zeta}\}} & x_i \\ s.t. & (12a) - (12g) \end{array} \right\}$$

Step 4: Solution of the linear interval system of equations

The objective of the optimization problem formulated in the previous section is to find the lower and upper bounds on the state vector $\bar{x}$ that satisfy the inequalities (12a)-(12g). To achieve this, a total of $2(n_l + n_u)$ LPs must be solved where each problem will derive either the lower or upper bound of an individual state variable, indicated by x^*_i. This procedure is described in Algorithm 1.

Algorithm 1 Solution of LISE using LP

```
begin
  for i = 1 to n_l + n_u do
      Minimize x_i by solving problem LPmin
      x_i^l = x_i^*
      Maximize x_i by solving problem LPmax
      x_i^u = x_i^*
  end for
  x̄ = [x̄^l, x̄^u]
return x̄
```

Step 5: Iterative solution of the nonlinear interval system of equations

In Algorithm 1 the linearized version of the original problem in (9) is solved. This is an outer interval solution to the nonlinear problem, which guarantees to include the interval hull solution. To find the smallest possible interval $\bar{\mathbf{x}} = \left[\bar{x}^{\,l},\ \bar{x}^{\,u}\right]$ that satisfies (9), an iterative method is used. At each iteration m, the range $\bar{\mathbf{x}}_{bnd}$ in which the optimization algorithm searches for an optimal solution becomes smaller and is redefined as $\bar{\mathbf{x}}_{bnd}^{(m+1)} = \bar{\mathbf{x}}_{bnd}^{(m)} \cap \bar{\mathbf{x}}^{\,(m+1)}$, where $\bar{\mathbf{x}}^{(m+1)}$ are the bounds calculated for the state vector $\bar{x}$ at iteration m. The iterations stop when the bounds on the state vector remain relatively unchanged, i.e. the change is smaller than a small number ϵ. The relative change in bounds is computed as follows:

$$e^{(m)} \triangleq \left|\left(\bar{x}^{u\,(m)} - \bar{x}^{l\,(m)}\right) - \left(\bar{x}^{u\,(m-1)} - \bar{x}^{l\,(m-1)}\right)\right|_1. \tag{13}$$

The overall procedure for the solution of the NISE is described in Algorithm 2. The resulting bounds are an OI solution of these equations that approximate closely the IH solution.

CASE STUDIES

In this section, the IHISE algorithm is applied on different WDN in order to demonstrate the calculated bounds and evaluate the performance of the algorithm. An illustrative example is given in the "Illustrative example" section, where the bounds on different states

Algorithm 2 Iterative solution of NISE

begin

Define initial bounds $\bar{\mathbf{h}}^{(0)}$ using physical properties.
Calculate initial bounds $\bar{\mathbf{q}}^{(0)}$ using the procedure in Step 1.
$\bar{\mathbf{x}}_{bnd}^{(0)} = \left[\bar{\mathbf{q}}^{(0)\top} \ \bar{\mathbf{h}}^{(0)\top}\right]^\top$
$m = 0$
while $e^{(m)} > \epsilon$ **do**
  Bounding linearization of (9) for $\bar{x} \in \bar{\mathbf{x}}_{bnd}^{(m)}$
  Formulate problems **LPmin** and **LPmax**
  Find $\bar{\mathbf{x}}^{(m+1)}$ using **Algorithm 1**
  $\bar{\mathbf{x}}_{bnd}^{(m+1)} = \bar{\mathbf{x}}_{bnd}^{(m)} \cap \bar{\mathbf{x}}^{(m+1)}$
  $m = m + 1;$
end while

return $\bar{\mathbf{x}}^{(m)}$

are shown graphically and compared with bounds obtained by MCS. In the section "Results from benchmark networks" a more extensive analysis of the algorithm is presented, as it is applied on different benchmark networks with varying characteristics. The performance of the algorithm is evaluated by defining appropriate performance metrics and by comparing the IHISE bounds with bounds obtained by MCS. In the simulations we assumed a demand-driven modeling approach with no leakages, which translates into $f_{ext,j}(h_j) = 1$ and $\boldsymbol{q}_{leak,j}(h_j) = 0,\ j \in \mathcal{N}_u$ in Problem (9). This modelling approach allows us to evaluate the performance of the algorithm using the established WDN simulation software EPANET (Rossman 2000).

Illustrative example

The benchmark network "Net1" shown in Fig. 1 provided by EPANET, is used to demonstrate the bounds on hydraulic states produced by the IHISE algorithm. The network parameters are shown in Table 1. Realistic water demand patterns, are assigned at each demand node.

The IHISE algorithm is used to generate bounds on water flows in pipes and hydraulic heads at nodes of the network. The measurement uncertainty is defined as $\pm 5\%$ on the given water demands at nodes, which is the typical error given by manufacturers of water flow meters. Modeling uncertainty is also considered and it is defined as $\pm 5\%$ on pipe Hazen-Williams coefficients. The simulation duration is 24 hours, with a discrete time step of one hour.

Additionally, the same bounds are generated using Monte-Carlo Simulations (MCS) of the network in EPANET. The demands are randomly varied at each simulation within a range of $\pm 5\%$ of the given water demands at nodes. The uncertainty on pipe Hazen-Williams coefficients is achieved by analogously varying pipe lengths, as the Hazen-Williams coefficients are linearly depended on this parameter. Uncertainty on pipe roughness coefficients and pipe diameter can also be considered, but the effect on Hazen-Williams coefficient will not be linear. The maximum and minimum value of each state is saved, defining the upper and lower bounds. The number of simulations is set to 30 000. Note that MCS provide an inner approximation of the bounds on each state and how close they are to the true bounds depends on the number of simulations. Given the possible variations of the same network for the given uncertainty, a sufficiently large number of simulations need to be performed in order for the MCS to converge to the true bounds. Nevertheless, especially in small networks, the MCS bounds can be useful to evaluate the IHISE bounds, which are an outer approximation of the true bounds, by: 1) verifying the correctness of the IHISE bounds by checking if the MCS bounds are always a subset of the IHISE bounds and 2) evaluating the conservativeness of the IHISE bounds by measuring their distance from the MCS bounds.

Simulation results for selected states which reflect the results for all the states are given in Fig. 2. The IHISE bounds are compared with bounds generated using MCS for each state. The figure illustrates that the MCS bounds are a subset of the IHISE bounds, while they are also closely approximated. Note that, the true unknown bounds are enclosed between the IHISE and MCS upper and lower bounds.

Results from benchmark networks

To evaluate the ability of IHISE to compute state bounds, five benchmark networks with varying characteristics are used: “Anytown”, which was used as the basis for the original

"Battle of the Network Optimization Models" (Walski et al. 1987), "Net1", "Net2" and "Net3", which are example networks in EPANET (Rossman 2000), and "ky3" from the Kentucky Infrastructure Authority database of water distribution models (Jolly et al. 2014). The networks and their characteristics are listed in Table 1.

The networks were carefully selected to have multiple varying characteristics in terms of size, topology and types of elements they contain. Varying the network size, i.e. the number of nodes and links, demonstrates the scalability of the IHISE algorithm by considering the performance in networks with different number of *states*, i.e. heads at nodes and water flows at links. Varying the number of reservoirs and tanks, as well as the number of pumps in the network, reveals the ability of the algorithm to deal with these components. The topology of the networks is also considered, specifically the complexity that arises in calculating hydraulic states when the networks contain loops. This is quantified by calculating the *circuit rank* of the network indicated by γ, which is then normalized by the number of links n_l of each network. The resulting metric is defined as the *Loop Ratio*, given by $LR = \gamma/n_l,\ 0 \leq LR < 1$. A value of LR equal to zero means that there are no loops in the network, while LR approaches the value of one in the case of a fully connected graph.

Note that the circuit rank of an undirected graph is defined as the number of independent cycles, or the minimum number of edges that must be removed from the graph to break all its cycles making it into a tree. It is calculated as $r = m - n + c$, where m is the number of edges in the given graph, n is the number of vertices and c is the number of connected components. The circuit rank is also known as the cyclomatic number and is used to indicate the complexity of a program's source code (McCabe 1976).

Monte-Carlo and IHISE Simulations

For each network, a random demand scenario is assigned which produces a feasible solution in EPANET, i.e. there are no negative pressures. Similar to the "Illustrative example" section, the demand and modeling uncertainty is consider equal to $\pm 5\%$. The simulation duration for all networks is 24 hours, with a discrete time step of one hour. The total number of simulation steps is defined as $N_s = 24$.

MCS were performed for all networks using EPANET. The varying parameters are the nodal demands and pipe parameters, in the range defined by the assumed uncertainty. At each simulation, the minimum and maximum value of each state for each time step is saved, thus defining the lower and upper bounds on the state. Additionally, a robust estimate of the state under the considered uncertainties is calculated by taking all the simulated scenarios and calculating the mean value of each state at each time step. This is indicated by $q^{\mu}_{MC,i}(k)$ for flow states and $h^{\mu}_{MC,j}(k)$ for head states and will be referred to as the *MCS state estimate*.

In order to perform an appropriate number of MCS and obtain quantifiable bounds for each network, a stopping criterion for the simulations is imposed, such that the change in all upper and lower flow-state and head-state bounds are less than $\Delta q(m^3/h)$ and $\Delta h(m)$ respectively, for at least 5 000 consecutive simulations. The flows and heads in each network may belong in different value ranges, so Δq and Δh are calculated as a percentage of the absolute mean value of the MCS state estimates. The absolute mean value of flow-states and head-states respectively is given by:

$$\begin{aligned} \mu_q &= \frac{1}{N_s} \sum_{k=1}^{N_s} \left(\frac{1}{n_l} \sum_{i=1}^{n_l} |q^{\mu}_{MC,i}(k)| \right) \\ \mu_h &= \frac{1}{N_s} \sum_{k=1}^{N_s} \left(\frac{1}{n_n} \sum_{j=1}^{n_n} |h^{\mu}_{MC,j}(k)| \right) \end{aligned} \tag{14}$$

The defined accuracies are then calculated as $\Delta q = 1\%(\mu_q)$ and $\Delta h = 1\%(\mu_h)$. Using this approach, it is assumed that the bound accuracy is given by Δq and Δh. Note that while this approach provides a degree of confidence for the accuracy of MCS bounds, it is not guaranteed that the deviation of these bounds from the actual bounds cannot be larger.

Using the IHISE algorithm, bounds for the state of all networks are computed, using the assumed uncertainty. As a technical note, due to the fact that the IHISE algorithm was designed from scratch without the use of other hydraulic solvers, the networks had to satisfy specific conditions in order for the current version of the algorithm to work and be compared with the results from EPANET. The tank levels have to be measurable, so any tanks in the network model are replaced with variable head reservoirs. Control rules that open and close pumps depending on tank levels were removed from the model. Additionally, the head-loss formula should be set to Hazen-Williams. These limitations will be removed in future versions of the algorithm.

Evaluation of bounds mean value

An estimated value of the each state at each time step k is derived using the mean value of the IHISE bounds, which we will refer to as the *IHISE state estimate*. For flow-states this is calculated as $q^{\mu}_{IH,i}(k) = \left(q^{u}_{IH,i}(k) + q^{l}_{IH,i}(k)\right)/2$, and for head-states this is calculated as $h^{\mu}_{IH,j}(k) = \left(h^{u}_{IH,j}(k) + h^{l}_{IH,j}(k)\right)/2$. The IHISE state estimates are compared with the MCS state estimates, by calculating the Absolute Percentage Deviation (APD) of the IHISE state estimates to the MCS state estimates. Note that data from time steps where the MCS state estimate is close to zero are excluded from the evaluation, as they produced large percentages that are not representative of the results. The comparison shows that, for all networks, the Mean APD was less than 1%, while 99% of the time, the APD was less than 8%. This indicates that, when no statistical characterization of the uncertainties is available, the mean value of the IHISE algorithm bounds can be used as a robust estimate of the system state.

Evaluation of bounds

For the evaluation of the IHISE algorithm bounds, we compare the lower bounds $x^{l}_{IH} = [q^{l}_{IH}\ h^{l}_{IH}]$ and upper bounds $x^{u}_{IH} = [q^{u}_{IH}\ h^{u}_{IH}]$ of the IHISE algorithm with the lower bounds $x^{l}_{MC} = [q^{l}_{MC}\ h^{l}_{MC}]$ and upper bounds $x^{u}_{MC} = [q^{u}_{MC}\ h^{u}_{MC}]$ derived from MCS respectively. First, the validity of the IHISE bounds was checked, i.e. $x^{l}_{IH,i}(k) < x^{l}_{MC,i}(k)$ and $x^{u}_{IH,i}(k) > x^{u}_{MC,i}(k)$ for all states i and all time steps k. The test indicated that there were no bound violations for networks "Net1", "Net2" and "Anytown". For "Net3", bound violations occur for two flow-states, in time steps when the MCS flow estimate is less than $10^{-4}(m^3/h)$ and the MCS bound width is less than $10^{-2}(m^3/h)$. For "ky3", bound violations occur in 1.77%

of the time, in specific states where the difference in width of the IHISE bounds and MCS bounds is less than $0.4(m^3/h)$ for flows and less than $0.005(m)$ for heads. The violation magnitude for any state and time step is less than 0.5% of the corresponding MCS state estimate. All the observed bound violations occur in cases where the IHISE bounds are very close to the MCS bounds. This can be explained by the fact that the IHISE algorithm uses a completely independent hydraulic solver than EPANET, thus differences in solutions may exist. Despite this fact, the differences in solutions made apparent by the violations are insignificant and the validity test of the IHISE can be considered succesful if these are attributed to modeling uncertainty which has not been taken into account.

Next, the Absolute Deviation (AD) of the two sets of bounds is evaluated separately for flow-states and head-states, as they are measured in different units. The AD for flow-state lower and upper bound is defined as $e^u_{q,i}(k) = q^u_{IH,i}(k) - q^u_{MC,i}(k)$ and $e^l_{q,i}(k) = q^l_{MC,i}(k) - q^l_{IH,i}(k)$ respectively. Similarly, the AD for head-state lower and upper bounds is defined as $e^l_{h,j}(k)$ and $e^u_{h,j}(k)$ respectively. An illustration of lower and upper bound ADs is shown in Fig. 3. In Table 1 the mean AD for all states and time steps are shown for each network. The results indicate mean errors for upper and lower bounds that are close to the accuracy of MCS, ΔQ and Δh.

The area defined by the IHISE bounds is also an important evaluation metric. Since the duration of simulations is the same for all states, evaluating the area is equivalent to evaluating the width of the bounds. The width of the bounds for each time step is defined as the difference between the upper and lower bound for each time step. For IHISE flow-state bounds, for state i and time step k, the width is given by $w^q_{IH,i}(k) = q^u_{IH,i}(k) - q^l_{IH,i}(k)$. Similarly, the width of IHISE head-state bounds, for state j and time step k, is defined as $w^h_{IH,j}(k)$. The corresponding widths for MCS bounds are denoted by $w^q_{MC,i}(k)$ and $w^h_{MC,j}(k)$. An illustration of bound widths is given in Fig. 3 In Table 1 the mean bound widths for all states and time steps are shown for each network. The mean IHISE bound widths is indicated by w^q_{IH} for flow-states and w^h_{IH} for heads-states, while the MCS bounds are similarly indicated by w^q_{MC} and w^h_{MC}. As shown in Table 1, the difference between IHISE and MCS mean bound width for flow-states in different networks varies from $0.6(m^3/h)$ in "Net2" to $37.5(m^3/h)$ in "Net3". Similarly, the mean bound width for head-states varies from $0.04(m)$ in "Net2" to $2.05(m)$ in "Net1".

In order for the bounds width to give meaningful insight into the accuracy of the algorithm, they must be normalized relative to the absolute mean value of states for each network, i.e. μ_q for flow-states and μ_h for head-states. Using this normalization, the bound width can be viewed as a percentage of each state's uncertainty. The *percent state uncertainty (PSU)* is calculated using the bound width-to-mean ratio as follows:

$$\eta^s_{alg} = \pm \frac{\left(w^s_{alg}/2\right)}{\mu_s} 100\%, \tag{15}$$

where $alg = \{IH, MC\}$ depending on the algorithm used, and $s = \{q, h\}$ depending on the type of state. This will allow the comparison between the calculated state uncertainty and the uncertainty on the network inputs, i.e. the demand and parameter uncertainty. It is recalled that the uncertainty on demands and parameters is defined as a percentage of their estimated value, which was set at $\pm 5\%$ in these simulations.

The average PSU for each network, indicated by $\eta^s_{alg} : alg = \{IH, MC\}, s = \{q, h\}$, is given in Table 1. For flow states, the PSU is, for both methods, close to the $\pm 5\%$ input uncertainty which will be used as a reference point. Typically MCS have slightly less uncertainty and IHISE slightly more, with the exclusion of the looped network "Anytown" where both methods have more uncertainty, and the small network "Net1", where both methods have more. For head-states, the results are much different, as both methods produce much less state uncertainty than the reference point, except in the case of "Net1" where the uncertainty is near $\pm 5\%$. The IHISE average PSU is at the worst case 2 times larger than the MCS average PSU. The worst cases present at the large network "ky3" but also in the highly looped network "Anytown".

For additional insight, the maximum PSU for each network is calculated. This is calculated using the MCS state estimate for each state i and time step k as follows: $\max(\eta^s_{alg}) = \max(w^s_{alg,i}(k)/s^{\mu}_i(k)) : alg = \{IH, MC\}, s = \{q, h\}$. Note that time steps when MCS state estimates have values close to zero, were excluded from the evaluation as they produced large percentages that are not representative of the results. As observed in Table 1, the IHISE maximum PSU is at worst 3 times larger than the maximum PSU obtained by MCS. However, the maximum values of PSU occur in only a few occasions. This is illustrated in Appendix S3 of Supplemental Data, where the distribution of the PSU for IHISE and MCS is plotted for network "ky3".

The different operating scenarios of the networks resulting by the changing demands may also affect the bounds of the IHISE algorithm. To evaluate this factor, the average difference in PSU for all flow-states $\bar{\eta}^q_{IH}(k) - \bar{\eta}^q_{MC}(k)$ at each time step k is calculated. This is then compared to the average nodal demand in the network $\bar{q}_{ext}(k)$ at each time step. By performing correlation analysis, we obtain a correlation of 0.9483, 0.9995, 1.0000, 0.9273 and 0.9975 between these data, for the networks "Net1", "Anytown", "Net2", "Net3" and "ky3" respectively. In Appendix S4 of Supplemental Data, this correlation is illustrated by plotting the average difference in PSU as a function of the average nodal demand for network "ky3". The average difference in PSU follows the pattern of average nodal demands. This can be explained by the fact that the uncertainty on demands is proportional to the demand value, as it was assumed in the design of the simulations, and MCS bounds become less accurate when uncertainty in the network is larger, thus deviating more from the IHISE bounds.

Simulation times

The simulation *time* of either the IHISE algorithm or MCS for a single time step is also evaluated, along with the average *iterations* needed by the IHISE algorithm to solve a single time instance of the specific network and the number of MCS. The simulations were performed on a personal computer with Intel Core i5-2400 CPU at 3.10GHz. Simulation times of the IHISE algorithm are mainly depended on the size of the network, as observed in

Table 1. In Appendix S5 of Supplemental Data, an extrapolation of the simulation times for the IHISE algorithm compared to the simulation times of MCS based on the five networks of this case study is given. The estimated simulation time of the IHISE algorithm is always less than the MCS time with the defined accuracy, while the time difference becomes larger for larger networks.

The simulation time also depends on the complexity of the network, as it is evident in Table 1 from the simulation time of the looped network "Anytown". The IHISE algorithm needs more iterations to converge to a solution for this network compared to the "Net2" which has similar number of states but is less looped. Similarly, more MCS are needed for the looped network "Anytown" than "Net2" for them to converge to a defined bound accuracy.

CONCLUSIONS

In this work the problem of estimating bounds on WDN hydraulic states is addressed. A new methodology is proposed that generates interval state estimates. The proposed *Iterative Hydraulic Interval State Estimation* (IHISE) algorithm generates bounds on hydraulic states of the network, by taking into account the water demand uncertainty and modeling uncertainty in the form of uncertain pipe parameters. The uncertainties are modeled as intervals. The results show that the proposed methodology is able to generate tight bounds on hydraulic states and can be used in place of randomized methods such as Monte-Carlo Simulations (MCS).

The advantage of this methodology over MCS is that the calculated bounds guarantee the inclusion of the true system state, while the iterative nature of the algorithm makes these bounds as tight as possible. An extension of this work is to use the generated bounds for fault diagnosis methods that detect and localize leakages in the network. The proposed methodology can be naturally used with model based fault-diagnosis and robust control methodologies, because many of these rely on the availability of bounding state estimates which are calculated by some knowledge of the system uncertainties. In the case of fault-diagnosis, the bounding state estimates are used to create thresholds, which when violated is an indication of a fault (Puig 2010). Additionally, the bounds on hydraulic states of the network can be used to generate bounds on water quality states, since the dynamics of hydraulic and quality states of a water network are interconnected.

A limitation of this methodology is that it does not model elements whose head function is depended on pressure, such as pressure reduction valves. This is something that will be considered in future work. Other elements that are used in WDN and are not modeled in this work, are pressure control valves, flow control valves etc. Future work will model a variety of additional components to be used with this methodology, and an interval hydraulic state estimation toolkit will be released. Additionally, more extensive simulations on how this methodology deals with pressure-driven demands and pressure-dependent leakages will be provided.

ACKNOWLEDGMENTS

This research is partially funded by the European Union Horizon 2020 programme under grant agreement no. 739551 (KIOS CoE), and by the Interreg V-A Greece-Cyprus 2014-2020 programme, co-financed by the European Union (ERDF) and National Funds of Greece and Cyprus, under the project SmartWater2020.

SUPPLEMENTAL DATA

A detailed description of the derivation of initial bounds on flow states, as described in Step 1 of the IHISE algorithm in Section 3, is provided in Appendix S1. Appendix S2 contains a detailed description of the bounding linearization procedure of Section 3, an illustrative example showing the convergence over successive iterations and illustrative examples of the application of this procedure on nonlinear functions of different hydraulic elements, pressure dependent demand functions and leakage functions. Appendices S3–S5 provide additional simulation data from the Case Study Section. Appendices S1–S5 are available online in the ASCE Library (www.ascelibrary.org).

REFERENCES

Andersen, J. H., Powell, R. S., and Marsh, J. F. (2001). "Constrained state estimation with applications in water distribution network monitoring." *International Journal of Systems Science*, 32, 807–816.

Arsene, C. T., Al-Dabass, D., and Hartley, J. (2011). "Confidence limit analysis of water distribution systems based on a least squares loop flows state estimation technique." *Proceedings of 2011 UKSim 5th European Symposium on Computer Modeling and Simulation*, IEEE, 94–101.

Avni, N., Fishbain, B., and Shamir, U. (2015). "Water consumption patterns as a basis for water demand modeling." *Water Resources Research*, 51(10), 8165–8181.

Bargiela, A. (1985). "An algorithm for observability determination in water-system state estimation." *IEE Proceedings D (Control Theory and Applications)*, Vol. 132, 245–250.

Bargiela, A. and Hainsworth, G. D. (1989). "Pressure and flow uncertainty in water systems." *Journal of Water Resources Planning and Management*, 115(2), 212–229.

Bargiela, A., Pedrycz, W., and Tanaka, M. (2003). "A study of uncertain state estimation." *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans.*, 33(3), 288–301.

Boulos, P. F., Lansey, K. E., and Karney, B. W. (2006). *Comprehensive water distribution systems analysis handbook for engineers and planners*. American Water Works Association.

Brdys, M. and Chen, K. (1994). *Joint state and parameter estimation of dynamic water supply systems with unknown but bounded uncertainty*. John Wiley & Sons, Inc.

Chinneck, J. W. and Ramadan, K. (2000). "Linear programming with interval coefficients." *The Journal of the Operational Research Society*, 51(2), 209–220.

Daumas, M., Lester, D., and Munoz, C. (2009). "Verified real number calculations: A library for interval arithmetic." *IEEE Transactions on Computers*, 58(2), 226–237.

Díaz, S., González, J., and Mínguez, R. (2016). "Observability analysis in water transport networks: Algebraic approach." *Journal of Water Resources Planning and Management*, 142(4), 04015071.

Díaz, S., Mínguez, R., and González, J. (2017). “Topological observability analysis in water distribution systems.” *Journal of Water Resources Planning and Management*, 143(5), 06017001.

Díaz, S., Mínguez, R., González, J., and Savic, D. (2018). “Explicit expressions for state estimation sensitivity analysis in water systems.” *Journal of Water Resources Planning and Management*, 144(4).

Eliades, D. G., Stavrou, D., Vrachimis, S. G., Panayiotou, C. G., and Polycarpou, M. M. (2015). “Contamination event detection using multi-level thresholds.” *Proceedings of 13th Computing and Control for the Water Industry Conference, CCWI 2015*, Vol. 119, 1429–1438.

Gabrys, B. and Bargiela, A. (1997). “Integrated neural based system for state estimation and confidence limit analysis in water networks.” *Proceedings of European Simulation Symposium ESS'96*, 398–402.

Gao, T. (2017). “Pipe roughness estimation in water distribution networks using head loss adjustment.” *Journal of Water Resources Planning and Management*, 143(5), 04017007.

Giustolisi, O., Kapelan, Z., and Savic, D. (2008). “Algorithm for automatic detection of topological changes in water distribution networks.” *Journal of Hydraulic Engineering*, 134(4), 435–446.

Giustolisi, O. and Laucelli, D. (2011). “Water distribution network pressure-driven analysis using the enhanced global gradient algorithm (EGGA).” *Journal of Water Resources Planning and Management*, 137(6), 498–510.

Huang, G. H. and Cao, M. F. (2011). “Analysis of solution methods for interval linear programming.” *Journal of Environmental Informatics*, 17(2), 54–64.

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L., and Savic, D. A. (2014). “Dealing with uncertainty in water distribution system models: A framework for real-time modeling and data assimilation.” *Journal of Water Resources Planning and Management*, 140, 169–

183.

Jiang, C., Han, X., and Liu, G. P. (2008). “A sequential nonlinear interval number programming method for uncertain structures.” *Computer Methods in Applied Mechanics and Engineering*, 197(49), 4250–4265.

Jolly, M. D., Lothes, A. D., Sebastian Bryson, L., and Ormsbee, L. (2014). “Research database of water distribution system models.” *Journal of Water Resources Planning and Management*, 140(4), 410–416.

Kang, D. and Lansey, K. (2009). “Real-time demand estimation and confidence limit analysis for water distribution systems.” *Journal of Hydraulic Engineering*, 135(10), 825–837.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2003). “Multiobjective sampling design for water distribution model calibration.” *Journal of Water Resources Planning and Management*, 129(6), 466–479.

Kearfott, R. B. (1996). “Interval computations: Introduction, uses, and resources.” *Euromath Bulletin*, 2(1), 95–112.

Klise, K. A., Hart, D., Moriarty, D. M., Bynum, M. L., Murray, R., Burkhardt, J., and Haxton, T. (2017). “Water network tool for resilience (WNTR) user manual.” *Report no.*, Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA (United States).

Kolev, L. V. (2004a). “A method for outer interval solution of parametrized systems of linear interval equations.” *Reliable Computing*, 10, 227–239.

Kolev, L. V. (2004b). “An improved interval linearization for solving nonlinear problems.” *Numerical Algorithms*, 37, 213–224.

Łangowski, R. and Brdys, M. (2007). “Monitoring of chlorine concentration in drinking water distribution systems using an interval estimator.” *International Journal of Applied Mathematics and Computer Science*, 17(2), 199–216.

Machell, J., Mounce, S. R., and Boxall, J. B. (2010). “Online modelling of water distribution systems: A UK case study.” *Drinking Water Engineering and Science*, 3(1), 21–27.

McCabe, T. (1976). "A complexity measure." *IEEE Transactions on Software Engineering*, SE-2(4), 308–320.

Moore, R. E., Kearfott, R. B., and Cloud, M. J. (2009). *Introduction to interval analysis*. Society for Industrial and Applied Mathematics.

Nagar, A. and Powell, R. (2000). "Observability analysis of water distribution systems under parametric and measurement uncertainty." *Joint Conference on Water Resource Engineering and Water Resources Planning and Management*, Vol. 104, 1–10.

Neumaier, A. and Pownuk, A. (2007). "Linear systems with large uncertainties, with applications to truss structures." *Reliable Computing*, 13(2), 149–172.

Okeya, I., Kapelan, Z., Hutton, C., and Naga, D. (2014). "Online burst detection in a water distribution system using the kalman filter and hydraulic modelling." *Procedia Engineering*, 89, 418–427.

Pasha, M. F. K. and Lansey, K. (2010). "Effect of parameter uncertainty on water quality predictions in distribution systems-case study." *Journal of Hydroinformatics*, 12(1), 1.

Pérez, R., Puig, V., Pascual, J., Peralta, A., Landeros, E., and Jordanas, L. (2009). "Pressure sensor distribution for leak detection in Barcelona water distribution network." *Water Science & Technology: Water Supply*, 9(6), 715.

Powell, R. S., Irving, M. R., Sterling, M. J. H., and Usman, A. (1988). "A comparison of three real-time state estimation methods for on-line monitoring of water distribution systems." *Computer applications in water supply: vol. 1—systems analysis and simulation*, Research Studies Press Ltd., 333–348.

Preis, A., Whittle, A. J., Ostfeld, A., and Perelman, L. (2011). "An efficient hydraulic state estimation technique using reduced models of urban water networks." *Journal of Water Resources Planning and Management*, 137(4), 343–351.

Puig, V. (2010). "Fault diagnosis and fault tolerant control using set-membership approaches: Application to real case studies." *International Journal of Applied Mathematics and Computer Science*, 20(4), 619–635.

Rao, Z. and Salomons, E. (2007). “Development of a real-time, near-optimal control process for water-distribution networks.” *Journal of Hydroinformatics*, 9(1), 25.

Rossman, L. A. (2000). “Epanet 2: Users manual.

Savic, D. A., Kapelan, Z. S., and Jonkergouw, P. M. (2009). “Quo vadis water distribution model calibration?.” *Urban Water Journal*, 6, 3–22.

Todini, E. and Pilati, S. (1987). “A gradient algorithm for the analysis of pipe networks.” *Proceedings of International Conference on Computer Applications for Water Supply and Distribution*.

Vrachimis, S. G., Eliades, D. G., and Polycarpou, M. M. (2015). “The backtracking uncertainty bounding algorithm for chlorine sensor fault detection.” *Proceedings of 13th Computing and Control for the Water Industry Conference, CCWI 2015*, Vol. 119, 613–622.

Vrachimis, S. G., Eliades, D. G., and Polycarpou, M. M. (2016). “Real-time hydraulic interval state estimation for water transport networks : a case study.” *Proceedings of 14th Computer Control for Water Industry Conference, CCWI 2016*.

Wagner, J. M., Shamir, U., and Marks, D. H. (1988). “Water distribution reliability: Simulation methods.” *Journal of Water Resources Planning and Management*, 114(3), 276–294.

Walski, T. M., Brill, E. D., Gessler, J., Goulter, I. C., Jeppson, R. M., Lansey, K., Lee, H., Liebman, J. C., Mays, L., Morgan, D. R., and Ormsbee, L. (1987). “Battle of the Network Models: Epilogue.” *Journal of Water Resources Planning and Management*, 113(2), 191–203.

Zieniuk, E., Kapturczak, M., and Kużelewski, A. (2015). “Solving interval systems of equations obtained during the numerical solution of boundary value problems.” *Computational and Applied Mathematics*, 35(2), 629–638.

List of Tables

Networks:	**Net1**	**Anytown**	**Net2**	**Net3**	**ky3**
States	24	63	76	216	646
Loop Ratio	0.23	0.49	0.13	0.19	0.26
Junctions	9	19	35	92	269
Reservoirs	1	3	0	2	3
Tanks	1	0	1	3	3
Pipes	12	40	40	117	366
Pumps	1	1	0	2	5
Flow-states:					
$\boldsymbol{\mu_q}$ (m^3/h)	551.98	75.93	13.33	469.02	42.12
$\boldsymbol{\Delta q}$ (m^3/h)	5.52	0.76	0.13	4.69	0.42
$\boldsymbol{e_q^u}$ (m^3/h)	5.77	6.49	0.32	19.01	1.60
$\boldsymbol{e_q^l}$ (m^3/h)	6.82	6.54	0.32	20.04	1.62
$\boldsymbol{w_{MC}^q}$ (m^3/h)	32.55	10.85	0.96	42.64	2.91
$\boldsymbol{w_{IH}^q}$ (m^3/h)	42.75	23.54	1.56	80.15	6.08
$\boldsymbol{\eta_{MC}^q}$ (%)	±2.95	±7.15	±3.59	±4.55	±3.46
$\boldsymbol{\eta_{IH}^q}$ (%)	±3.87	±15.50	±5.86	±8.54	±7.22
$\mathbf{max}(\boldsymbol{\eta_{MC}^q})$ (%)	±19.33	±78.62	±23.12	±95.71	±62.53
$\mathbf{max}(\boldsymbol{\eta_{IH}^q})$ (%)	±37.86	±205.80	±40.52	±190.35	±190.07
Head-states:					
$\boldsymbol{\mu_h}$ (m)	63.06	42.92	43.69	52.89	48.04
$\boldsymbol{\Delta h}$ (m)	0.52	0.37	0.41	0.50	0.49
$\boldsymbol{e_h^u}$ (m)	1.34	0.06	0.02	0.74	0.08
$\boldsymbol{e_h^l}$ (m)	1.35	0.07	0.02	0.61	0.09
$\boldsymbol{w_{MC}^h}$ (m)	6.10	0.11	0.03	1.49	0.29
$\boldsymbol{w_{IH}^h}$ (m)	8.16	0.24	0.07	2.76	0.46
$\boldsymbol{\eta_{MC}^h}$(%)	±4.84	±0.13	±0.04	±1.41	±0.30
$\boldsymbol{\eta_{IH}^h}$ (%)	±6.47	±0.28	±0.08	±2.61	±0.48
$\mathbf{max}(\boldsymbol{\eta_{MC}^h})$ (%)	±12.99	±0.29	±0.10	±1.82	±1.71
$\mathbf{max}(\boldsymbol{\eta_{IH}^h})$ (%)	±17.44	±0.61	±0.21	±3.60	±2.35
Times:					
MCS Number	5849	28993	21805	13977	32695
MCS (min)	1.71	3.84	4.36	11.74	34.49
IHISE (min)	0.01	0.12	0.08	0.96	13.72
IHISE Iterations	7.44	13.48	9.00	14.56	16.40

TABLE 1. Results of the IHISE algorithm on benchmark networks.

List of Figures

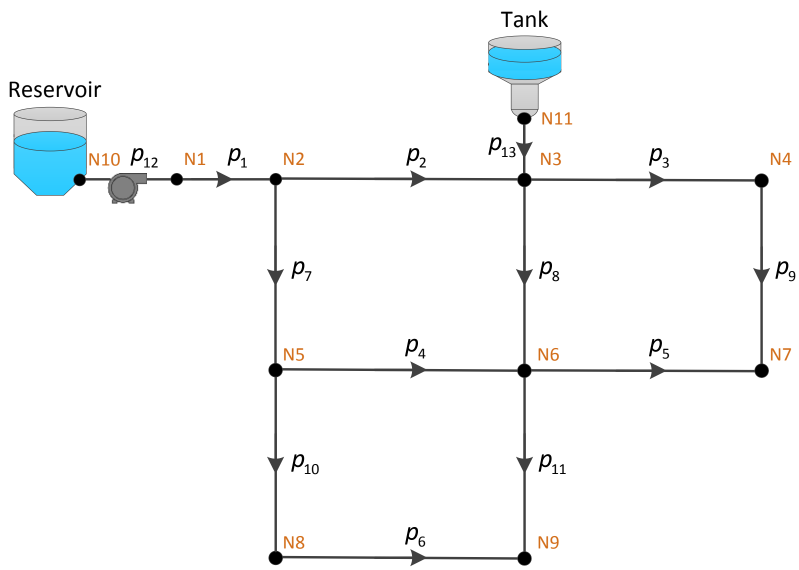

Fig. 1. The benchmark network "Net1", on which the IHISE algorithm is demonstrated.

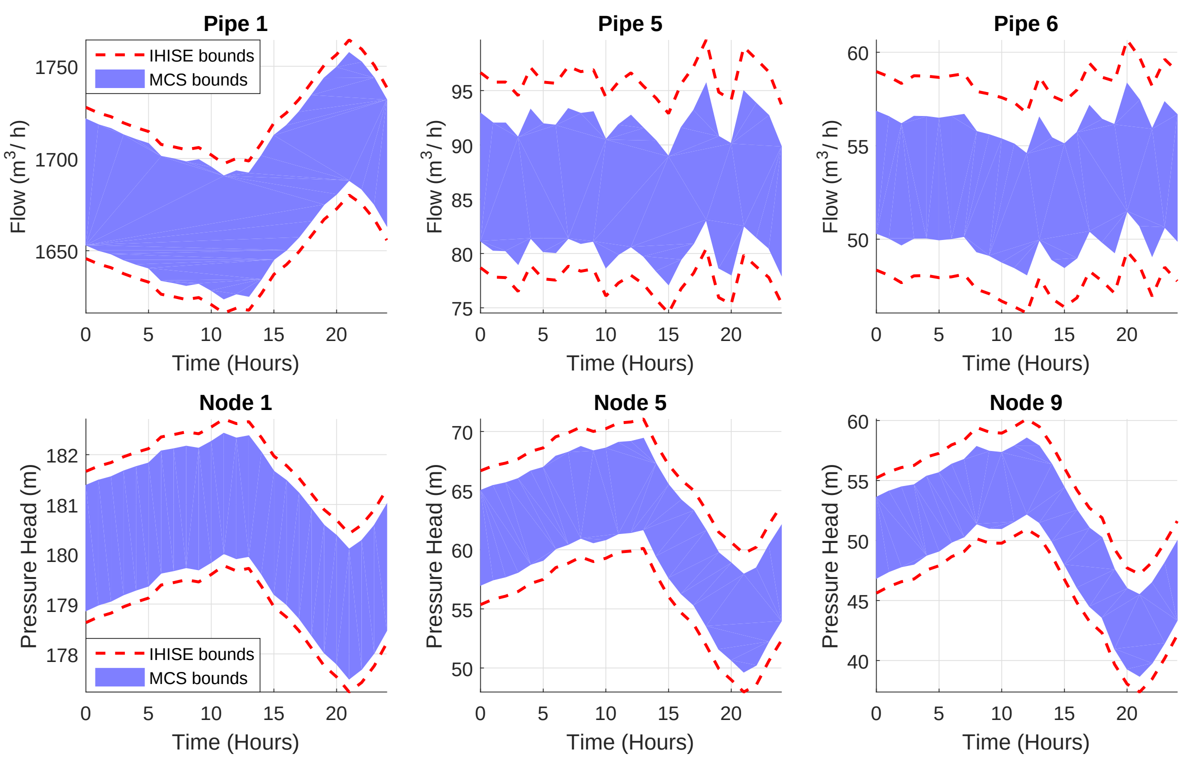

Fig. 2. Comparison of selected pipes water flow bounds (above) and selected nodes hydraulic head bounds (below), generated by Monte-Carlo Simulations (blue solid area) and the IHISE algorithm (red dashed lines).

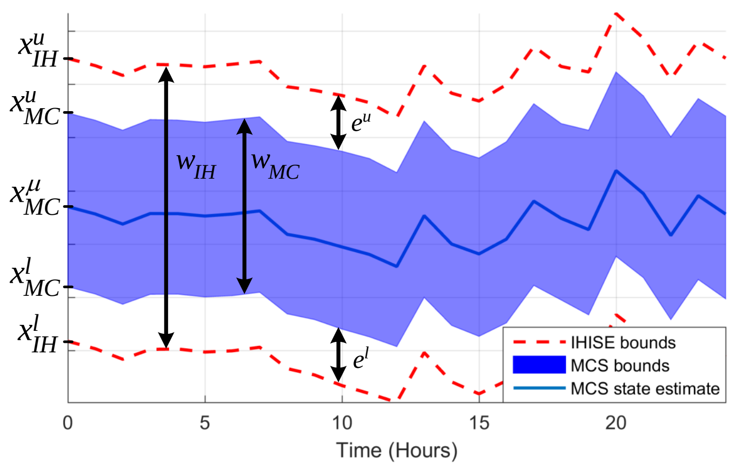

Fig. 3. Illustration of bound evaluation parameters.