Water distribution network calibration using enhanced GGA and topological analysis

O. Giustolisi and L. Berardi

ABSTRACT

The calibration of hydraulic models of water distribution networks (WDN) is of preeminent importance for their analysis and management. It is usually achieved by solving a constrained optimization problem based on some priors on decision variables and the demand-driven simulation of the entire network, given the observations of some hydraulic status variables (i.e. typically nodal heads and sometimes pipe flows). This paper presents a framework to perform the calibration of pipe hydraulic resistances considering two main issues: (i) the enhancements of WDN simulation models allowing us to simplify network topology with respect to serial nodes/trunks and/or to account for a more realistic representation of distributed demands and (ii) a different formulation of the calibration problem itself.

Depending on the available measurements, the proposed calibration strategy reduces the hydraulic simulation model size and can permit the decomposition of the network. On the one hand, such a procedure allows for numerical and computational advantages, especially for large size networks. On the other hand, it allows a prompt analysis of observability of calibration decision variables based on actual observations and might help identifying those pipes (i.e. hydraulic resistances) which are more important for the whole network behaviour.

Key words | calibration, global gradient algorithm, simulation, water distribution networks

O. Giustolisi (corresponding author)
L. Berardi
Civil and Environmental Engineering Dept.,
Technical University of Bari,
v. E. Orabona 4,
70123 Bari,
Italy
E-mail: *o.giustolisi@poliba.it*

ABBREVIATIONS AND NOTATION

The following symbols are used in the paper:

$\overline{\mathbf{A}}_{p,n}$ = general topological matrix of the WDN model;
$\mathbf{A}_{p,n}$, $\mathbf{A}_{n,p}$,
$\mathbf{A}_{p,0}$ = topological incidence sub-matrices of the WDN model;
$\mathbf{A}_{p,p}$ = diagonal matrix of the WDN model;
$\mathbf{B}_{p,p}$ = diagonal matrix of the WDN model;
$\mathbf{d}_{n,1}$ = vector of nodal demands in WDN model;
$\mathbf{d}_{n,1}^{orig}$ = vector of demands at non-serial nodes in the original WDN topology;
$\mathbf{d}_{n,1}^{serial}$ = vector of demands at serial nodes in the original WDN;
$d_{k,j}$ = demand of the jth serial node of the kth pipe;
$\mathbf{D}_{p,p}$ = derivative diagonal matrix used in GGA or EGGA;
$D_{k,i}$ = internal diameter of the ith trunk and kth pipe;
e = absolute (equivalent sand) pipe roughness;
EGGA = Enhanced Global Gradient Algorithm;
EPS = Extended Period Simulation;
$f_{k,\mathrm{i}}$ = friction factor of the ith trunk of the kth pipe;
$f_{k,\mathrm{i}}^{\infty}$ = fully turbulent flow friction factor of the ith trunk of the kth pipe;
$\mathbf{F}_{n,1}$ = temporary matrix used in GGA or EGGA;
$F_{k,I}$ = parameter for friction head losses through the ith trunk;
GGA = Global Gradient Algorithm;

doi: 10.2166/hydro.2010.088

g = gravitational acceleration;
$\mathbf{H}_{n,1}$ = vector of total network heads (i.e. internal nodes);
$\mathbf{H}_{0,1}$ = vector of total fixed (i.e. known) network heads;
$\mathbf{H}_x$ = vector of x measured total nodal heads;
i = index for serial trunks;
j = index for serial nodes/demands;
$iter$ = counter in iterative search of the GGA or EGGA;
k = index for pipes;
Ke = equivalent roughness;
$K_{k,i}^{ml}$ = minor loss coefficient of the ith trunk along of the kth pipe;
$K_{k,i}$ = unitary pipe hydraulic resistance of the ith trunk and kth pipe;
$K_{k,i}^{\infty}$ = unitary hydraulic resistance of rough fully turbulent flow of the ith trunk of the kth pipe;
$K_{k,\infty}$ = unitary hydraulic resistance of fully rough turbulent flow of the kth pipe;
$L_{k,i}$ = length of the ith trunk within the kth pipe;
L_k = total length of the kth pipe;
m_k = number of serial nodes of the kth pipe;
n = head loss equation exponent;
n_0 = total number of known heads;
n_p = total number of unknown flows;
n_n = total number of unknown heads;
n_k = number of calibration decision variables;
$O.F.$ = objective function;
$obj1$, $obj2$, $obj3$ = objective functions;
$\mathbf{P}_{p,1}$ = vector of total demand distributed along pipes;
P_k = total distributed demand of the kth pipe;
$Q_{k,i}$ = flow rate in the ith trunk of the kth pipe;
Q_k = flow rate of the kth pipe;
$\mathbf{Q}_{p,1}$ = vector of pipe flow rates;
$\mathbf{Q}_y$ = vector of y measured pipe flow rates;
$\mathbf{R}_{p,1}$ = vector of pipe hydraulic resistances;
$R_{k,\infty}$ = hydraulic resistance of the kth pipe of turbulent flow;
$r_{k,i}^{I}, r_{k,i}, \omega_{k,i}, H_{k,i}, \gamma_{i,}, on$ = parameters of the ith pump within the kth pipe.
x = number of measured total nodal heads;
y = number of measured pipe flow rates;
α_k = lumping coefficients of the kth pipe;
ΔH_k = head loss between terminal nodes of the kth pipe;
$\mathbf{\Lambda}_{n,p}$, $\mathbf{\Lambda}_{n,p}^{new}$ = matrix of α_k and (1-α_k) to generate lumped nodal demands;
ϵ_k = pipe hydraulic resistance correction factor of the kth pipe;
ϕ_k = ratio between $K_{k,\infty}^{cal}$ and $K_{k,\infty}$ used in case study 2;
WDN = Water Distribution Network;
$|.|$ = absolute value;
$\|.\|_2$ = Euclidean/quadratic distance;
$(\,)^{\mathrm{T}}$ = vector/matrix transpose operator;
$(\,)^{-1}$ = matrix inverse operator;

INTRODUCTION

The calibration of a water distribution network (WDN) is essential to perform reliable model simulations for maintenance and/or operational purposes. WDN model calibration consists of determining various model parameters that, when input into a hydraulic simulation model, will yield a reasonable match between measured and predicted pressures and flows in the network (Shamir & Howard 1968). Several approaches have been proposed so far that reflect different issues of WDN model calibration and resort to progressively increasing computing capacities. Walski (1983) first reported that the selection of the parameters to calibrate should be performed by considering field observations corresponding to more than one flow rate, while knowing pump pressures, tank elevations and valve settings corresponding to that time. In the opposite case (when just a single set of observations is used) model calibration would be just an error compensation and the model will give poor results if compared with other observations. Similar to other works (Rahal *et al.* 1980; Bhave 1988) the methodology proposed by Walski was trial-and-error and proposed two important criteria to locate WDN monitoring: (i) pressure should be monitored close to high demand locations and (ii) on the perimeter of the skeletonized network (i.e. far from water sources). Such criteria were aimed at reducing the propagation of errors from observations to the calibrating parameters.

Other authors (Ormsbee & Wood 1986; Boulos & Wood 1990) faced WDN model calibration by solving a set of equations representing both mass and energy balance conservation (in steady-state conditions) and constraints represented by available observations. Due to the need for an even-determined system of equations, such methodologies hypothesized a number of parameters (usually pipe roughness only) equal to the number of measurements. This eventually led to pipe grouping for the calibration of pipe roughness, as already proposed by Walski (1983) and further investigated by Mallick *et al*. (2002). Ormsbee (1989) considered extended-period simulation (EPS) of the network to calibrate the roughness of pipes, nodal demands and hydraulic grades at sources and pressure regulating devices. The same author applied some explicit constraints on minimum and maximum bounds of the parameters.

Although using different optimization approaches, other methodologies (Pudar & Liggett 1992; Datta & Sridharan 1994; Reddy *et al*. 1996) tried to minimize the differences between observed and predicted nodal heads, flows and tank levels by using least-squares function types. Savic & Walters (1995) successfully solved the problem of pipe roughness calibration by using genetic algorithms followed by Kapelan *et al*. (2003), while Lingireddy & Ormsbee (1999) used a similar approach for the demand adjustment factor accounting for EPS analysis.

The issue of including noise in input data was faced in Reddy *et al*. (1996) that reported a sensitivity analysis of parameters to be calibrated. Greco & Del Giudice (1999) further emphasized the inclusion of uncertainty in WDN calibration by assuming that some prior roughness coefficients can be estimated based on engineering knowledge of the WDN. Accordingly, the objective function was formulated as a sum of squared differences between the model-predicted and the *a priori* estimated pipe friction coefficients.

The quantification of uncertainty in nodal demand has been investigated in several works (Bargiela & Hainsworth 1989; Xu & Goulter 1996; Gargano & Pianese 2000) and considered in WDN calibration by resorting to many different mathematical approaches. For example, the analysis of the covariance matrix of the parameter estimates through the first-order second-moment method was proposed by Bush & Uber (1998) within the context of the *D*-optimal sampling design problem for WDN calibration. A similar statistical approach was used by Lansey *et al*. (2001). Kapelan *et al*. (2007) used the Metropolis algorithm within a Bayesian approach to achieve the probability distributions of pipe roughness. In a recent work Alvisi & Franchini (2010) proposed the calibration of pipe roughness by using grey numbers which allow for representing uncertainty through intervals, without specifying any probability distribution.

Finally, Ozawa (1986), Carpentier & Cohen (1991, 1993) and Todini (1999) studied the problem of observability. In particular, Ozawa, and Carpentier & Cohen addressed the problem of topological observability by graph theory using the spamming tree concept. The observable variables are those for which the inverse problem of calibration is well-posed. The unobservable variables are those for which the problem is ill-posed. Thus, the unobservability of the network can occur for numerical (parametric) and/or graph-theoretic (topological) reasons. Todini converted the WDN model formulation into a linear estimation problem for which a Kalman filter approach was developed. It was emphasized that a unique set of steady-state data is not sufficient to guarantee the network observability for looped systems even if all the nodal heads and demands are assumed as known. Thus, the use of several independent sets of steady-state observations of the hydraulic system (such as, for example, the use of the EPS) is mandatory. In addition, Todini emphasized that the topological observability is mainly dominated by the network topology when available observations of nodal heads is scarce. This is consistent with the studies by Ozawa (1986) and Carpentier & Cohen (1991, 1993).

Almost every work mentioned above ends up with the conclusion that reliable WDN calibration strongly depends on the location of available measurements which should guarantee the topological observability of the parameters to be calibrated when some independent sets of observations are available. Nonetheless, a commonly adopted approach is to solve the inverse problem (i.e. estimate WDN parameters in order to find a reasonable matching between simulated and observed WDN status) by considering the simulation of the entire WDN and then drawing conclusions about the actual observability of decision variables (e.g. even in terms of sensitivity analysis).

STRATEGY OF WDN ANALYSIS FOR PIPE RESISTANCE CALIBRATION

The procedure proposed here for calibration is based on the following elements.

- The use of Enhanced GGA (EGGA) in order to simplify the network with respect to serial nodes while correctly accounting for energy balance. This allows both numerical and computational advantages, especially for the calibration of large size networks and/or when EPS is performed. In addition, the use of EGGA permits the assumption of any demand pattern (included the "average" hypothesis of uniformly distributed demands along pipes when information about the actual connections/demands is not available) by correcting the possible systematic energy balance error due to representation of demands as concentrated withdrawals in pipe terminal nodes, which is adopted in the classical simulation models (Giustolisi & Todini 2009).
- The proposal of a different formulation of the calibration problem based on either available observations or priors which allows a prompt analysis of pipe observability and might help identifying those pipes (i.e. hydraulic resistances) which are most important for reproducing the whole network behaviour. Moreover, such a strategy might allow for further reducing the simulation model size during the calibration process. For example, it is demonstrated that the network can be skeletonized with respect to unmonitored components of the network which are not only its branched sections. The procedure is consistent with the graph manipulation (deletion of arcs for which the flows are known) operated by Ozawa (1986) and Carpentier & Cohen (1991, 1993). The main outcomes confirm the previous findings of Todini (1999) and Carpentier & Cohen (1991, 1993) about the importance of the network topology for observability of pipe hydraulic resistances.

ENHANCED GLOBAL GRADIENT ALGORITHM

The steady-state simulation of a network of n_p pipes with unknown discharges/flows, n_n nodes with unknown heads (internal nodes) and n_0 nodes with known heads (tank levels, for example) can be formulated in the following nonlinear and linear system of equations based on energy and mass balance, respectively:

$$\begin{array}{lcl} \mathbf{A}_{p,p}\mathbf{Q}_{p,1} + \mathbf{A}_{p,n}\,\mathbf{H}_{n,1} & = & -\mathbf{A}_{p,0}\mathbf{H}_{0,1} \\ \mathbf{A}_{n,p}\mathbf{Q}_{p,1} & = & \mathbf{d}_{n,1} \end{array} \quad (1)$$

$\mathbf{Q}_{p,1}$ is the $[n_p,1]$ column vector of unknown pipe flows, $\mathbf{H}_{n,1}$ is the $[n_n,1]$ column vector of unknown nodal heads, $\mathbf{H}_{0,1}$ is the $[n_0,1]$ column vector of known nodal heads, $\mathbf{d}_{n,1}$ is the $[n_n,1]$ column vector of demands lumped at nodes, $\mathbf{A}_{p,n} = \mathbf{A}^T_{n,p}$ and $\mathbf{A}_{p,0}$ are topological incidence sub-matrices of size $[n_p,n_n]$ and $[n_p,n_0]$, respectively, derived from the general topological matrix $\overline{\mathbf{A}}_{p,n} = [\mathbf{A}_{p,n}\,|\,\mathbf{A}_{p,0}]$ of size $[n_p,n_n+n_0]$ as defined in Todini & Pilati (1988), and $\mathbf{A}_{p,p}$ is a diagonal matrix whose elements are given by the entry-wise or Hadamard product $\mathbf{R}_{p,1}|\mathbf{Q}_{p,1}|$, $\mathbf{R}_{p,1}$ being the vector of the pipe hydraulic resistances. Thus, the nonlinear mathematical problem of network simulation has unknowns ($\mathbf{Q}_{p,1}$; $\mathbf{H}_{n,1}$) and its boundary conditions are ($\mathbf{R}_{p,1}$; $\mathbf{d}_{n,1}$; $\mathbf{H}_{0,1}$). It is noteworthy that $\mathbf{R}_{p,1}$ is an asset state variable which can be considered invariable among different observations (unless they are protracted over a long time interval). For this reason WDN calibration is commonly referred to $\mathbf{R}_{p,1}$ while $\mathbf{d}_{n,1}$ and $\mathbf{H}_{0,1}$ are dynamically varying during time, although it is possible to use them as state variables to be estimated by means of an inverse problem based on the simulation model. For this purpose, the vector $\mathbf{d}_{n,1}$ is assumed known as well as network topology ($\overline{\mathbf{A}}_{p,n}$) and water sources levels ($\mathbf{H}_{0,1}$) during the single steady-state simulation.

WDN calibration is referred herein to the issue of finding the n_p unitary pipe hydraulic resistances, $K_{k,\infty} = R_{k,\infty}/L_k$ (with $k = 1,\ldots,n_p$), where $R_{k,\infty}$ are pipe resistances related to rough fully turbulent flow ($k = 1, \ldots, n_p$). Calibrating $K_{k,\infty}$ is actually more correct than referring to pipe internal roughness only since $K_{k,\infty}$ encompasses all uncertainties surrounding relative roughness ($Ke = e/D$) and diameter D. In particular, the dependence on the power 5 of D (from the Darcy–Weisbach friction equation) hints that a small uncertainty in pipe internal diameter could result in a large uncertainty on $K_{k,\infty}$ which would be neglected if Ke only was calibrated. Note that uncertainty in internal diameter is a plausible effect for aged pipes due to encrustation/corrosion, mainly as a consequence of water quality (Vassiljev *et al.* 2009).

The Global Gradient Algorithm (GGA) (Todini & Pilati 1988) as reported in Equation (2) has been assumed as the reference simulation model to introduce its expanded formulation in the Enhanced GGA (EGGA) (Giustolisi 2010; Giustolisi & Todini 2009):

$$\begin{aligned}
\mathbf{B}_{p,p}^{iter} &= (\mathbf{D}_{p,p}^{iter})^{-1}\mathbf{A}_{p,p}^{iter} \\
\mathbf{F}_{n,1}^{iter} &= -\mathbf{A}_{n,p}(\mathbf{Q}_{p,1}^{iter} - \mathbf{B}_{p,p}^{iter}\mathbf{Q}_{p,1}^{iter}) + \mathbf{d}_{n,1} + \mathbf{A}_{n,p}(\mathbf{D}_{p,p}^{iter})^{-1}(\mathbf{A}_{p,0}\mathbf{H}_{0,1}) \\
\mathbf{H}_{n,1}^{iter+1} &= -\left(\mathbf{A}_{n,p}(\mathbf{D}_{p,p}^{iter})^{-1}\mathbf{A}_{p,n}\right)^{-1}\mathbf{F}_{n,1}^{iter} \\
\mathbf{Q}_{p,1}^{iter+1} &= (\mathbf{Q}_{p,1}^{iter} - \mathbf{B}_{p,p}^{iter}\mathbf{Q}_{p,1}^{iter}) - (\mathbf{D}_{p,p}^{iter})^{-1}(\mathbf{A}_{p,0}\mathbf{H}_{0,1} + \mathbf{A}_{p,n}\mathbf{H}_{n,1}^{iter+1})
\end{aligned} \tag{2}$$

where *iter* is a counter of the iterative solving algorithm and $\mathbf{D}_{p,p}$ is a diagonal matrix whose elements are the derivatives of the head loss function with respect to $\mathbf{Q}_p$.

EGGA framework

As recently demonstrated by Giustolisi & Todini (2009), one drawback of the classical WDN simulation models is the assumption of nodal demands $\mathbf{d}_{n,1}$ without accounting for actual demand distribution along the pipes which in turn might cause coarse errors in pipe resistance calibration due to errors in the energy balance equation of the model system. Giustolisi & Todini (2009) accounted for such error considering a dimensionless correction factor ϵ_k in the energy balance equations:

$$\Delta H_k = K_{k,\infty}(1+\epsilon_k)Q_k|Q_k|^{n-1}L_k \tag{3}$$

where Q_k is the pipe flow; n is the exponent of the adopted head loss formula, L_k is the pipe length and ΔH_k is the head loss between terminal nodes of the kth pipe. Berardi *et al.* (2010) and Giustolisi (2010) generalized the original formulation to account for the actual flow regime through pipes and provided formulations of the pipe hydraulic resistance correction factor ϵ_k.

This work shows how to use the EGGA also in order to simplify network topology with respect to serial nodes/trunks. This is an additional optional feature of the EGGA which can be used during WDN calibration to correctly represent energy balance (Giustolisi & Todini 2009) while providing network simplification.

Without losing the generality of the presentation, let's consider the pipeline between nodes A and B in the upper part of Figure 1. It comprises four serial trunks having lengths L_i and unitary hydraulic resistances K_i ($i=1, \ldots, 4$), three serial nodes (i.e. 1, 2 and 3) joining two trunks each and two non-serial nodes (i.e. A and B) joining three or more trunks. As usual in WDN modeling practice, demands are originally assigned at both serial (i.e. d_1, d_2 and d_3) and non-serial nodes (i.e. $d_A{}^{orig}$ and $d_B{}^{orig}$). The EGGA formulation allows us to represent such a configuration as a unique pipe k (between non-serial nodes A and B) having length L_k equal to the total pipeline length, unitary hydraulic resistance $K_{k,\infty}=K_1$ by convention, and demands d_A and d_B lumped at non-serial nodes as reported in the lower part of Figure 1. The choice of fractions α_k and $(1-\alpha_k)$ of total serial nodal demands ($P_k=d_1+d_2+d_3$) that are attributed to non-serial nodes (A and B) is not unique and, in general, $\alpha_k \neq 0.5$ for non-symmetric serial demand deployment (Giustolisi 2010). The hydraulic resistance correction factor ϵ_k reported in the head loss expression of lower part of Figure 1 takes charge of demand deployment through serial nodes in terms of correction of the energy balance to preserve the original one.

This way, the nodes of the new simplified topology are the non-serial nodes of the original network joining three or more pipes or being the terminal node of a single pipe. All serial

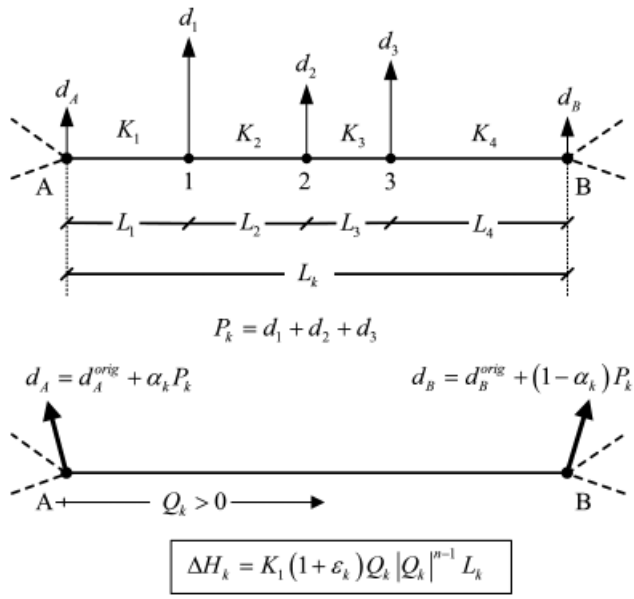

Figure 1 | EGGA representation – serial nodes.

nodes of the original topology, joining just two pipes, disappear in the new network configuration.

Such a simplification eventually results in a lower dimension of the topological matrices of system (1) and even faster simulation runs, especially for real large size networks. This is more useful when multiple WDN simulations are performed, like in trial-and-error calibration procedures.

In the general case of m_k serial nodes supplying demands $d_{k,j}$ along the kth pipe ($j = 1, \ldots, m_k$), the vector of demands distributed along the pipes of the simplified network topology is

$$\mathbf{P}_{p,1} = \left[P_1, ..., P_k, ..., P_{n_p}\right]^T \text{ with } P_k = \sum_{j=1}^{m_k} d_{k,j} \tag{4}$$

where the size of $\mathbf{P}_{p,1}$ is $[n_p, 1]$. The vector of nodal demands $\mathbf{d}_{n,1}$ in the new simplified topology can be obtained from the original nodal demands as follows:

$$\mathbf{d}_{n,1} = \mathbf{d}_{n,1}^{orig} + \mathbf{d}_{n,1}^{serial} \tag{5}$$

where $\mathbf{d}_{n,1}{}^{orig}$ is the vector of demand concentrated at non-serial nodes in the original network; thus, the size of vector $\mathbf{d}_{n,1}{}^{orig}$ is the same as vector $\mathbf{d}_{n,1}$ in the simplified topology. $\mathbf{d}_{n,1}{}^{serial}$ is a vector containing the demands of original serial nodes to be lumped at non-serial nodes; it is computed as

$$\mathbf{d}_{n,1}^{serial} = \mathbf{\Lambda}_{n,p}^{new}\, \mathbf{P}_{p,1} \tag{6}$$

where $\mathbf{\Lambda}_{n,p}{}^{new}$ is built from the typological matrix $\mathbf{A}_{n,p}{}^{new}$ of the new simplified network by substituting into each kth column -1 and $+1$ by α_k and $(1-\alpha_k)$, respectively. Actually, the matrix $\mathbf{A}_{n,p}{}^{new}$ is obtained from the original topological matrix $\mathbf{A}_{n,p}$ by eliminating the rows representing serial nodes and after merging together columns representing serial trunks.

In this work the EGGA is described without using the pipe hydraulic resistance correction factor ϵ_k (as reported in Giustolisi (2010) and Appendix A), but proposing a new alternative formulation. Although it is not as compact as the original one, it serves to clarify all the contributions of head loss through each kth pipe after eliminating the m_k serial nodes, as exemplified in Figure 1. Equation (7) reports the head loss through the kth pipe considering friction head losses and the potentially existing minor loss devices (valves) and pumps into each ith serial trunk:

$$\Delta H_k = \sum_{i=1}^{m_k+1} Q_{k,i}\left(R_{k,i}^{\infty} F_{k,i} |Q_{k,i}|^{n-1} + K_{k,i}^{ml} |Q_{k,i}| + r_{k,i}^{I} |Q_{k,i}|^{\gamma_i - 1}\right) - H_{k,i}^{I}. \tag{7}$$

The terms in brackets account for every head loss contributions by using parameters $F_{k,i}$ for friction head losses, $K_{k,i}{}^{ml}$ for minor head losses and $r_{k,i}{}^{I}$ for pumps (whose speed factor and three parameters of pump curve are $\omega_{k,i}$, $H_{k,i}$, $r_{k,i}$ and γ_i – see Appendix A for further details). Relevant expressions are reported below:

$$R_{k,i}^{\infty} = K_{k,i}^{\infty} L_{k,i}; \quad K_{k,i}^{\infty} = \frac{8 f_{k,i}^{\infty}}{g\pi^2 D_{k,i}^5}; \quad F_{k,i} = \frac{f_{k,i}(\mathrm{Re}, \mathrm{Ke})}{f_{k,i}^{\infty}};$$
$$on = \left[r_{k,i}(Q_{k,i}) \geq 0\right]; \quad H_{k,i}^{I} = on \cdot \omega_{k,i}^2 H_{k,i}; \quad r_{k,i}^{I} = on \cdot \omega_{k,i}^{2-\gamma_i} r_{k,i}. \tag{8}$$

In Equation (8) $D_{k,i}$ is trunk diameter, Re is Reynolds number, $f_{k,i}{}^{\infty}$ and $f_{k,i}$ are friction factors of the ith trunk under fully rough turbulent flow regime and any Re value, respectively; on is a Boolean variable which accounts for the pump installation direction.

Then EGGA can be obtained from the GGA (see Equation (2)) by conveniently writing matrices $\mathbf{A}_{p,p}$, $\mathbf{D}_{p,p}$ and $\mathbf{B}_{p,p}$ and considering that $H^{I}_{k,i}$ are known terms of the energy balance equations:

$$A_{p,p}(k,k) = \frac{1}{Q_k}\left[\sum_{i=1}^{m_k+1} Q_{k,i}\left(R_{k,i}^{\infty} F_{k,i} |Q_{k,i}|^{n-1} + K_{k,i}^{ml} |Q_{k,i}| + r_{k,i}^{I} |Q_{k,i}|^{\gamma_i - 1}\right)\right]$$
$$D_{p,p}(k,k) = \sum_{i=1}^{m_k+1}\left[\left(\frac{R_{k,i}^{\infty}}{f_{k,i}^{\infty}} \frac{df_{k,i}(\mathrm{Re}, \mathrm{Ke})}{dQ_{k,i}} Q_{k,i} + n R_{k,i}^{\infty}, F_{k,i}\right)|Q_{k,i}|^{n-1} + 2K_{k,i}^{ml}|Q_{k,i}| + \gamma_i r_{k,i}^{I} |Q_{k,i}|^{\gamma_i - 1}\right]$$
$$B_{p,p}(k,k) = \frac{A_{p,p}(k,k)}{D_{p,p}(k,k)} \tag{9}$$

where flows $Q_{k,i}$ through the ith serial trunk of the kth pipe are computed as follows, by accounting for the demands of serial nodes $d_{k,j}$ up to the ith trunk (i.e. $\Sigma d_{k,i}$):

$$Q_{k,i} = Q_k + \alpha_k P_k - \Sigma d_{k,i} \text{ with } \Sigma d_{k,i} = \sum_{j=1}^{i-1} d_{k,j} \text{ and } \Sigma d_{k,1} \triangleq 0. \tag{10}$$

It is worth observing that such an EGGA formulation holds also for $m_k = 0$, thus falling into the classical GGA

representation without serial trunks (i.e. null hydraulic resistance correction factor α_k).

Furthermore, for serial trunks distributing any water (i.e. $d_{k,j}=0$ for every j along the kth pipe, so $P_k=0$) $Q_k=Q_{k,1}$ and Equation (7) becomes

$$\Delta H_k = Q_{k,1}\left[\left(\sum_{i=1}^{m_k+1} R_{k,i}^{\infty}F_{k,i}\right)|Q_{k,1}|^{n-1}+\left(\sum_{i=1}^{m_k+1} K_{k,i}^{ml}\right)|Q_{k,1}| + \left(\sum_{i=1}^{m_k+1} r_{k,i}^{I}\right)|Q_{k,1}|^{\gamma_i-1}\right]-\sum_{i=1}^{m_k+1} H_{k,i}^{I} \quad (11)$$

which is the formulation for serial pipes, minor losses and pumps without serial demands.

Some remarks on using EGGA for WDN calibration

From the previous subsection it is evident that the EGGA strategy allows reducing the size of the hydraulic simulation model (i.e. hydraulic state variables, namely pipe discharges and nodal heads explicitly represented in the network topology, which is the size of the model) without reducing the number of unitary hydraulic resistances to be calibrated. Moreover, the elimination of serial nodes/trunks performed in EGGA does not introduce new uncertainties beyond those already existing in the original network.

For the sake of simplicity and without losing the generality of the discussion, let's consider a pipe having one intermediate node, without pumps and minor losses. Suppose that the two trunks have different hydraulic resistances R_1 and R_2, so that the total head loss of the entire pipe is: $\Delta H = R_1Q_1^2 + R_2Q_2^2$. If the pipe flows are known (i.e. by the difference between flow entering the pipe and the demand from the intermediate node) and the total head loss ΔH is known (i.e. from two pressure measurements at the terminal nodes), the resistances R_1 and R_2 cannot be predicted. In fact, the solutions are one infinity (i.e. the inverse problem is ill-posed and unobservability holds). On the one hand, from a global perspective, the prediction of ΔH along the entire pipe is sufficient to analyze the behavior of the remaining network. On the other hand, from a local viewpoint, it is necessary to have additional information/observations to analyze the hydraulic behavior along the pipe. Such information might result from additional measurement points (i.e. at an intermediate node), from some grouping strategies (e.g. presuming some relations between R_1 and R_2 based on the material and aging for adjacent trunks) or using multiple independent sets of measurements in EPS (as suggested by Todini (1999)). Thus, the number of hydraulic resistances to be calibrated in the example remains two even after eliminating the intermediate node by EGGA.

Consequently, EGGA simplification of the topological representation of the hydraulic system emphasizes that the measurement points in the non-serial nodes are more useful in observing the global network behavior; while observations in serial (internal) nodes are more helpful for improving local pipe observability. In such a sense, the EGGA allows modelling the whole network behaviour by considering the most important pipes/links as those between non-serial nodes. This is somehow consistent with the need of proceeding during calibration from coarser to finer network meshes when nodal pressures are scarce (Todini 1999). Thus, it is possible to preliminarily identify the main pipe resistances using a very coarse mesh (i.e. simplified WDN topology), and then proceed to the analysis of finer meshes with additional internal measurement points, independent sets of observations and/or using some prior assumptions (e.g. grouping of similar pipes).

In addition, it can be argued that the simulation of the simplified network (made up of the most important links only) results in an averaging effect of all uncertainties surrounding unitary hydraulic resistances ($K_{k,i}{}^{\infty}$), lengths ($L_{k,i}$) and/or demands of the internal (serial) nodes. In fact, according to Equation (7) the head loss ΔH_k is obtained by summing the products between unitary hydraulic resistances, lengths and a power of serial trunk discharges (i.e. dependent on serial node demands $d_{k,j}$ – see Equation (10)). Thus, if one or more of these values are uncertain together with pumps and minor losses' parameters, the EGGA averages these uncertainties in energy and mass balance equations. For the sake of clarity Equation (12) reports energy and mass balance equations for the new simplified network:

$$\begin{aligned} \mathbf{A}_{p,p}^{new}\mathbf{Q}_{p,1} + \mathbf{A}_{n,p}^{new}\mathbf{H}_{n,1} &= -\mathbf{A}_{p,0}^{new}\mathbf{H}_{0,1} \\ \mathbf{A}_{n,p}^{new}\mathbf{Q}_{p,1} &= \mathbf{d}_{n,1}^{orig} + \mathbf{\Lambda}_{n,p}^{new}\mathbf{P}_{p,1} \end{aligned} \quad (12)$$

where the size of matrices (i.e. subscripts n and p) refers to the number of (non-serial) nodes and pipes of the new simplified network topology.

As a final remark, the simplification based on EGGA has some consistencies with the graph manipulation (contraction of arcs for which nothing is measured) operated by Ozawa (1986) and Carpentier & Cohen (1991, 1993) for observability studies.

WDN MODEL CALIBRATION FORMULATION

The classical way to calibrate a WDN model is to search those boundary conditions that minimize the distance between the hydraulic status computed through WDN simulation and some measurements (nodal pressures and/or flows) (Shamir & Howard 1968). The scheme for the calibration of pipe hydraulic resistances proposed herein consists of the following key aspects:

- different formulations of the WDN simulation model induced by available observations which are consistent with graph manipulations adopted to study topological observability (Ozawa 1986; Carpentier & Cohen 1991, 1993);
- potential decomposition of the network into sub-systems that can be separately simulated with a further reduction of the simulation problem size;
- analysis of topological observability of the resulting components which can also be useful to draw recommendations on sampling design, consistent with the need of separately analyzing each connected graph (i.e. a single component) as in Ozawa (1986) and Carpentier & Cohen (1991, 1993).

It is worth observing that such a calibration strategy is actually independent from the EGGA representation of demands and/or simplification of the network reported in the previous section.

For ease of presentation, the proposed WDN calibration strategy is reported here by studying the case of some available observations of nodal heads $\mathbf{H}_x$ as separate from the case of some measurements of pipe flows $\mathbf{Q}_y$. A unique calibration framework, where both types of measurements might be available, is provided at the end of this section.

Some H_x are known

From a simulation viewpoint, nodes related to the observation of $\mathbf{H}_x$ can be regarded as tank nodes since their head is known. Thus, the related nodal head is moved to the right-hand side of the mathematical system (1) (i.e. known terms). From an uncertainty analysis perspective, those observations can be considered like other measured boundary conditions (e.g. tank levels). The only difference from real tanks is represented by the nodal demands which require mass balance equations to be verified at the x nodes (i.e. the outflow of those nodes is constrained to be the nodal demands $\mathbf{d}_x$), see Figure 2. Such x mass balance equations are then excluded from Equations (1) and (2) in order to preserve the balance between the number of unknown hydraulic parameters and equations as well as the symmetrical properties of the simulation model in (2) (to be solved by either EGGA or GGA). Figure 2 (right) reports in grey the tank added at node 2 of the network to be simulated and the constraint on mass balance as a grey-dashed arrow representing the original nodal demand.

This way, the balance of equations/unknowns remains in equilibrium and the x mass balance equations could be used for determining at most x pipe resistances. Considering n_k decision variables (e.g. pipe unitary hydraulic resistances), the condition $n_k \leq x$ is necessary to have the chance of a well-posed calibration problem (topological observability of the

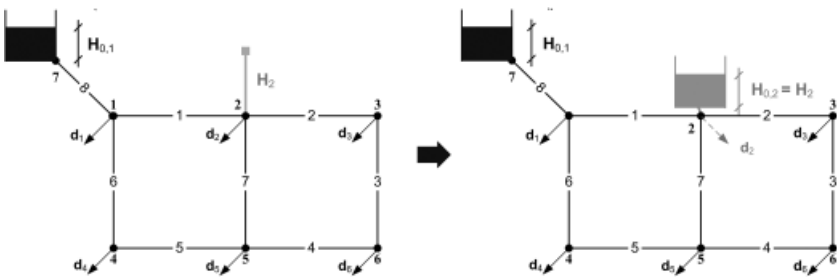

Figure 2 | Network simplification for known nodal heads.

network), bearing in mind that it also depends on network topology and the location of the observation points (Todini 1999). The condition $n_k \leq x$ can be achieved by increasing the number of equations x using more than one steady-state observation of the hydraulic system (e.g. in EPS) or by reducing the number of decision variables n_k (e.g. by grouping them based on some similarities). In both cases, the topological observability of all the decision variables remains mandatory to guarantee network observability.

Thus, the simulation of WDN to be adopted for calibration purposes is based on the iterative solution of the following equations (formally similar for GGA and EGGA):

$$\begin{aligned}
&\mathbf{B}_{p,p}^{iter} = \left(\mathbf{D}_{p,p}^{iter}\right)^{-1}\mathbf{A}_{p,p}^{iter}\\
&\mathbf{F}_{n-x,1}^{iter} = -\mathbf{A}_{n-x,p}\left(\mathbf{Q}_{p,1}^{iter} - \mathbf{B}_{p,p}^{iter}\mathbf{Q}_{p,1}^{iter}\right) + \mathbf{d}_{n-x,1}\\
&\qquad + \mathbf{A}_{n-x,p}\left(\mathbf{D}_{p,p}^{iter}\right)^{-1}\left(\mathbf{A}_{p,0+x}\mathbf{H}_{0+x,1}\right)\\
&\mathbf{H}_{n-x,1}^{iter+1} = -\left(\mathbf{A}_{n-x,p}\left(\mathbf{D}_{p,p}^{iter}\right)^{-1}\mathbf{A}_{p,n-x}\right)^{-1}\mathbf{F}_{n-x,1}^{iter}\\
&\mathbf{Q}_{p,1}^{iter+1} = \left(\mathbf{Q}_{p,1}^{iter} - \mathbf{B}_{p,p}^{iter}\mathbf{Q}_{p,1}^{iter}\right)\\
&\qquad - \left(\mathbf{D}_{p,p}^{iter}\right)^{-1}\left(\mathbf{A}_{p,0+x}\mathbf{H}_{0+x,1} + \mathbf{A}_{p,n-x}\mathbf{H}_{n-x,1}^{iter+1}\right)
\end{aligned} \tag{13}$$

where $\dim(\mathbf{A}_{n\text{-}x,p}) = [n_n - x,\ n_p]$, $\dim(\mathbf{A}_{p,0+x}) = [n_p,\ n_0 + x]$ and $\dim(\mathbf{H}_{0+x,1}) = [n_0 + x,\ 1]$. The decision variables are those $K_{k,i}{}^{\infty}$ that minimize, for example, the following objective function:

$$O.F. = \min_{K_{k,i}^{\infty}}\left\{\sum\left|\mathbf{A}_{x,p}\mathbf{Q}_{p,1} - \mathbf{d}_{x,1}\right|\right\} \tag{14}$$

with $\dim(\mathbf{A}_{x,p}) = [x,\ n_p]$. Such an approach allows accounting for demand uncertainty while the error on mass balance is the performance indicator for the calibration. In other words, the values $\mathbf{d}_x$ represent a sort of prior on demand at x nodes and the uncertainty surrounding the remaining nodal demands (and other boundary conditions) affects pipe discharges at these nodes (i.e. $\mathbf{A}_{x,p}\mathbf{Q}_{p,1}$) and is definitely expressed as the distance from such priors.

It is noteworthy that minimizing mass balance at observed nodes (as in Equation (14)) confirms the previous findings of Walski (1983) who proposed monitoring pressure close to the high-demand locations. In fact, the error in mass balance is likely to be larger in high demand points (i.e. large $\mathbf{d}_x$).

It can be argued that the condition $n_k << x$ is preferable in order to deal with uncertainty.

Finally, starting from the x mass balance equations and considering the case of n_n pressure measurements (i.e. all nodal heads are known) the following expressions can be written:

$$\begin{aligned}
&\mathbf{A}_{x,p}\mathbf{B}_{p,p}\mathbf{Q}_{p,1} = -\mathbf{A}_{x,p}\left(\mathbf{D}_{p,p}\right)^{-1}\left(\mathbf{A}_{p,0+x}\mathbf{H}_{0+x,1} + \mathbf{A}_{p,n-x}\mathbf{H}_{n-x,1}\right)\\
&\text{if } x = n_n \text{ and } \Delta\mathbf{H}_{p,1} = -\overline{\mathbf{A}}_{p,n}\mathbf{H}_{0+n,1}\\
&\mathbf{d}_{n,1} = \mathbf{A}_{n,p}\left(\mathbf{A}_{p,p}\right)^{-1}\Delta\mathbf{H}_{p,1}
\end{aligned} \tag{15}$$

which is the same result obtained by Todini (1999) by substituting $\Delta\mathbf{H}_{p,1}$ by a diagonal matrix (whose elements are $\Delta\mathbf{H}_{p,1}$), the diagonal matrix $\mathbf{A}_{p,p}$ with a vector (whose elements are its diagonal) and adding the term dependent on measurement errors. Clearly the problem in Equation (15) is ill-posed for one steady-state observation (topological unobservability of some state variables) as shown by Todini (1999).

Some Q_y are known

The energy balance equations related to y pipes, where water flow is measured or known (e.g. from demands of branched portions of the WDN), are removed from the system of Equation (1) since $\mathbf{Q}_y$ is no longer unknown. The reason is the same as in the previous section and the choice is also supported by works of Ozawa (1986) and Carpentier & Cohen (1991, 1993) since it corresponds to the arc deletion used while studying topological observability.

Mass balance equations at terminal nodes of the y pipes must be rewritten considering a new set of known demands $\mathbf{d}_y{}^{*} = \mathbf{d}_y - \mathbf{Q}_y$. In Figure 3 the modified network to be simulated for calibration purposes is reported on the right; the modified demands $d_2{}^{*}$ and $d_3{}^{*}$ are reported in grey while pipe 2 is removed.

In this case, the initial number of parameter to be estimated is reduced by y because the resistances of the y pipes could be removed from the calibration problem. However, in order to use these observations for the whole calibration problem, it is preferable to maintain those y state variables and assemble y equations based on observed $\mathbf{Q}_y$ and those flows computed by $\Delta\mathbf{H}_{p,1}$ and $\mathbf{R}_{y,1}$, as reported in the remainder of the paper.

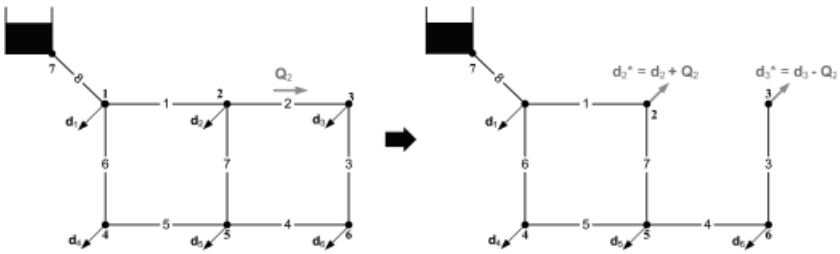

Figure 3 | Network simplification for known pipe discharge.

The discussion of the previous section about the network observability (i.e. equations/unknowns balance) still holds now considering the n_k state variables to be estimated and the previous y equations.

The equations representing the EGGA (or GGA) simulation can be re-written as follows:

$$\begin{aligned}
\mathbf{B}^{iter}_{p-y,p-y} &= \left(\mathbf{D}^{iter}_{p-y,p-y}\right)^{-1}\mathbf{A}^{iter}_{p-y,p-y}\\
\mathbf{F}^{iter}_{n,1} &= -\mathbf{A}_{n,p-y}\left(\mathbf{Q}^{iter}_{p-y,1} - \mathbf{B}^{iter}_{p-y,p-y}\mathbf{Q}^{iter}_{p-y,1}\right) +\\
&\quad + \mathbf{d}^{*}_{n,1} + \mathbf{A}_{n,p-y}\left(\mathbf{D}^{iter}_{p-y,p-y}\right)^{-1}\left(\mathbf{A}_{p,0}\mathbf{H}_{0,1}\right)\\
\mathbf{H}^{iter+1}_{n,1} &= -\left(\mathbf{A}_{n,p-y}\left(\mathbf{D}^{iter}_{p-y,p-y}\right)^{-1}\mathbf{A}_{p-y,n}\right)^{-1}\mathbf{F}^{iter}_{n,1}\\
\mathbf{Q}^{iter+1}_{p-y,1} &= \left(\mathbf{Q}^{iter}_{p-y,1} - \mathbf{B}^{iter}_{p-y,p-y}\mathbf{Q}^{iter}_{p-y,1}\right)\\
&\quad - \left(\mathbf{D}^{iter}_{p-y,p-y}\right)^{-1}\left(\mathbf{A}_{p,0}\mathbf{H}_{0,1} + \mathbf{A}_{p-y,n}\mathbf{H}^{iter+1}_{n,1}\right)
\end{aligned} \tag{16}$$

with $\dim(\mathbf{A}_{n,p\text{-}y}) = [n_n, n_p{-}y]$.

Decision variables of calibration problem $K_{k,i}{}^{\infty}$ should minimize, for example, the following objective function:

$$O.F. = \min_{K^{\infty}_{k,i}}\left(\sum\left|\mathbf{Q}^{comp}_{y,1} - \mathbf{Q}^{known}_{y,1}\right|\right) \tag{17}$$

where $\dim(\mathbf{A}_{y,y}) = [y,y]$, $\dim(\mathbf{A}_{y,n}) = [y, n_n]$ and $\dim(\mathbf{A}_{y,0}) = [y, n_0]$. $\mathbf{Q}_{y,1}{}^{comp}$ is the vector of the y pipe flows computed by using the nodal heads coming from the simulation of the network and some possible priors on pipe unitary hydraulic resistances (e.g. based on similarities with other pipes).

In the remainder of the text it is shown that the most important consequence of such strategy consists of using pipe flow measurements to potentially indentify separate components of the network, which allows a prompt analysis of pipe observability.

Some H_x and Q_y are known

The combination of the previous equations is an easy task. The EGGA (or GGA) system of equations is

$$\begin{aligned}
\mathbf{B}^{iter}_{p-y,p-y} &= \left(\mathbf{D}^{iter}_{p-y,p-y}\right)^{-1}\mathbf{A}^{iter}_{p-y,p-y}\\
\mathbf{F}^{iter}_{n-x,1} &= -\mathbf{A}_{n-x,p-y}\left(\mathbf{Q}^{iter}_{p-y,1} - \mathbf{B}^{iter}_{p-y,p-y}\mathbf{Q}^{iter}_{p-y,1}\right)\\
&\quad + \mathbf{d}^{*}_{n-x,1} + \mathbf{A}_{n-x,p-y}\left(\mathbf{D}^{iter}_{p-y,p-y}\right)^{-1}\left(\mathbf{A}_{p-y,0+x}\mathbf{H}_{0+x,1}\right)\\
\mathbf{H}^{iter+1}_{n-x,1} &= -\left(\mathbf{A}_{n-x,p-y}\left(\mathbf{D}^{iter}_{p-y,p-y}\right)^{-1}\mathbf{A}_{p-y,n-x}\right)^{-1}\mathbf{F}^{iter}_{n-x,1}\\
\mathbf{Q}^{iter+1}_{p-y,1} &= \left(\mathbf{Q}^{iter}_{p-y,1} - \mathbf{B}^{iter}_{p-y,p-y}\mathbf{Q}^{iter}_{p-y,1}\right)\\
&\quad - \left(\mathbf{D}^{iter}_{p-y,p-y}\right)^{-1}\left(\mathbf{A}_{p-y,0+x}\mathbf{H}_{0+x,1} + \mathbf{A}_{p-y,n-x}\mathbf{H}^{iter+1}_{n-x,1}\right)
\end{aligned} \tag{18}$$

where subscripts of matrices $\mathbf{A}_{n-x,\,p-y}$, $\mathbf{A}_{\,p-y,0+x}$ and $\mathbf{H}_{0+x,1}$ indicate their dimensions, as in previous sections. The objective function for calibration could be formulated by minimizing the sum of arguments of Equations (14) and (17). Alternatively, it is possible to explicitly account for uncertainties in observations and priors of decision variables (i.e. unitary hydraulic resistances) by using an objective function like that in Equation (19):

$$\begin{aligned}
O.F. = \min_{K^{\infty}_{k,i}}\Bigg\{&\frac{1}{\sigma^2_Q}\sum\|\mathbf{Q}^{comp}_{y,1} - \mathbf{Q}^{known}_{y,1}\|_2\\
&+\frac{1}{\sigma^2_d}\sum\|\mathbf{A}_{x,p}\mathbf{Q}^{comp}_{p,1} - \mathbf{d}_{x,1}\|_2 + \frac{1}{\sigma^2_K}\sum\|K^{\infty}_{k,i} - K^{\infty,\,prior}_{k,i}\|_2\Bigg\}
\end{aligned} \tag{19}$$

For the sake of simplicity, the uncertainty of each type of parameter is assumed to be drawn from normal distributions whose standard deviations are σ_Q, σ_d and σ_K. In particular, $1/\sigma_K{}^2$ can also be seen as a regularization parameter

that smooths the error surface of the inverse problem by cancelling local minima. This smoothing effect increases while reducing the standard deviation (i.e. increasing the confidence on priors $K_{k,i}^{\infty,prior}$). On the other hand, large σ_K^2 values consistently reproduce the adoption of weak priors within the objective function.

It is worth observing that the comprehensive strategy proposed here can also accommodate the analysis of uncertainty in observations and state parameters by resorting to a Bayesian framework (Kapelan *et al.* 2007) or grey numbers (Alvisi & Franchini 2010), provided some appropriate and consistent modifications are made.

REMARKS ON WDN DECOMPOSITION AND OBSERVABILITY

The system of Equations (18) shows that the network simulation model could be resized by removing:

- the mass balance equations of pressure measurement nodes whose heads ($\mathbf{H}_x$) are known. Those nodes are regarded as tank nodes;
- the energy balance equations of pipes whose flows are known ($\mathbf{Q}_y$). Nodal demand of their two terminal nodes is changed in order to encompass known flows.

In particular, the second transformation (due to known pipe flows) is of preeminent importance because it can help in analyzing topological observability. In fact, it can be argued that the new topology induced by the set of flow measurements/priors (as described by the matrix $[\mathbf{A}_{p-y,\ n-x}|\mathbf{A}_{p-y,0+x}\ ]$) can generate the partitioning of the original WDN into components.

Let's consider Figure 4, where the flow Q_9 is known from the demand d_B (sum of known demands of portion B) and the only water source is the tank at node 7. According to previous considerations, pipe 9 can be removed and the component analysis allows the identification of the network portions A and B. The automatic identification of network components can be performed by resorting to graph theory (Brualdi & Ryser 1991) or by using an adjacency matrix based on $[\mathbf{A}_{p\text{-}y,n}\ |\ \mathbf{A}_{p\text{-}y,0}]$ (Giustolisi & Savic 2010).

In this case (i.e. one pipe generating components) the nodal head H_8 derives from the behavior of network portion A. In fact, H_8 can be computed from H_3 based on a prior on

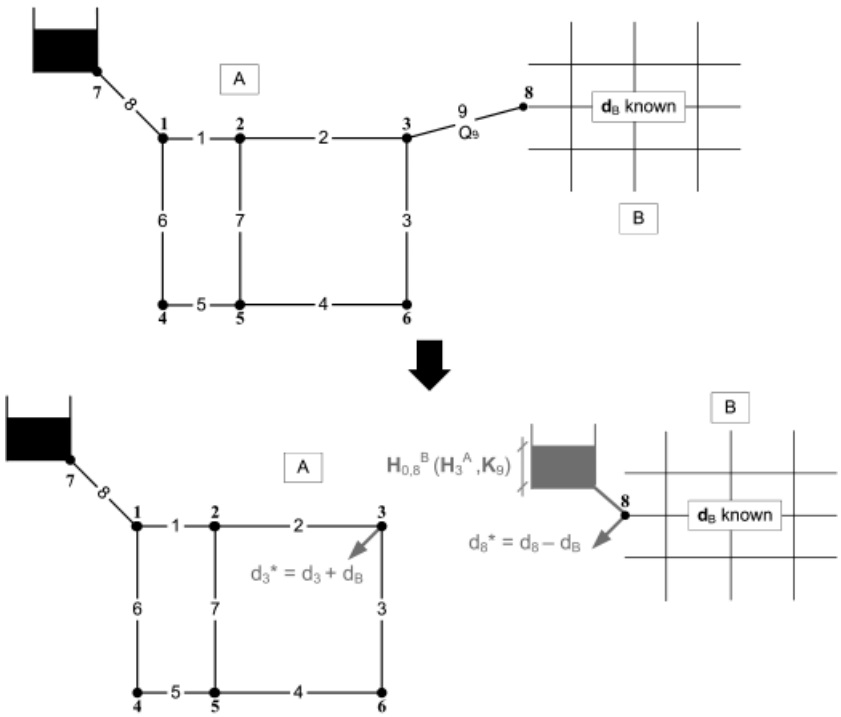

Figure 4 | Topological remarks: one-pipe connection.

R_9. In turn, the portion B can be simulated as a hydraulic system with a tank at node 8 and a mass balance constraint with respect to d_8^*.

Two outcomes can be emphasized from such a simple case:

- the mass balance in node 8 is always satisfied. In fact, d_8 is the nodal demand to which the mass balance must be constrained by definition, $Q_9 = d_B$ and $\Delta Q_j = d_B - d_8$, where j are the indices of pipes in portion B directly connected to node 8. Consequently, the mass balance at node 8 is $Q_9 - \Delta Q_j = d_B - d_B + d_8 = d_8$.
- the value of H_3 simply acts as an offset on nodal heads in hydraulic portion B, and the hydraulic behavior in portion B does not affect portion A.

Note that such a scheme holds because there is no other source of water in portion B; thus portion B is fed by portion A through pipe 9. Then, the value of Q_9 (i.e. the demands $\mathbf{d}_B$) is sufficient to simulate the behavior of portion B with respect to A, since the hydraulic resistances in B do not influence the nodal head H_3. Therefore, the hydraulic resistances of portion B and pipe 9 cannot be calibrated by using the measurements in A since they are not observable from A. Thus, the topological unobservability of portion B occurs.

onsequently, if a grouping of pipes is assumed, the calibration of portion A allows us to calibrate only those pipes in B (and pipe 9) belonging to the same group of pipes in A. Otherwise, the prior hydraulic resistance hypothesized for pipes in B not yet included in the groups of A cannot be improved using the measurements in A. Moreover, the unobservability of those pipes cannot be eliminated by neither adding measurement points in portion A nor adopting EPS analysis (i.e. several independent sets of steady-state observations) with the same sampling points in A.

Afterwards, the situation reported in Figure 4 asks for resizing the calibration problem with respect to a more appropriate portion of the WDN, i.e. by removing portion B from the network simulation.

Now it is possible to discuss two further cases: (i) portion B is connected to portion A with at least one more pipe whose flow is known (Figure 5); (ii) portion B has at least one pressure measurement point or a source of water (i.e. a tank):

(i) In the first case the network portion A needs to be calculated at first and the hydraulic status of portion B

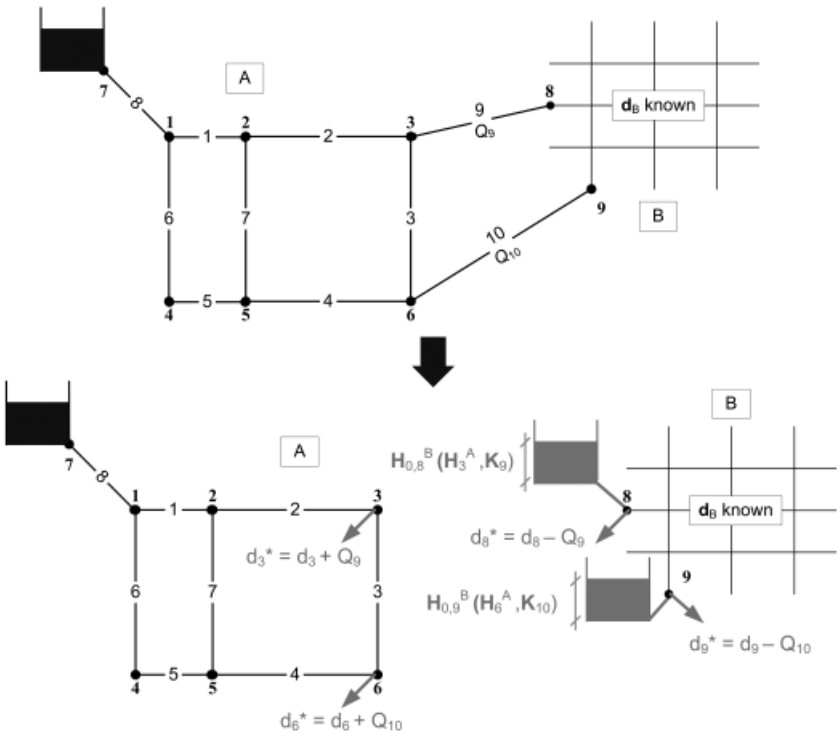

Figure 5 | Topological remarks: two-pipe connection.

depends on the nodal heads at terminal nodes of the connecting pipes. Figure 5 emphasizes this concept reporting heads $H_{0,8}$ and $H_{0,9}$ as functions of heads H_3^A and H_6^A and unitary hydraulic resistances K_9 and K_{10}, respectively.

(ii) In the second case there are two network components, A and B, connected by one pipe (i.e. pipe 9), but pipe 9 is strategic for the calibration. An error in the hydraulic resistance of pipe 9 would strongly bias the results of calibration of the whole network. In this circumstance a strong prior on unitary hydraulic resistance under rough fully turbulent flow of pipe 9 ($K_{9,\infty}$) is particularly useful. In this case the objective function to be minimized should account for the simulation of both portions of the network.

As a vestige from these topological remarks, some considerations can be drawn on sampling design. The analysis of the network is useful in discovering those portions of the network which are connected by means of one pipe only. When such portions (i.e. portion B of the previous example) do not include any source of water it is mandatory to have some pressure measurements in order to make pipes in the same portion topologically observable for calibration purposes.

For pipes connecting separate network components strong priors on unitary hydraulic resistance or their inclusion into an homogeneous group of pipes are strongly recommended to accomplish reliable and consistent calibration.

Moreover, due to the interpretation of nodal pressure measurements as tanks, a plausible criterion to deploy pressure sampling points inside network components could be to decide a minimum number of paths between two nodal observations (e.g. two pipes: no adjacent sampling points). Such an analysis could be easily accomplished by using the node adjacency matrix (Giustolisi & Savic 2010).

It is worth observing that network components could degenerate in a single tank or pressure measurement point. These cases will be treated in the following case study 1 as special ones, leading to further advice on system monitoring (in terms of both sampling design and priors on pipe hydraulic resistances).

It is evident that pipes forming closed loops represent portions of the network that cannot be calibrated, unless there is a measurement point along the loop, in a node different from that connecting the loop to the network.

As a concluding remark, the observability analysis allows us to promptly identify those network components that are actually useful for calibration purposes as they reflect the whole network functioning.

In such a context, the EGGA simplification of serial trunks allows an easier identification of those links (even comprising many trunks) that induce network decomposition, thus being the most crucial to be monitored (e.g. by collecting some flow measurements through them) and accurately calibrated since they greatly affect network behavior. Also this point is further emphasized in the following case study 1.

CASE STUDY 1 – HANOI NETWORK

The advantages of using the proposed calibration scheme are demonstrated on the well-known Hanoi network (e.g. Abebe & Solomatine 1998). The original network topology is reported in Figure 6 and comprises 1 tank, 34 pipes and 31 internal nodes.

As a preliminary analysis, the adoption of the EGGA and its energy balance correction strategy allows us to consider a simplified network topology comprised of 11 pipes and 8 nodes, as reported at the bottom of Figure 6. Clearly, the demands in the remaining nodes of the simplified network account for distribution along the new pipes (P_k) as detailed by Equations (4)–(6). From this point on, indices of nodes and pipes refer to the simplified topology (bottom of Figure 6).

It is evident that using EGGA greatly simplifies network topology, thus allowing the identification of links and nodes which are crucial to describe hydraulic network behavior.

The next step is to analyze the network in Figure 6 provided that the flows in pipes 1, 2 and 9 are known from the assumed known demands at all nodes. The component analysis (e.g. performed using the adjacency matrix) allows us to obtain the four network components A, B, C and D.

Some preliminary recommendations can be drawn about the sampling design for the Hanoi network. In fact, the components A, C and D are comprised of one node only and, as discussed above, components C and D need a pressure measurement at nodes 4 and 8 in order to make observable hydraulic resistances at pipes 2 and 9, respectively.

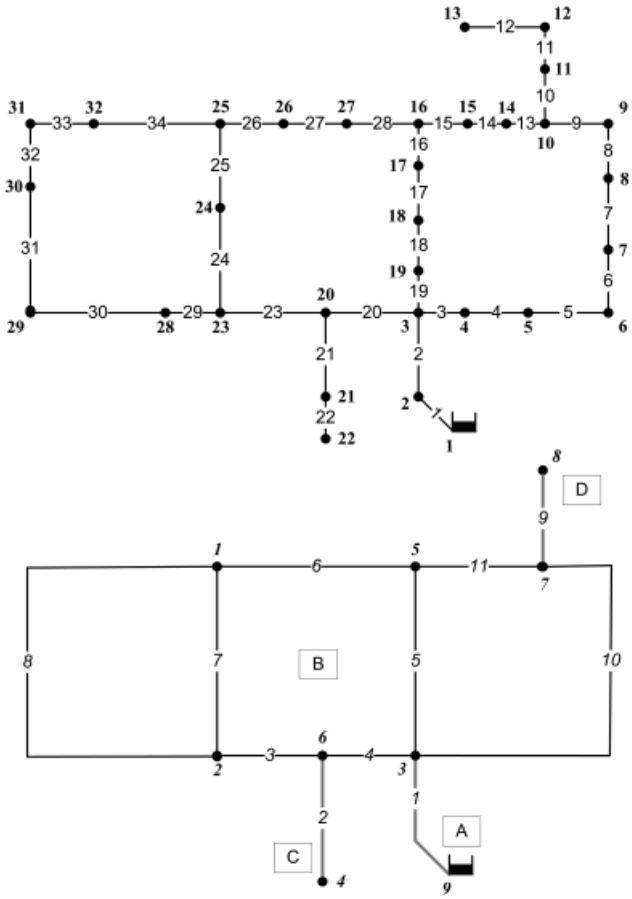

Figure 6 | Hanoi: network topology.

Furthermore, for the same reason, the closed loop component B requires at least one measurement (pipe flow or nodal pressure) since it does not include any known head (e.g. water tank). For example a pressure measurement could be set in node 1 being not adjacent to nodes 6, 7 and 3, as remarked in the previous section.

Moreover, to achieve unbiased estimates of pipe resistances in component B and, consequently, in C and D a strong prior on the resistance of pipe 1 or the grouping of homogeneous pipes is required. In fact, the overestimation/underestimation of the pipe 1 resistance would be compensated by the underestimation/overestimation of pipe resistances of component B while minimizing the mass balance error in node 1 derived from the pressure measurement.

Based on the above deployment of sampling points, the calibration of the Hanoi network in Figure 6 can be performed by minimizing the error on mass balance objective functions in nodes 1, 3, 4, 6, 7 and 8 and simulating component B only, which comprises 5 nodes and 8 pipes (while components A, C and D are comprised of 1 node only). The distance between actual and prior values of unitary resistances for some pipes (e.g. pipe 1) can be easily added to the objective functions to be minimized in a multi-objective optimization context or as a further term in a single-objective weighted sum (e.g. see Equation (19)).

As a concluding remark let's observe that the identification of network components also induces a sort of ranking of pipe resistances based on hydraulic network behavior. In fact, supposing that components C and D had a more articulated topology, their hydraulic functioning would be affected by the calibration of component B. Similarly, it can be argued that calibration of pipe 1 is essential since it directly affects component B and, in turn, C and D.

CASE STUDY 2 – APULIAN2

This section proposes a numerical case study where the proposed calibration methodology (based on the resized WDN modelling) and the classical calibration approach are compared. The network considered herein is named Apulian2 and is chosen since it conjugates the ease of analysis due to its small size with a realistic variation of hydraulic functioning due to the assumed daily demand pattern. Figure 7 depicts the topology of Apulian2 and Table 1 reports data on pipes (length, internal diameter and total distributed demand) and nodes (elevation HL and tank water levels P_0). As

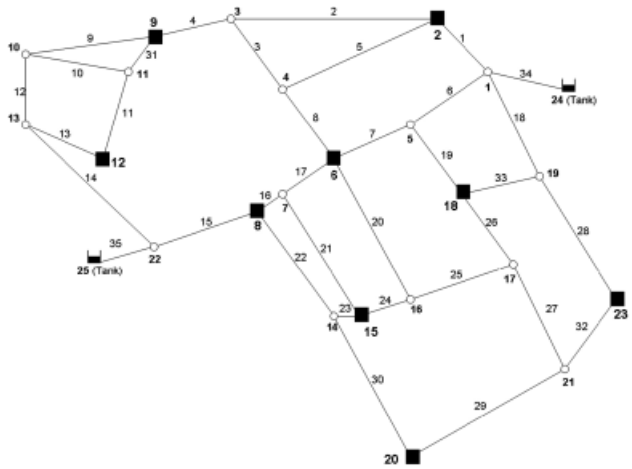

Figure 7 | Apulian2: network topology.

Table 1 | Apulian2: pipes and nodes data

Pipe ID	L_k (m)	D_k (mm)	P_k ($m^3\ s^{-1}$)	Node ID	HL (m)	P_0 (m)	Mean (ΔP_0) (m)
1	348.5	163.6	0.0057	**1**	6.4	–	–
2	955.7	100.0	0.0155	**2**	7.0	–	+1.82
3	483.0	100.0	0.0078	**3**	6.0	–	–
4	400.7	100.0	0.0065	**4**	8.4	–	–
5	791.9	100.0	0.0129	**5**	7.4	–	–
6	404.4	163.6	0.0066	**6**	9.0	–	+1.92
7	390.6	100.0	0.0063	**7**	9.1	–	–
8	482.3	100.0	0.0078	**8**	9.5	–	−1.94
9	934.4	100.0	0.0152	**9**	8.4	–	+1.86
10	431.3	184.0	0.0070	**10**	10.5	–	–
11	513.1	100.0	0.0083	**11**	9.6	–	–
12	428.4	204.4	0.0070	**12**	11.7	–	+2.13
13	419.0	100.0	0.0068	**13**	12.3	–	–
14	1023.1	257.6	0.0166	**14**	10.6	–	–
15	455.1	327.2	0.0074	**15**	10.1	–	−2.01
16	182.6	204.4	0.0030	**16**	9.5	–	–
17	221.3	184.0	0.0036	**17**	10.2	–	–
18	583.9	229.0	0.0095	**18**	9.6	–	+1.99
19	452.0	100.0	0.0073	**19**	9.1	–	–
20	794.7	100.0	0.0129	**20**	13.9	–	−2.13
21	717.7	100.0	0.0117	**21**	11.1	–	–
22	655.6	204.4	0.0107	**22**	11.4	–	–
23	165.5	100.0	0.0027	**23**	10.0	–	+1.83
24	252.1	100.0	0.0041	**24**	15.0	21.4	–
25	331.5	100.0	0.0054	**25**	15.0	21.4	–
26	500.0	100.0	0.0081				
27	579.9	100.0	0.0094				
28	842.8	184.0	0.0137				
29	792.6	100.0	0.0129				
30	846.3	163.6	0.0138				
31	164.0	163.6	0.0027				
32	427.9	163.6	0.0070				
33	379.2	100.0	0.0062				
34	158.2	327.2	–				
35	158.2	368.2	–				

happens in many real networks, most of the pipes have small diameters (i.e. 100 mm). The assumed demand pattern is reported in Figure 8 in terms of the ratio between actual (P_k (t)) and maximum demand (i.e. P_k reported in Table 1) distributed along the pipes for every hour of the day (Giustolisi *et al.* 2008).

Since no real measurements are available for such a network, a fictitious set of measurements has been obtained

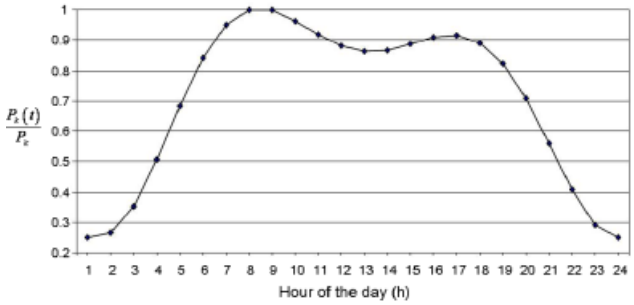

Figure 8 | Apulian2: daily demand pattern.

by performing the 24-h EPS of the network using the EGGA under the hypothesis that all pipe hydraulic resistances ($K_{k,\infty}$) were known. For example, each pipe diameter has been assigned a unitary hydraulic resistance as reported in Table 2 (i.e. 8 groups of $K_{k,\infty}$). The nodal heads ($\mathbf{H}_x$) obtained from such EPS are hypothesized to represent as many pressure observations for the remainder of this section.

To perform the numerical analysis, only a subset of such pressure "measurements" are used, namely pressure values at all hours of the day at nodes 2, 6, 8, 9, 12, 15, 18, 20 and 23 (black squares in Figure 7), so that the minimum path length between them is two (i.e. there are no adjacent pressure measurement nodes).

Moreover, these values in $\mathbf{H}_x$ (in 2, 6, 8, 9, 12, 15, 18, 20 and 23) have been modified by randomly adding or subtracting a perturbation associated with error measurements. Such perturbations ΔP_0 are sampled from a normal distribution N(2.1, 0.1) whose mean and variance reflect the accuracy and precision of commercial pressure data loggers. The sign of the perturbation does not change in time for a given node.

Table 2 | Apulian2: hydraulic resistances

Group	D (mm)	$K_{k,\infty}$ (s^2 m^{-6})
1	100.0	265.1467
2	163.6	18.5649
3	184.0	9.8824
4	204.4	5.6291
5	229.0	3.0681
6	257.6	1.6389
7	327.2	0.4605
8	368.2	0.2466

The last column of Table 1 reports the mean values of perturbations considered. For the sake of simplicity, no pipe flow measurements are considered here.

The search for *calibrated* unitary hydraulic resistances $K_{k,\infty}{}^{\text{cal}}$ is performed by assuming values in Table 2 as priors $K_{k,\infty}$, while actual decision variables are coefficients $\phi_k = K_{k,\infty}{}^{\text{cal}}/K_{k,\infty}$. Such setting of the problem allows minimizing also the difference between calibrated and prior (true) unitary hydraulic resistances in terms of $|\phi_k - 1|$ for each pipe group.

According to the classical methodology (referred to as "Problem 1" in the remainder), the calibration has been performed by simultaneously minimizing the following two objective functions:

$$obj\ 1 = \min_{\phi_k}\left\{\frac{\sum_k|\phi_k - 1|}{n_k}\right\};$$
$$obj\ 2 = \min_{\phi_k}\left\{\frac{\sum_t\left|\mathbf{H}_x^{sim}(t) - \mathbf{H}_x(t)\right|}{216}\right\} \quad (20)$$

where *obj2* minimizes the distance between measured pressures and simulated pressures for a given set of unitary hydraulic resistances (i.e. ϕ_k); 216 is the number of available observations (i.e. 24 h for 9 sampling points). In contrast, the new calibration (referred to as "Problem 2") consists of minimizing the objective functions

$$obj\ 1 = \min_{\phi_k}\left\{\frac{\sum_k|\phi_k - 1|}{n_k}\right\};$$
$$obj\ 3 = \min_{\phi_k}\left\{\frac{\sum_t\left|\mathbf{A}_{x,p}\mathbf{Q}_{p,1}(t) - \mathbf{d}_{x,1}(t)\right|}{216}\right\}. \quad (21)$$

Note that the objective function *obj*1 does not change and represents a measure of the mean distance between actual unitary hydraulic resistances and the priors $K_{k,\infty}$. The decision variables of calibration are the 8 values of $K_{k,\infty}{}^{\text{cal}}$ (i.e. ϕ_k). Therefore, a single value of *obj*1 actually might reflect several combinations in 8D space of $K_{k,\infty}{}^{\text{cal}}$ resulting into as many values of *obj*2 and *obj*3. Thus, the set of $K_{k,\infty}{}^{\text{cal}}$ that minimizes *obj*2 might result in a non-minimum value of *obj*3, and vice versa.

Both calibration procedures started from the same initial point representing the priors (i.e. $K_{k,\infty}{}^{\text{cal}} = K_{k,\infty}$ and $\phi_k = 1$, $k = 1, \ldots, 8$); incidentally, the value of *obj*2 resulting

from such an initial solution coincides with the mean value of the perturbations at measurement points over the 24 h EPS (i.e. the last column of Table 1).

Solving Problem 1 (classical calibration) requires the simulation of a WDN comprising of 2 tanks, 23 internal nodes and 35 pipes (i.e. 58 simulation unknowns), while Problem 2 (new calibration) is based on the simulation of a resized WDN with 11 tanks (2 + 9 measured heads), 14 internal nodes and 35 pipes (i.e. 49 simulation unknowns).

Figure 9 depicts the solutions of the two separate problems. Due to the remarks on *obj*1, the comparison between the two Pareto fronts is not straightfprward as the same *obj*1 could reflect several possible sets of $K_{k,\infty}$ in 8D space. Thus, any conclusion can be drawn about the mutual Pareto dominance of the solutions, apart from that it seems that the Pareto front of the new calibration formulation (Problem 2) is simply more spread because of the lowest dimensionality of the multi-objective optimization problem.

In Problem 2 (new calibration), *obj*3 represents the distance between the outflows obtained from mass balance at nodes x and nodal demands $\mathbf{d}_x$. Referring *obj*3 to the nodes where $\mathbf{d}_x$ is the highest would mean minimizing the largest expected mismatching of mass balance and definitely

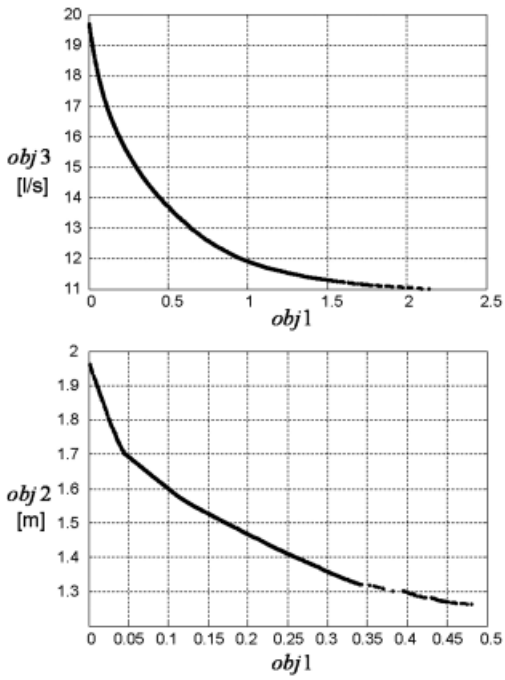

Figure 9 | Calibration results: Problem 1 (down) and Problem 2 (top).

achieves the more plausible calibration of $K_{k,\infty}$, as recommended by Walski (1983).

Actually, $\mathbf{d}_{x,1}$ are rarely measured and their values usually reflect prior assumptions on demand patterns along the pipes. *Obj*3 accounts for uncertainty in such assumption both implicitly, since $\mathbf{Q}_{p,1}$ are obtained from WDN simulation based on the assumed $\mathbf{d}_{x,1}$, and explicitly in terms of the distance between simulated outflows and priors $\mathbf{d}_{x,1}$. Then, when demand $\mathbf{d}_{x,1}$ is significantly uncertain, solving Problem 2 returns a set of candidate unitary pipe hydraulic resistances which reflect such uncertainty. Moreover, $\mathbf{d}_{x,1}$ being derived from the assumed demand distributed along the pipes (i.e. through parameters ϕ_k), its uncertainty can be thought of as reflecting the uncertainty on adjacent nodes (not yet monitored).

In contrast, uncertainty on $\mathbf{d}_{x,1}$ is only implicitly included in Problem 1 since simulated heads $\mathbf{H}_x^{sim}$ result from the assumed demand patterns along the pipes.

$\mathbf{H}_x$ observations are usually assumed as true values and their uncertainty could potentially derive from measurement errors only. In Problem 1 (classical calibration), *obj*2 represents the distance between simulated heads $\mathbf{H}_x^{sim}$ and measurements $\mathbf{H}_x$, while Problem 2 (new calibration) implicitly accounts for uncertainty in $\mathbf{H}_x$, since pipe flows $\mathbf{Q}_{p,1}$ result also from values of $\mathbf{H}_x$ (water levels in fictitious tanks) hypothesized for simulation.

In summary, the uncertainties surrounding $\mathbf{d}_{x,1}$ and $\mathbf{H}_x$ are of different types: usually values on $\mathbf{d}_{x,1}$ are assumed as priors and uncertainty surrounding actual demands could be significant; in contrast measurements of $\mathbf{H}_x$ are usually assumed as true values, apart from measurement errors. Problem 2 (new calibration) accounts both implicitly and explicitly for the distance from priors $\mathbf{d}_{x,1}$ (i.e. *obj*3) and only implicitly for error measurements in $\mathbf{H}_x$ (i.e. through simulated $\mathbf{Q}_{p,1}$). Vice versa, Problem 1 (classical calibration) accounts explicitly for the distance from $\mathbf{H}_x$ (i.e. *obj*2) and only implicitly for uncertainty in demands (i.e. through simulated $\mathbf{H}_x^{sim}$).

Another remarkable difference between objectives *obj*2 and *obj*3 stems from the number of hydraulic variables involved in their calculation. Such a number is equal to the number of measured heads (i.e. 9 in this case study) in *obj*2; while *obj*3 explicitly considers all pipe flows converging to pressure measurement nodes (i.e. 22 in this case study),

which are usually many more than the nodes themselves. In addition, *obj*3 explicitly consists of 9 linear equations of 22 hydraulic variables; while *obj*2 implicitly derives from 22 energy balance equations (i.e. solved during WDN simulation) which are proportional to $K_{k,\infty} L_k Q_k^2$.

Accordingly, in order to make pressure measurements maximally informative for calibration, the monitored nodes should not be adjacent and join the largest possible number of pipes. This fact further confirms the opportunity to measure pressure at high demand nodes, because pipes joining multiple pipes are likely to correspond to large nodal demands (see Equation (2)). In more general terms, calibrating a WDN according to Problem 2 provides a practical criterion for pressure sampling design which avoids redundant information being collected.

CONCLUSIONS

The calibration of pipe hydraulic resistances used in WDN simulation models is of paramount importance to achieve reliable results for network management. In recent years many procedures for calibration and sampling design have resorted to trial-and-error approaches requiring multiple WDN simulations. Unfortunately, two key issues of model calibration have been seldom considered: the correctness of demand representation within the simulation model and the analysis of actual topological observability of pipes achievable through simulation under a given set of pressure/flow measurements.

This paper leverages some recent developments in WDN hydraulic modeling to demonstrate the need for a more correct representation of demands for calibration purposes. In particular, a comprehensive formulation of the EGGA is proposed here. It includes the effect of pumps and valves (minor losses) beyond the actual flow regime and demand pattern through the pipes already included in the original formulations. The advantages of using the EGGA for calibration purposes can be summarized as follows: (i) considerable reduction of the simulation problem size due to the elimination of serial nodes in a simplified network topology; (ii) identification of those pipes/links which are most important to describe global network hydraulic behaviour and (iii) averaging effect of uncertainties surrounding boundary conditions in WDN simulations.

The analysis of topological observability is based on the observation that measurements of nodal pressure and/or pipe flow induce some modifications in the balance between unknowns and energy/mass conservation equations of the simulation problem. Known heads and pipe flows have been basically represented as equivalent fictitious tanks and interrupted pipes, respectively, provided that as many equations are used as the objectives of model calibration. The advantages of the proposed calibration procedure are as follows: (i) identification of network components induced by potentially available pipe flow measurements; (ii) prompt detection of pipes actually observable using the available observations and (iii) possible ranking of pipe roughness to be calibrated based on hydraulic WDN behaviour. In addition, such a strategy may result in a reduced computational burden for WDN simulation, especially for large size real networks, and provides a pragmatic support for sampling design.

REFERENCES

Abebe, A. J. & Solomatine, D. P. 1998 Application of global optimization to the design of pipe networks. In: *Proc. Hydroinformatics '98*. Balkema, Rotterdam, pp. 989–996.

Alvisi, S. & Franchini, M. 2010 Pipe roughness calibration in water distribution systems using grey numbers. *J. Hydroinformatics* **12**(4), 424–445.

Bargiela, A. & Hainsworth, G. D. 1989 Pressure and flow uncertainty in water systems. *J. Wat. Res. Plann. Mngmnt.* **115**(2), 212–229.

Berardi, L., Giustolisi, O. & Todini, E. 2010 Accounting for uniformly distributed pipe demand in WDN analysis: Enhanced GGA. *Urban Wat. J.* **7**(4), 243–255.

Bhave, P. R. 1988 Calibrating water distribution network models. *J. Environ. Engng.* **114**(1), 120–136.

Boulos, P. F. & Wood, D. J. 1990 Explicit calculation of pipe network parameters. *J. Hydraul. Engng.* **116**(11), 1329–1344.

Brualdi, R. A. & Ryser, H. J. 1991 *Combinatorial Matrix Theory*. Cambridge University Press, Cambridge.

Bush, C. A. & Uber, J. G. 1998 Sampling design methods for water distribution model calibration. *J. Wat. Res. Plann. Mngmnt.* **124**(6), 334–344.

Carpentier, P. & Cohen, G. 1991 State estimation and leak detection in water distribution networks. *J. Civil Engng. Environ. Syst.* **8**(4), 247–257.

Carpentier, P. & Cohen, G. 1993 Applied mathematics in water supply network management. *Automatica* **29**(5), 1215–1250.

Datta, R. S. N. & Sridharan, K. 1994 Parameter estimation in water-distribution systems by least squares. *J. Wat. Res. Plann. Mngmnt.* **120**(4), 405–422.

Ferrante, M., Todini, E., Massari, C., Brunone, B. & Meniconi, S. 2011 Equivalent hydraulic resistance to simulate pipes subject to diffuse outflows. *J. Hydroinformatics*, in press.

Gargano, R. & Pianese, D. 2000 Reliability as tool for hydraulic network planning. *J. Hydraul. Engng.* **126**(5), 354–364.

Giustolisi, O. 2010 Considering actual pipe connections in WDN analysis. *J. Hydraul. Engng.* **136**(11), 889–900.

Giustolisi, O., Kapelan, Z. & Savic, A. 2008 Extended period simulation analysis considering valve shutdowns. *J. Wat. Res. Plann. Mngmnt.* **134**(6) 527–537.

Giustolisi, O. & Savic, D. A. 2010 Identification of segments and optimal isolation valve system design in water distribution networks. *Urban Wat. J.* **7**(1), 1–15.

Giustolisi, O. & Todini, E. 2009 Pipe hydraulic resistance correction in WDN analysis. *Urban Wat. J.* **6**(1), 39–52.

Greco, M. & Del Giudice, G. 1999 New approach to water distribution network calibration. *J. Hydraul. Engng.* **125**(8), 849–854.

Kapelan, Z., Savic, D. A. & Walters, G. 2003 Multiobjective sampling design for water distribution model calibration. *J. Wat. Res. Plann. Mngmnt.* **129**(6), 466–479.

Kapelan, Z., Savic, D. A. & Walters, G. A. 2007 Calibration of water distribution hydraulic models using a Bayesian-type procedure. *J. Hydraul. Engng.* **133**(8), 927–936.

Lansey, K. E., El-Shorbagy, W., Ahmed, I., Araujo, J. & Haan, C. T. 2001 Calibration assessment and data collection for water distribution networks. *J. Hydraul. Engng.* **127**(4), 270–279.

Lingireddy, S. & Ormsbee, L. E. 1999 *Optimal Network Calibration Model Based on Genetic Algorithms*. University of Kentucky. In: *Proc. 29th Annual Water Resources Planning and Management Conference WRPMD (1999) – Preparing for the 21st Century*, Tempe, Arizona, USA, ASCE.

Mallick, K. N., Ahmed, I., Tickle, K. S. & Lansey, K. E. 2002 Determining pipe grouping for water distribution networks. *J. Wat. Res. Plann. Mngmnt.* **128**(2), 130–139.

Ormsbee, L. E. 1989 Implicit network calibration. *J. Wat. Res. Plann. Mngmnt.* **115**(2), 243–257.

Ormsbee, L. E., & Wood, D. J. 1986 Explicit pipe network calibration. *J. Wat. Res. Plann. Mngmnt.* **112**(2), 166–182.

Ozawa, T. 1986 The principal partition of a pair of graphs and its applications. *Discr. Appl. Math.* **17**, 163–186.

Pudar, R. S., & Liggett, J. A. 1992 Leaks in pipe networks. *J. Hydraul. Engng.* **118**(7), 1031–1046.

Rahal, C. M., Sterling, M. J. H. & Coulbeck, B. 1980 Parameter tuning for simulation models of water distribution networks. *Proc. ICE* **69**(2), 751–762.

Reddy, P. V. N., Sridharan, K. & Rao, P. V. 1996 WLS method for parameter estimation in water distribution networks. *J. Wat. Res. Plann. Mngmnt.* **122**(3), 157–164.

Rossman, L. A. 2000 *Epanet2 Users Manual*. US EPA.

Savic, D. A., & Walters, G. A. 1995 *Genetic Algorithm Techniques for Calibrating Network Models*. Rep. No. 95/12, Centre for Systems and Control Engineering, University of Exeter. Available at: http://centres.exeter.ac.uk/cws/downloads/cat_view/27-reports

Shamir, U. & Howard, C. D. D. 1968 Water distribution systems analysis. *J. Hydraul. Div, ASCE* **94**(1), 219–234.

Swamee, P. K. & Jain, A. K. 1976 Explicit equations for pipe flow problems. *J. Hydraul. Engng.* **102**(5), 657–664.

Todini, E., 1999 Using a Kalman filter approach for looped water distribution network calibration. In: *Proc. Water Industry Systems: Modelling and Optimisation Applications, University of Exeter* (Savic, D. A., Walters, G. A.)). vol. 1, pp. 327–336. Taylor & Francis, London.

Todini, E. & Pilati, S. 1988 A gradient method for the solution of looped pipe networks. In: *Computer Applications in Water Supply*. John Wiley & Sons, New York, vol 1, pp. 1–20.

Vasiljev, A., Martaud, M. & Koppel, T. 2009 Use of relative roughness and changed diameters for old pipes when calibrating an operational water distribution system. In: *Proc. Computer and Control in Water Industry (CCWI) – Integrating Water Systems, University of Sheffield* (Boxall, J., Maksimovic, C.). Taylor & Francis, London, pp 283–286.

Walski, T. M. 1983 Technique for calibrating network models. *J. Wat. Res. Plann. Mngmnt.* **109**(4), 360–372.

Xu, C. & Goulter, I. 1996 Uncertainty analysis of water distribution networks. In: *Proc. Stochastic Hydraulics '96* (Tickle, K. S., Goulter, I. C., Xu, C., Wasimi, S. A., Bouchart, F.). Balkema, Rotterdam, pp 609–616.

First received 9 October 2009; accepted in revised form 22 January 2010. Available online 4 October 2010

APPENDIX A

A comprehensive dimensionless formulation of the EGGA is provided herein. The error in pipe energy balance can be obtained from Equation (3) (Giustolisi & Todini 2009; Giustolisi 2010):

$$E_k^{\epsilon} = K_{k,\infty}\epsilon_k Q_k |Q_k|^{n-1} L_k = K_{k,\infty} z_k P_k^n L_k. \quad \text{(A1)}$$

Consequently, for a given set of $K_{k,\infty}$, ϵ_k ensures the two-way relationship between the actual hydraulic status of the network and its representation within the simulation model, provided that the solution to the system (1) is unique (Todini & Pilati 1988).

The modification of the matrices of the simulation model in Equation (2) is obtained by introducing four dimensionless parameters (Giustolisi 2010):

$$z_k = \epsilon_k \delta_k |\delta_k|^{n-1}; \qquad zA_k = \frac{z_k}{\delta_k}; \qquad zD_k = \frac{dz_k}{d\delta_k};$$

$$zB_k = \frac{|\delta_k|^{n-1} + zA_k}{n|\delta_k|^{n-1} + zD_k} \tag{A2}$$

with $\delta_k = Q_k/P_k$ and P_k total supplied demand through the kth pipe. The elements of key diagonal matrices in Equation (2) can be rewritten as

$$\begin{aligned} A_{p,p}(k,k) &= R_{k,\infty}|Q_k|^{n-1} + \{R_{k,\infty} zA_k P_k^{n-1}\} \\ D_{p,p}(k,k) &= nR_{k,\infty}|Q_k|^{n-1} + \{R_{k,\infty} zD_k P_k^{n-1}\} \\ B_{p,p}(k,k) &= \frac{A_{p,p}(k,k)}{D_{p,p}(k,k)} = \frac{|Q_k|^{n-1} + \{zA_k P_k^{n-1}\}}{n|Q_k|^{n-1} + \{zD_k P_k^{n-1}\}} = zB_k. \end{aligned} \tag{A3}$$

The formulation of the correction factor ϵ_k (Giustolisi & Todini 2009; Giustolisi 2010) for the general case of m_k connections along the kth pipe delivering constant demands $d_{k,j}$ (i.e. in demand-driven simulation) is based on the following assumptions:

$$\begin{aligned} &\kappa_{k,i} = \frac{K_{k,i}}{K_{k,\infty}} && \forall k \in [1,...,n_p], \forall i \in [1,...,m_k+1] \\ &\lambda_{k,i} = \frac{L_{k,i}}{L_k} && \forall k \in [1,...,n_p], \forall i \in [1,...,m_k+1] \\ &\delta_k + \alpha_k = \frac{Q_{k,1}}{P_k}; \qquad \delta_k = \frac{Q_k}{P_k}; \\ &P_k = \sum_{j=1}^{m_k} d_{k,j} && \forall k \in [1,...,n_p] \\ &\pi_{k,i} = \frac{\sum_{j=1}^{i-1} d_{k,j}}{P_k}; \qquad \pi_{k,1} \triangleq 0 && \forall k \in [1,...,n_p], \forall i \in [1,...,m_k] \end{aligned} \tag{A4}$$

where $K_{k,i}$, $f_{k,i}$, $L_{k,i}$ and $D_{k,i}$ are the unitary hydraulic resistance, the friction factor, the length and the internal diameter, respectively, of the ith trunk between two connections of the kth pipe. $K_{k,i}$ and $f_{k,i}$ depend on the Reynolds number (Re) and equivalent roughness (Ke $= e/D$, with e the absolute roughness), while $K_{k,\infty}$ and $f_{k,\infty}$ are the unitary hydraulic resistance and the friction factor respectively, under a rough fully turbulent flow regime (not dependent on Re). $K_{k,\infty}$ and $f_{k,\infty}$ are both computed with respect to diameter $D_{k,1}$. In the third of Equations (A4) $Q_{k,1}$ is pipe discharge in the first trunk of the kth pipe, while Q_k is the flow through the kth pipe as computed in the EGGA.

The corrections factors proposed by Giustolisi (2010) to account for the actual pipe flow regime are

$$\begin{aligned} \epsilon_k &= \frac{\sum_{i=1}^{m_k+1} \kappa_{k,i}(\delta_k + \alpha_k - \pi_{k,i})|\delta_k + \alpha_k - \pi_{k,i}|^{n-1} \lambda_{k,i}}{\delta_k |\delta_k|^{n-1}} - 1 \\ z_k &= \sum_{i=0}^{m_k+1} \kappa_{k,i}(\delta_k + \alpha_k - \pi_{k,i})|\delta_k + \alpha_k - \pi_{k,i}|^{n-1} \lambda_{k,i} - \delta_k |\delta_k|^{n-1} \\ zA_k &= \frac{\sum_{i=0}^{m_k+1} \kappa_{k,i}(\delta_k + \alpha_k - \pi_{k,i})|\delta_k + \alpha_k - \pi_{k,i}|^{n-1} \lambda_{k,i}}{\delta_k} - |\delta_k|^{n-1} \\ zD_k &= \frac{dz_k}{d\delta_k} = \sum_{i=0}^{m_k+1} \left[\left(P_k \frac{d\kappa_{k,i}}{dQ_{k,i}} (\delta_k + \alpha_k - \pi_{k,i}) + n\kappa_{k,i} \right) \right. \\ &\qquad \left. |\delta_k + \alpha_k - \pi_{k,i}|^{n-1} \lambda_{k,i} \right] - n|\delta_k|^{n-1} \end{aligned} \tag{A5}$$

where the derivative can be computed using, for example, the Swamee–Jain (1976) approximation of the Colebrook–White formula of the friction factor.

In the case of a pump installed before the connection $i = p$ and a minor loss before the connection $i = v$ it is possible to further generalize the corrections:

$$\begin{aligned} \epsilon_k &= \frac{-sign(r_{k,p}) \left(H^I_{k,p} - r^I_{k,p} |\delta_k + \alpha_k - \pi_{k,p}|^{\gamma_p} \right)}{\delta_k |\delta_k|^{n-1}} + \\ &\qquad \frac{\frac{A^2_{k,1}}{A^2_{k,v}} K^{ml}_{k,v} G_k (\delta_k + \alpha_k - \pi_{k,v}) |\delta_k + \alpha_k - \pi_{k,v}|}{\delta_k |\delta_k|^{n-1}} - 1; \\ on &= [r_{k,p}(\delta_k + \alpha_k - \pi_{k,p}) \geq 0]; \quad H^I_{k,p} = on \frac{\omega^2_{k,p} H_{k,p}}{R_{k,\infty} P_k^n}; \\ r^I_{k,p} &= on \frac{\omega_{k,p}^{2-\gamma_p} |r_{k,p}|}{R_{k,\infty} P_k^{n-\gamma_p}}; \\ G_k &= \frac{D_{k,1}}{f_{k,\infty} L_k P_k^{n-2}} \end{aligned} \tag{A6}$$

where $D_{k,p}$ and $A_{k,p}$ are the internal diameter and area of the pth trunk, $\omega_{k,p}$, $H_{k,p}$, $r_{k,p}$ and γ_p are the pump speed factor and the three parameters of pump curve; $K_{k,v}{}^{ml}$ is the minor loss coefficient and on is a Boolean variable that accounts for the pump installation direction. For further details see the EPANET tutorial (Rossman 2000). The sign of $r_{k,p}$ relates to pump installation direction; a positive $r_{k,p}$ means that the

installation of the pump is coherent with the positive direction assigned to flow in the WDN simulation model. Clearly, if the sign $r_{k,p}$ differs on actual flow, $on<0$, the pumps does not work and it should be substituted by a closed valve assuming that pumps are always equipped with a non-return valve. Hence, the remaining correction factors are

$$\begin{aligned}
z_k &= -sign(r_{k,p})\left(H^I_{k,p} - r^I_{k,p}\left|\delta_k+\alpha_k-\pi_{k,p}\right|^{\gamma_p}\right) \\
&\quad + \frac{A^2_{k,1}}{A^2_{k,v}} K^{ml}_{k,v} G_k\left(\delta_k+\alpha_k-\pi_{k,v}\right)\left|\delta_k+\alpha_k-\pi_{k,v}\right| - \delta_k|\delta_k|^{n-1} \\
zA_k &= \frac{-sign(r_{k,p})\left(H^I_{k,p} - r^I_{k,p}\left|\delta_k+\alpha_k-\pi_{k,p}\right|^{\gamma_p}\right)}{\delta_k} \\
&\quad + \frac{\frac{A^2_{k,1}}{A^2_{k,v}} K^{ml}_{k,v} G_k\left(\delta_k+\alpha_k-\pi_{k,v}\right)\left|\delta_k+\alpha_k-\pi_{k,v}\right|}{\delta_k} - |\delta_k|^{n-1} \\
zD_k &= +sign(r_{k,p})\gamma_p r^I_{k,p}\left|\delta_k+\alpha_k-\pi_{k,p}\right|^{\gamma_p-1} \\
&\quad + 2\frac{A^2_{k,1}}{A^2_{k,v}} K^{ml}_{k,v} G_k\left|\delta_k+\alpha_k-\pi_{k,v}\right| - n|\delta_k|^{n-1}
\end{aligned} \tag{A7}$$

It is possible to consider also the static pressure regain due to water withdrawal at each connection (Ferrante *et al.* 2011) by imposing the Bernoulli equation across sections i and $i+1$ of the kth pipe:

$$\Delta H_i = \frac{V^2_{k,i}}{2g} - \frac{V^2_{k,i+1}}{2g} - \beta\frac{(V_{k,i}-V_{k,i+1})\left|V_{k,i}-V_{k,i+1}\right|}{2g} \tag{A8}$$

where the parameter β accounts for minor head losses due to the change of water velocity as a consequence of the withdrawal in the connection point. $V_{k,i}$ and $V_{k,i+1}$ are the upstream and downstream velocities and ΔH_i is the head/pressure increase. Then, assuming

$$\begin{aligned}
&v_{k,i} = \frac{A_{k,i}}{A_{k,i+1}}; \quad \rho_{k,i} = \frac{Q_{k,i+1}}{Q_{k,i}} = 1 - \frac{q_{k,i}/P_k}{\delta_k+\alpha_k-\pi_{k,i}} \\
&C^{\beta}_{k,i} = (\rho_{k,i}v_{k,i})^2 + \beta(1-\rho_{k,i}v_{k,i})\left|1-\rho_{k,i}v_{k,i}\right| \\
&\forall k \in [1,\ ...,\ n_p], \forall i \in [1,\ ...,\ m_k]
\end{aligned} \tag{A9}$$

the correction factors are

$$\epsilon_k = G_k\frac{A^2_{k,1}\sum_{i=1}^{m_k}(C^{\beta}_{k,i}-1)\frac{(\delta_k+\alpha_k-\pi_{k,i})^2}{A^2_{k,i}}}{\delta_k|\delta_k|^{n-1}} - 1$$

$$\begin{aligned}
z_k &= G_k A^2_{k,1}\sum_{i=1}^{m_k}(C^{\beta}_{k,i}-1)\frac{(\delta_k+\alpha_k-\pi_{k,i})^2}{A^2_{k,i}} - \delta_k|\delta_k|^{n-1} \\
zA_k &= \frac{G_k}{\delta_k}A^2_{k,1}\sum_{i=1}^{m_k}(C^{\beta}_{k,i}-1)\frac{(\delta_k+\alpha_k-\pi_{k,i})^2}{A^2_{k,i}} - |\delta_k|^{n-1} \\
zD_k &= \frac{dz_k}{d\delta_k} + n|\delta_k| = 2G_k A^2_{k,1}\sum_{i=1}^{m_k}\frac{(\delta_k+\alpha_k-\pi_{k,i})}{A^2_{k,i}} - n|\delta_k|
\end{aligned} \tag{A10}$$

where $A_{k,i}$ is the internal area of the pipe. Finally, $\beta=0$ is assumed when $\rho_{k,i}<0$, because the inversion of flow occurs in the ith connection (i.e. $q_{k,i} > Q_{k,i}$), and $\beta=0.5$ otherwise.

The comprehensive correction factors can be obtained from Equation (9) and Equations (A1)–(A5). For example the overall correction ϵ_k is

$$\begin{aligned}
\epsilon_k = &\overbrace{\frac{\sum_{i=1}^{m_k+1}\kappa_{i+1}(\delta_k+\alpha_k-\pi_{k,i})\left|\delta_k+\alpha_k-\pi_{k,i}\right|^{n-1}\lambda_{i+1}}{\delta_k|\delta_k|^{n-1}}}^{\text{due to actual connections and flow regimes}} \\
&+\overbrace{\frac{\sum_{i=0}^{m_k+1} - sign(r_k)\left(H^I_{k,i} - r^I_k\left|\delta_k+\alpha_k-\pi_{k,i}\right|^{\gamma}\right)}{\delta_k|\delta_k|^{n-1}}}^{\text{due to pump}} \\
&+\overbrace{\frac{\sum_{i=0}^{m_k+1}\frac{A^2_{k,1}}{A^2_{k,i}}K^{ml}_{k,i}G_k(\delta_k+\alpha_k-\pi_{k,i})\left|\delta_k+\alpha_k-\pi_{k,i}\right|}{\delta_k|\delta_k|^{n-1}}}^{\text{due to minor loss}} \\
&+\overbrace{G_k\frac{A^2_{k,1}\sum_{i=1}^{m_k+1}(C^{\beta}_{k,i}-1)\frac{(\delta_k+\alpha_k-\pi_{k,i-1})^2}{A^2_{k,i}}}{\delta_k|\delta_k|^{n-1}}}^{\text{due to static pressure regain}} - 1.
\end{aligned} \tag{A11}$$