UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Ingegneria Civile, Edile e Ambientale (ICEA)

CORSO DI DOTTORATO DI RICERCA IN
SCIENZE DELL'INGEGNERIA CIVILE ED AMBIENTALE
XXIX CICLO

Data Assimilation Techniques for Leakage Detection in Water Distribution Systems

Coordinatore della Scuola: Ch.mo Prof. Stefano Lanzoni

Supervisore: Ch.mo Prof. Paolo Salandin

Dottoranda: Valentina Ruzza

Abstract

Leakage in pressurized water distribution systems is a major issue for water utilities today, because of the huge concerns over public health risks and the economic constraints on energy and resources. This thesis investigates innovative techniques for the detection of leakages in water distribution systems, relying on the calibration of network hydraulic models. The main goal is to suggest a method to reduce the costs of the field surveys currently required from the leakage detection activity on real systems.

An inverse model, based on the coupling between Kalman Filter based data assimilation techniques and network hydraulic models, is proposed and critically analyzed. The model is based on the knowledge of pressure heads, pipe flow rates and volume measurements, which can be easily obtained in any network with a limited effort and no technical troubles, with exception of the flow rate measurements.

The present work investigates different aspects of the proposed coupled model, related to the data assimilation technique used (Ensemble Kalman Filter or Ensemble Smoother), the type of hydraulic analysis developed (demand driven analysis through standard EpaNET or pressure driven analysis), the type of model parameters to be calibrated (the nodal leakage flow rates or the EpaNET emitter coefficients responsible for the nodal leakage flow rates), besides distinctions on the type of assimilated data and on the number and locations of available measurements.

Despite the fact that the success of the proposed technique depends on the specific features and topological structure of the network analyzed, this coupled model applied to synthetic water distribution systems proves to be effective for leakage detection and could be a competitive solution compared to the traditionally used district metering procedures in real world cases.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis investigates innovative techniques for the detection of leakages in water distribution systems (WDSs).

To some extent leakages plague all the WDSs and the problem turns mainly into water quality issues and concerns over public health risk, since the bursts are potential points for the transfer of contaminants from the water table into the system when pressure decreases during a shortcoming due to management or maintenance operations. Leakages result not only in loss of treated drinkable water, but also in wasting the energy and resources invested in its supply, transportation and distribution, which are major issues for water utilities today. Moreover, leakages reduce the system efficiency in satisfying the consumer requirements in terms of delivered discharge and pressure, besides causing long term damages to infrastructures and to the distribution system itself due to the hidden pressurized outflows.

A first estimate of the leakage scope in WDSs is often obtained from a balance between non-revenue water and the overall system intake, but the actual losses may however differ significantly from this estimate, which does not account for the apparent losses due to non-revenue consumption (e.g. water used for washing the streets or for the fire service). Alternative techniques to the traditional balance are thus required for a proper leakage assessment, but in any case the economic benefit deriving from the application of a leakage detection technique actually depends on its availability and effectiveness. At present, the audits help to identify parts of the WDS that have excessive leakage, however they do not provide information about the exact location of leaks requiring attention, thus leak detection surveys are undertaken via the preferential use of acoustic methods (Hunaidi et al., 2000; AWWA, 2008), which are costly both in terms of number of pipes to be examined and working time. The spatial assessment of water losses through network models is a challenging issue that can help to analyze real world problems when, as usual, a proper network segmentation and the detailed knowledge of water demand are not available, making the

identification of losses through district metering cumbersome. Moreover, the spatial distribution of water losses in WDSs strongly affects the calibration of the network models, that should be accomplished on the basis of data readily available as the metered consumption, the global flow rates and the pressure heads on a limited number of locations.

1.1 Purpose and scope

This thesis investigates a leakage localization method based on coupling data assimilation techniques based on the Kalman Filter, as the Ensemble Kalman Filter (EnKF) (Evensen, 1994) or the Ensemble Smoother (ES) (Van Leeuwen and Evensen, 1996), to the EpaNET (Rossman et al., 2000) software as a tool for the hydraulic simulation of the network. The Kalman Filter data assimilation techniques are able to manage the different uncertainties linked to the context, and at the same time are able to work as inverse models to calibrate the parameters involved in the modelling procedure.

This thesis proves that the proposed model allows for the definition of the network segments where significant water losses are more likely to occur, through the assimilation of a suitable number of measurements of nodal pressure (Pudar and Liggett, 1992; Chen and Zhang, 2006), pipe flow rate and cumulative incoming flow volume. The total inflow is commonly known, while nodal pressures can be monitored with a limited technical and economical effort. Although accurate measurements of pipe flow rates are not usual and require a good preliminary knowledge of the WDS, good practice and the newly developed design and renewal techniques generally assure the division of the water supply network into hydraulically independent districts that can be continuously monitored.

A comparison between EnKF and ES is realized and the adequacy of the EnKF for the analyzed problem is shown, the algorithm being able to manage the non-linear physical relationship that links water losses to the indirect measurements of pressure heads or volumes in a WDS. The EnKF technique coupled with a pressure driven hydraulic model, which better describes the physical relationship between pressures and flow rates in a WDS, is then verified on a more complicated system.

Given the great potential of the information technology tools currently available, the use of the proposed coupled model to assess the spatial distribution of leakages in a WDS is a promising tool that can help to solve real world problems reducing the costs associated with onsite acoustic surveys and, at the same time, making the calibration of hydraulic network models a more reliable procedure.

1.2 Literature review

1.2.1 Inverse models for leakage detection

Leak detection methodologies have gained the interest of the scientific community and many papers have been published on this topic in recent years, thanks to the increase in computer power and the availability of new numerical methods for WDSs analysis.

Leak detection in WDSs can be accomplished by solving an inverse problem using measurements of pressure and/or flow. In the work developed by Pudar and Liggett (1992), the problem is formulated with equivalent orifice areas of possible leaks as the unknowns. Minimization of the difference between measured and calculated heads produces a solution for the areas. In this analysis the usual assumption is made that all the demands occur at the nodes and the leaks are at the nodes as well. In this way leaks become simply additional demands, the location and quantities being unknown. In the ill posed inverse problem we know the characteristics of the system (pressures) and the demands but some quantities - the unaccounted for nodal outflows, leaks - are unknown. If the known quantities are extended to a sufficient number of pressures (i.e. pressure measurements), we can find the leaks. The leaks L_i are usually the unknowns, but can be expressed in terms of pressure by an orifice formula

$$L_i = C_{oi} A_{oi} \sqrt{\frac{2gp_i}{\gamma_w}} \tag{1.1}$$

in which C_{oi} is an orifice coefficient and A_{oi} is an equivalent orifice area. The values of A_{oi} become the unknowns instead of the L_i.

Inverse techniques can be used to calibrate WDSs, as discussed by Boulos and Wood (1990), who consider the case of unknown pipe parameters and boundary conditions.

Pérez et al. (2011) propose a leakage localization method based on the pressure measurements and pressure sensitivity analysis of nodes in a network. The methodology of leakage localization proposed in this paper is mainly based on standard theory of model-based diagnosis described for example in Gertler (1998) that has already been applied to water networks to detect faults in flow meters (Ragot and Maquin, 2006).

The work developed by Morosini et al. (2014) shows a leak detection approach based on a Bayesian calibration method. The methodology uses a newly formulated index which takes into account the variation of roughness in pipes between the calibrated models with and without leaks. The approach proved to be effective in finding leaks without a high computational cost, but the results depend crucially on the number and quality of the observed data.

Because the system characteristics are never known perfectly and no measurement can

be considered without error, and because the different kinds of measurements available are usually indirectly related to the model parameters, a suitable approach is necessary to reconcile information from multiple sources. A comprehensive approach to the problem of uncertainty in WDSs has been proposed within the frame of leak detection problem (Poulakis et al., 2003; Rougier, 2005) and real time management (Hutton et al., 2012).

Recently it has become common the solution of the inverse problem through the application of Kalman Filter based techniques, as the Ensemble Kalman Filter (EnKF) or the Ensemble Smoother (ES). These techniques follow a Monte Carlo approach, using an ensemble of model realizations to evaluate necessary statistics.

The first formulation of the EnKF was given by Evensen (1994). The EnKF has been found useful in various applications (e.g. meteorology) which deal with large and non-linear problems. This technique shows a fairly low computational cost compared to optimization processes (that need the explicit calculation of an objective function). Another benefit of the EnKF is that existing codes can be used to obtain model predictions, such as pressure heads and flows. With the prediction of such codes, the system is sequentially updated by incorporating the available observations.

Chen and Zhang (2006) apply the EnKF in the field of subsurface hydrology to update the hydraulic conductivity field by assimilating hydraulic head measurements, providing the capabilities of the method.

Dealing with subsurface hydrology, in Hendricks Franssen and Kinzelbach (2008) the EnKF is used for the joint updating of parameters and states. In their work the filter inbreeding problem is investigated in details and some solutions are proposed for its reduction. In Nowak (2009) the EnKF procedure moves toward pure parameter updating.

Zhou et al. (2011) have concentrated on the Gaussian hypothesis of the variables probability distribution function, that ensures the optimal working conditions for the Kalman Filter based methods. In this work a technique called Normal-Score Transform is applied to parameters and state variables so that the Kalman filtering equations will be applied on Gaussian variates.

Data assimilation methods are used in Okeya et al. (2014) to improve predictions of water demand and WDS states with the assumption that pipe roughness values and other hydraulic model parameters are known and the system has no leakage losses.

In their very recent paper Bragalli et al. (2016) use an innovative cascade of Ensemble Kalman Filters to assimilate the information deriving from sensors measuring pressure heads, flow in pipes and demands, with the objective of increasing knowledge in WDSs.

In contrast to the EnKF, the ES (Van Leeuwen and Evensen, 1996) analysis incorporates

all previous measurements and model states to compute an updated model state estimate at all previous measurement times. The ES scheme is a promising alternative to other inverse modelling techniques because of the low computational burden and the ability to run the algorithm entirely independent from the transient model simulation. ES applications are reported in literature (e.g. Bailey and Baù, 2010) for the estimation of the system parameter values.

1.2.2 Optimal sampling design

In order to increase knowledge in a WDS, measurements are performed through pressure sensors and flowmeters. An interesting problem is to find the most appropriate measurement network that optimizes leakage detection, using a minimum number of sensors.

Pudar and Liggett (1992) point out that measurement programs can be guided by the sensitivity matrix for maximum effectiveness. A methodology based on this matrix is proposed by Pérez et al. (2011) for optimizing leakage detection. The leakage sensitivity analysis evaluates the effect of a leakage on the pressure in a node. If this process is repeated for each node and possible leak, the sensitivity matrix S is obtained as follows:

$$S = \begin{pmatrix} \frac{\partial p_1}{\partial L_1} & \cdots & \frac{\partial p_1}{\partial L_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial p_n}{\partial L_1} & \cdots & \frac{\partial p_n}{\partial L_n} \end{pmatrix} \tag{1.2}$$

where each element s_{ij} measures the effect of leak L_j in the pressure of node p_i. It is extremely difficult to calculate S analytically in a real network, because of the huge non explicit and non-linear systems of equations that describe its dynamics. In the mentioned work the sensitivity matrix is generated by simulation as follows: the same leakage is introduced in each node and the corresponding increment of pressure is measured. Because some sensors are much more sensitive to all leakages than others, a normalization of sensitivity is needed so that the information provided by any node is comparable. Each row corresponding to a node with a sensor is divided by the maximum value of this row that corresponds to the leakage most important for that node. This procedure leads to the normalized sensitivity matrix $\bar{S}$:

$$\bar{S} = \begin{pmatrix} \frac{\partial s_{11}}{\partial \sigma_1} & \cdots & \frac{\partial s_{1n}}{\partial \sigma_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_{n1}}{\partial \sigma_n} & \cdots & \frac{\partial s_{nn}}{\partial \sigma_n} \end{pmatrix} \tag{1.3}$$

where $\sigma_i = max\{s_{i1}, ..., s_{in}\}$, $i = 1, ..., n$. This matrix shows how the most relevant leak is the one on the node itself, the maximum normalized sensitivity being on the diagonal.

Columns correspond to nodes with leaks and rows correspond to nodes with sensors. Each element is equal to zero when leakage j does not affect pressure in node i and it is equal to 1 when leakage j affects node i. This approach is a possible way to define sensor placement avoiding the optimization process.

A significant amount of previous research on sampling design has focused on the reliability aspects of WDSs (Xu and Goulter, 1998) and on the model calibration (Bush and Uber, 1998). To solve the problem of optimal sampling design various methods are suggested in literature. Kapelan et al. (2005) propose a methodology in which the sampling design of pressure loggers is formulated as a multi-objective problem that minimizes the calibrated model prediction uncertainty and cost. Morosini et al. (2014) present a method for sampling design based on sensitivity analysis (D-optimality criteria based) to find pipes and nodes that affect the hydraulic behaviour of the entire system. The essence of all these methods is in discovering sensitivity nodes in the network, which represent behaviour of all other nodes.

In some recent works the sensor placement methodology was aimed at finding leaks. In the work developed by Pérez et al. (2009) sensors are placed in the Barcelona network to detect a discrepancy in pressure due to leakage depending on its location. Promising results are obtained by Quevedo Casín et al. (2011) by using a fault isolation algorithm which correlates the residuals (generated by comparing available pressure measurements with their estimation using a model) with the fault sensitivity matrix. In a more recent study Pérez et al. (2014) investigate the optimal sensor distribution considering that some sensors are already installed in real networks, as flow sensors at the control points. Casillas et al. (2013) propose a genetic algorithm-based sensor placement method for leak location, consisting in minimizing the number of non-isolable leaks. The EnKF approach recently adopted by Bragalli et al. (2016) allows for the selection of selective and affordable monitoring networks.

1.3 Thesis outline

Chapter 2 gives the theoretical fundamentals of the techniques used. The Kalman Filter theory is developed, focusing on two derivative data assimilation methods, the EnKF and the ES.

In Chapter 3 the gradient algorithm for the solution of pipe networks is presented. The derivation of the recursive algorithm is then extended to include pumps. A pressure driven hydraulic network model is developed, that is a FORTRAN program which performs the three step procedure proposed by Todini (2003). Two example networks are reported for

the model application.

In Chapter 4, it is investigated the possibility of retrieving the spatial distribution of water losses by assimilating pressure head measurements. A comparison is developed between the EnKF and the ES techniques on the synthetic Anytown benchmark system. The EnKF generally outperforms the ES when the number of available measurements is higher, the recursive structure of the EnKF allowing for an effective management of the problem non-linearities.

In Chapter 5, the Normal-Score EnKF technique is coupled with a pressure driven hydraulic network model to investigate the possibility of retrieving the spatial distribution of water losses through the calibration of the EpaNET emitter coefficients responsible for the nodal leakages. The assimilated measurements are pressure heads, flow rates and volumes. The procedure is tested on the two synthetic networks of Anytown and Net3, characterized by a different topological complexity. The numerical experiments demonstrate that the success of the technique is directly proportional to the topological complexity of the network and to the cross correlation relationship between the leakage variables to be estimated and the potentially measured system variables.

Finally the conclusions summarize the main results of the thesis.

Chapter 2

Kalman Filter based data assimilation techniques

2.0.1 Introduction

There are many sources of uncertainty in any mathematical model of a system. In fact, the model objective is to represent critical modes of system response, so many effects are left unmodelled. Even effects which are modelled are necessarily approximations to what is observed and their parameters are not determined absolutely. Then disturbances, which we can neither control nor model deterministically, contribute to drive the system. Moreover sensors do not provide perfect and complete data about a system, devices being always noise corrupted.

A mean of extracting valuable information from a noisy signal must be provided. This problem is known as *data assimilation* and a possible solution is made by Kalman Filter based techniques. The Kalman Filter (KF) and its variants, the Ensemble Kalman Filter (EnKF) and the Ensemble Smoother (ES), are herein described.

2.1 The Kalman Filter

The Kalman Filter (KF) (Kalman, 1960) is a sequential data assimilation algorithm for linear dynamics and measurement processes with Gaussian error statistics. Given a linear forecast model and a series of time dependent observations affected by errors, the KF can sequentially incorporate new observations at the point when they become available, thus obtaining a least square estimation of the state of the system.

The KF consists of three main components. The first is a state vector, which usually includes model parameters and dependent variables. The second is a forecast model. The

third is an assimilation model, whose purpose is to combine the information coming from the forecast model and the observed data. The difference between the two sources of information is called innovation. The weight assigned to the innovation is determined through the state error covariance matrix and error covariance matrix of the observations (Evensen, 2003). The KF is shortly described by the equations in the sequel.

$$X_{t_i}^f = AX_{t_i-1}^u + e_{1t_i} \tag{2.1}$$

$$P_{t_i}^f = AP_{t_i-1}^u A^T + W_{t_i} \tag{2.2}$$

$$z_{t_i}^m = MX_{t_i} + e_{2t_i} \tag{2.3}$$

$$K_{t_i} = P_{t_i}^f M^T \left(MP_{t_i}^f M^T + R_{t_i} \right)^{-1} \tag{2.4}$$

$$X_{t_i}^u = X_{t_i}^f + K_{t_i} \left(z_{t_i}^m - MX_{t_i}^f \right) \tag{2.5}$$

$$P_{t_i}^u = \left(I - K_{t_i} M \right) P_{t_i}^f \tag{2.6}$$

where X is the state vector, which represents the state of the system, including model parameters, (dependent) variables, and other observations; z^m denotes the observation vector; M is the observation operator which represents the relationship between the state vector and the observation vector; P denotes the state error covariance matrix; K denotes the Kalman gain; R is the error covariance matrix of the observations; W denotes the covariance matrix of the model noise; A stands for the linear transition matrix (also referred to as matrix derivative or sensitivity matrix), which is a linear operator to forward the state from one time step to the next time step; e_1 and e_2 are independent white noises for the forecast model and the observations, drawn from multi-normal distributions with zero mean and covariance W and R, respectively; t_i denotes the time step; the superscript T stands for transpose; the superscript f and u indicate the forecast and update procedure, respectively.

Equation (2.1) represents the forecast procedure of the KF system at the time step t_i. The forecast model will run until new observations become available. The observation vector z^m is transformed from the true field through the observation operator M, allowing for observation errors e_2 in equation (2.3). The optimal least squares Kalman gain is calculated through equation (2.4). From equation (2.5) the new state vector is obtained as a weighted summation of the forecast state and the innovation. After the data assimilation step of equation (2.5), the new state error covariance P^u is given by equation (2.6). In general, the trace of P^u should be less than that of the forecast error covariance P^f given by equation (2.2).

2.2 The Ensemble Kalman Filter

The EnKF (Evensen, 1994) is based on the Monte Carlo approach, meaning that an ensemble of realizations is used to describe the problem and the error statistics. The basic concept is that the information conveyed by a full probability density function can be exactly represented by an infinite ensemble of model states. Suppose that we have N model states, each of dimension n. Each model state represents a point in an $n-$dimensional state space. This cloud of points in the state space can, in the limit when N goes to infinity, be described using a probability density function

$$\Phi(x) = \frac{dN}{N} \tag{2.7}$$

where dN is the number of points in a small unit volume. With the knowledge about either Φ or the ensemble representing Φ, we can calculate whichever statistical moment.

The essence of the EnKF is very similar to the KF one. The difference is that in the KF the error covariance matrix for the forecast estimate, P^f, is explicitly computed through

$$P^f = \left\langle \left(X^f - X\right)\left(X^f - X\right)^T \right\rangle \tag{2.8}$$

where the brackets denote an expectation value. P^f is then propagated in time through equation (2.2). However, the true state X is not known, and we therefore define the ensemble covariance matrix P_e^f around the ensemble mean $\langle X^f \rangle$.

$$\langle X^f \rangle \approx \frac{1}{NMC} \sum_{mc=1}^{NMC} X_{mc}^f \tag{2.9}$$

$$P^f \approx P_e^f = \frac{1}{NMC-1} \sum_{mc=1}^{NMC} \left[\left(X_{mc}^f - \langle X^f \rangle\right)\left(X_{mc}^f - \langle X^f \rangle\right)^T \right] \tag{2.10}$$

where now the average is over the ensemble. The subscript mc denotes the index of the ensemble members, and NMC denotes the total number of the ensemble members.

Thus, we can use an interpretation where the ensemble mean is the best estimate and the spreading of the ensemble around the mean is a natural definition of the error in the ensemble mean. There will clearly exist infinitively many ensembles with an error covariance equal to P_e^f. Thus, instead of storing a full covariance matrix, we can represent the same error statistics using an appropriate ensemble of model states. Given an error covariance matrix, an ensemble of finite size will always provide an approximation to the error covariance matrix. However, when the size of the ensemble N increases, the errors in the Monte Carlo sampling will decrease proportional to $\frac{1}{\sqrt{N}}$ (Evensen, 2009b).

In the EnKF the forecast model is performed on each ensemble member independently. Equation (2.1) becomes:

$$X^f_{mc(t_i)} = AX^u_{mc(t_i-1)} + e_{1mc(t_i)} \tag{2.11}$$

where A is the forecast operator to integrate the state with time, that can be either linear or non-linear. The observation vector at the time step t_i for each ensemble member is similar to equation (2.3):

$$z^m_{mc(t_i)} = MX_{(t_i)} + e_{2mc(t_i)} \tag{2.12}$$

where $MX_{(t_i)}$ is the observation data obtained from the true field and $e_{2mc(t_i)}$ is the observation noises. Following the previous interpretation, each observation is represented by an ensemble, where the mean is the actual measurement and the variance of the ensemble represents the measurement errors. The Kalman gain has exactly the same expression as its KF counterpart (equation 2.4). The updated ensemble of states is then computed similarly to equation (2.5):

$$X^u_{mc(t_i)} = X^f_{mc(t_i)} + K_{t_i}\left(z^m_{mc(t_i)} - MX^f_{mc(t_i)}\right) \tag{2.13}$$

with the new (posterior) error covariance matrix still given by equation (2.6), which can also be conveniently computed from the ensemble of updated states (2.13) using the formulae like (2.9) and (2.10). In the latter procedure, the posterior mean and covariance of the state vector can be selectively calculated whenever and wherever needed, eliminating the needs of keeping track of the whole covariance matrix.

2.2.1 Non-linear model dynamics

For a non-linear model where we appreciate that the model is not perfect and contains model errors, we can write it as a stochastic differential equation:

$$dx = u\,(x)\,dt + k\,(x)\,dq \tag{2.14}$$

where x is a random vector, u is a deterministic non-linear operator and $k\,(x)\,dq$ is a stochastic forcing term representing the random contribution from the model errors. In detail, dq describes a vector Brownian motion process with assigned covariance Wdt, while k is an operator.

When additive Gaussian model errors forming a Markov process are used, one can derive the Fokker-Planck equation (also named Kolmogorov's equation) which describes the time evolution of the probability density $\Phi\,(x)$ of the model state,

$$\frac{\partial \Phi}{\partial t} = \sum_i \frac{\partial (u_i \Phi)}{\partial x_i} = \frac{1}{2}\sum_{ij} \frac{\partial^2 \Phi \left(kWk^T\right)_{ij}}{\partial x_i \partial x_j} \tag{2.15}$$

where $i = 1, ..., n$, thus u_i is the component number i of the model operator u and kWk^T is the covariance matrix for the model errors.

If equation (2.15) could be solved in terms of Φ, it would be possible to calculate statistical moments for the model forecast to be used in the analysis scheme. The EnKF applies a Markov Chain Monte Carlo (MCMC) method to solve equation (2.15), by integrating the ensemble of model states forward in time, according to the stochastic model dynamics described by equation (2.14). This ensemble prediction is equivalent to solving the Fokker-Planck equation using an MCMC method. This procedure forms the backbone for the EnKF.

An advantage of the EnKF is that the effect of non-linear terms is retained, since each ensemble member is integrated independently by the model. In fact, the error covariance of the ensemble evolves according to

$$P^f_{e(t_i)} = FP_{e(t_i-1)}F^T + W_e + o(2) \tag{2.16}$$

where F coincides with matrix A of the KF. Equation (2.16) is again of the same form as is used in the standard KF, except of the extra $o(2)$ terms that may appear if u is non-linear and that are implicitly retained in the EnKF.

2.2.2 Computation efficiency of the EnKF

In the EnKF scheme, it is easy to find out that the observation operator M and the state error covariance P always appear together during the updating process, thus the product of M and P can be calculated instead of the actual covariance P. M is a linear operator with only 0s and 1s as its entries, indicating the product of M and P is simply selecting several lines from the matrix P. Computing the product of M and P first can simplify the computation greatly.

If the state vector contains n components and the observation vector contains m components (usually m being much smaller than n), the matrix P has the size $m \times n$. Therefore, in the EnKF system only m lines related to the observable components of the state vector need to be computed out of the whole covariance matrix P.

The posterior mean and covariance of the state vector can be selectively evaluated from the ensemble of updated realizations as needed. The reduced dimensionality in the covariance matrix as well as the approach for obtaining the required covariance constitutes a major advantage for the EnKF.

The EnKF system is also suitable for parallel computation, since each ensemble member works independently when performing forecasting and updating. Communication is only

necessary when calculating the state error covariance and the Kalman gain.

2.3 The Ensemble Smoother

The Ensemble Smoother (ES) (Van Leeuwen and Evensen, 1996) is a derivative of the EnKF scheme. In contrast to the EnKF, which incorporates all previous measurements to provide an updated model state at only the current time, the ES analysis incorporates all previous measurements and model states to compute an updated model state estimate at all previous measurement times, using both the spatial and temporal covariance of model results (Evensen, 2009a). In this way, previous model states continue to be updated at each measurement assimilation time, and thus the method can be used to reconstruct historical conditions (McLaughlin, 2002).

As discussed by Van Leeuwen (2001), the ES gives superior results compared to the EnKF at earlier times, since the model states can be updated back in time, but provides identical estimates of the system state at the current assimilation time as the EnKF (Evensen and Van Leeuwen, 2000).

The ES algorithm still follows a Monte Carlo approach, that is an ensemble of realizations is used to describe the system state and the measurement data, whose error is assigned by the modeller. All the error statistics are assumed to follow a normal frequency distribution.

2.3.1 Forecast step

Each model state X is forecasted to time $t + \Delta t$ on the basis of the model state at the current time X_t, parameters C, forcing terms c, boundary conditions b, and solution to the mathematical model φ, generating the prior system information $X^f_{t+\Delta t}$.

$$X^f_{mc(t+\Delta t)} = \varphi\left(X_{mc(t)}, C, c_t, b_t\right) \tag{2.17}$$

If this step occurs at the beginning of the model simulation, then each X_t coincides to the initial model state X_0.

2.3.2 Update step

At time $t + \Delta t$, measurement data $z^m_{t+\Delta t}$ from the true state are collected and perturbed with a Gaussian error to create the perturbed measurement vector $Z^m_{t+\Delta t}$. The assimilation procedure generates a posterior state estimate $X^u_{t+\Delta t}$ where the superscript u represents update.

$$X^u_{mc(t+\Delta t)} = X^f_{mc(t+\Delta t)} + K_{mc(t+\Delta t)}\left(Z^m_{mc(t+\Delta t)} - MX^f_{mc(t+\Delta t)}\right) \tag{2.18}$$

The matrix M maps model results at measurement locations to actual measurements, creating the residual at measurement locations. The lower the error in the measurement data, the more heavily the residual is weighted, and the model forecast values thus approach the measurement values. A residual equal to 0 signifies complete agreement between the model state and the true state. K is the Kalman gain matrix, with the same formulation as equation (2.4) of the KF and EnKF. Equation (2.18) is the same as equation (2.5) of the KF or equation (2.13) of the EnKF.

The forecast model error covariance matrix for the ES is still calculated through equation (2.10), and it is composed of spatial covariance terms between states at the same simulation time, as well as space-time covariance terms between states from different times.

Because the ES update routine is run only once using all previous model states and measurement data, and as such can be applied exclusive of the model simulation, it is an appealing approach for estimation of time-independent parameters. Besides, the ES computational burden is lower than the EnKF one and no iterative procedure is required.

2.4 Augmented state for parameter estimation

The state matrix X of both the EnKF or ES scheme can be augmented to include model parameter values, allowing the spatial covariance between parameter and state variables to correct not only the state, but also the parameters, of the model-estimated system. By doing so, the model itself is corrected to more precisely imitate the workings of the true system (e.g. Nowak, 2009).

In all the cases analyzed in this thesis, the data assimilation techniques are implemented considering an augmented system state for the estimation of parameters. The EnKF and ES are applied to non the non linear relationship that links the leakage parameters and the observed system variables. As in other applications reported in literature (e.g. Zhou et al., 2011) the KF based techniques are used with non Gaussian variables.

Chapter 3

Modelling and simulation of WDSs

3.0.1 Introduction

Given the geometry (layout, topography), physical properties (pipe lengths, diameters, roughness, reservoir shapes and levels, pump and valve characteristic curves) and the nodal demands (inflows and outflows), the problem usually referred to as the analysis of a water distribution network (WDN) consists of determining both the flows in every link (pipe, pump, valve) and the piezometric head (and pressure) in every node of the system, under the assumption that steady state flow has been reached.

Because the flows in the links and piezometric heads in the nodes are interrelated via the head loss-flow formula being used, the problem is normally restricted to determining one of them, i.e. either link flows or nodal piezometric heads.

The WDN analysis problem has received considerable attention since early 1936 when Cross (1936) proposed two methods for tackling the problem, one of them well suited for hand computation. Later on, Warga (1954), Martin and Peters (1963) and others (Shamir and Howard, 1968), proposed solutions based on the Newton-Raphson method. Wood and Charles (1972) introduced the Linear Theory Method algorithm. The gradient algorithm was originally proposed by Todini (1979) and Pilati and Todini (1984) and subsequently extended to incorporate pumps and other devices.

Nowadays all the algorithms that solve the network analysis problem use some fixed piezometric head values as boundary conditions, from which the distribution of the flow rates in the system is derived. Only the method from Wood and Charles (1972) still remains, which is based on the use of flow rates as boundary conditions. Such a method is rather inefficient from the computational point of view, thus leaving space to the more recent and effective gradient algorithm.

3.1 Mathematical background for network analysis problem

The steady state flow in a WDN is usually presented as that corresponding to the simultaneous fulfillment of a mass conservation and an *energy conservation* law.

- Mass conservation law, also referred to as mass or flow continuity law:

$$\sum_i Q_{ij} = q_j \tag{3.1}$$

  for all the nodes, with $j = 1, 2, ...nn$. Q_{ij} is the flow in link connecting nodes i and j. q_j is the nodal demand in node j. The summation is carried out over all the nodes i connected with the j-th node. Equation (3.1) represents a system of nn linear equations in the unknown Q_{ij}, where nn is the number of nodes in the network.

- *Energy conservation* law, usually expressed in terms of head losses (gain) along a loop or energy path:

$$\sum_k h_{ij} = \delta E_k \tag{3.2}$$

  for all the paths, where $k = 1, 2, ...nl$. h_{ij} is the head loss in link connecting nodes i and j. The summation is carried out over one sequence (path) of links going from one known head node to another. Normally a reservoir will be the known head node. δE_k is the energy (level) difference between the starting and final node of the k-th path. nl is the number of loops plus energy paths in the network. A loop is the particular case of a path starting and ending at the same node (i.e. $\delta E_k = 0$). Eq. (3.2) represents a system of nl equations. Instead of being an *energy conservation* law, eq. (3.2) is actually establishing a continuity of potential condition, since it holds for any set of nodal piezometric head.

In addition to mass and potential continuity, a head loss-flow relationship must also be satisfied by each link. The Hazen-Williams and Darcy-Weisbach formulae are some of the most widely used head loss-flow relationships for computing head losses in pipe networks; quadratic functions are used in the case of valves and pumps. Because the relationships are non-linear in the flows, the system of simultaneous equations produced by (3.1) and (3.2) is a non-linear one and no direct solution is possible.

3.2 The global gradient formulation

The gradient algorithm may be regarded as a bridge between the optimization based and the Newton-Raphson based techniques in that it starts from the minimization of a slightly modified content model (Collins et al., 1978) in order to prove the existence and uniqueness of the solution, which is the key to the unconditional convergence of the method. The problem is algebraically reconducted to the recursive solution of a linear system of size equal to the number of unknown nodal heads and a matrix projection of the results over the unknown pipe flows. The special structure of the resulting system matrix, a sparse Stieltjes matrix, symmetrical, positive definite which non zero elements can be stored in number of nodes + number of pipes locations, allows for an efficient solution by using the Incomplete Cholesky Factorization/Modified Conjugate Gradient algorithm (ICF/MCG) due to Kershaw (1978).

3.2.1 Necessary conditions

Todini (1979) and Pilati and Todini (1984) show that the necessary conditions for the steady state flow are simply the simultaneous fulfilment of the mass conservation law (nodal balance) and a non-linear relationship describing the head loss-flow phenomena in pipes. Both conditions can be expressed in the following compact system of equations:

$$\begin{bmatrix} A_{12}H + F(Q) \\ A_{21}Q \end{bmatrix} = \begin{bmatrix} -A_{10}H_0 \\ q \end{bmatrix} \tag{3.3}$$

$A_{12} = A_{21}^T$: (np, nn) unknown head nodes incidence matrix
$A_{10} = A_{01}^T$: (np, no) fixed head nodes incidence matrix
$Q^T = [Q_1, ..., Q_{np}]$: $(1, np)$ flow rates in each pipe
$q^T = [q_1, ..., q_{nn}]$: $(1, nn)$ nodal demands
$H^T = [H_1, ..., H_{nn}]$: $(1, nn)$ unknown nodal heads
$H_0^T = [H_{01}, ..., H_{0no}]$: $(1, no)$ fixed nodal heads
$F^T(Q) = [f_1, ..., f_{np}]$: $(1, np)$ law expressing head losses in pipes $(f_i(Q_i))$
with nn the number of nodes with unknown heads, no the number of nodes with fixed head and np the number of pipes with unknown flow rate.

$$A_{12}(i,j) = \begin{array}{ll} 1 & \text{if flow of pipe } i \text{ enters node } j \\ 0 & \text{if pipe } i \text{ and node } j \text{ are not connected} \\ \text{-1} & \text{if flow of pipe } i \text{ leaves node } j \end{array}$$

A_{10} is defined similarly to A_{12} for fixed head nodes.

$$A_{10}(i,j) = \begin{array}{rl} 1 & \text{if flow of pipe } i \text{ enters node 0} \\ 0 & \text{if pipe } i \text{ and node 0 are not connected} \\ \text{-1} & \text{if flow of pipe } i \text{ leaves node 0} \end{array}$$

The system represented by equation (3.3) may have more than one solution depending upon the shape of $f_i(Q_i)$. If all $f_i(Q_i)$ are monotonically increasing functions, it can be proved that the solution of system (3.3) exists and is unique (Pilati and Todini, 1984). A more general approach to the proof of the existence and uniqueness of the solution can be derived as follows. Assuming the Hazen-Williams head loss function, one can write for each pipe:

$$f_i(Q_i) = R_i |Q_i|^{n_i - 1} Q_i \tag{3.4}$$

with R_i a constant. After integration of f_i from 0 to Q_i:

$$minC(Q) = \sum_{i=1}^{np} \frac{R_i |Q_i|^{n_i+1}}{n_i + 1} + \sum_{j=1}^{no} H_{0j} \sum_{i=1}^{np} A_{01}(j,i) Q_i \tag{3.5}$$

subjected to

$$\sum_{i=1}^{np} A_{21}(j,i) Q_i - q_j = 0 \tag{3.6}$$

with $j = 1, nn$. Due to the definition of $f_i(Q_i)$ there is no need for the additional constraints $Q_i \geq 0$. This allows to transform the constrained minimization into an unconstrained one by means of Lagrange multipliers, i.e.

$$min\Gamma(Q,\lambda) = \sum_{i=1}^{np} \frac{R_i |Q_i|^{n_i+1}}{n_i + 1} + \sum_{j=1}^{no} H_{0j} \sum_{i=1}^{np} A_{01}(j,i) Q_i + \sum_{j=1}^{nn} \lambda_j \sum_{i=1}^{np} (A_{21}(j,i) Q_i - q_j) \tag{3.7}$$

Since all the R_i are positive when all $n_i > 0$, Γ is convex and the solution of this problem exists and is unique. This coincides with the sufficient condition for a minimum. The solution can thus be found by imposing all the necessary conditions for an extreme:

$$\frac{\partial \Gamma}{\partial Q_i} = 0 \tag{3.8}$$

with $i = 1, np$ and

$$\frac{\partial \Gamma}{\partial \lambda_j} = 0 \tag{3.9}$$

with $j = 1, nn$, to get, in matrix form:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix} \begin{bmatrix} Q \\ \lambda \end{bmatrix} = \begin{bmatrix} -A_{10} H_0 \\ q \end{bmatrix} \tag{3.10}$$

where

$$A_{11} = \begin{bmatrix} R_1|Q_1|^{n_1-1} & & \\ & R_i|Q_i|^{n_i-1} & \\ & & R_{np}|Q_{np}|^{n_{np}-1} \end{bmatrix} \tag{3.11}$$

is an $A_{11}\,(np, np)$ diagonal matrix.

By comparison with equation (3.3) it is immediate to assign a physical meaning to the Lagrange multipliers: they represent in fact the unknown nodal heads. Finally after the substitution $\lambda = (n+1)\,H$ one gets:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix} \begin{bmatrix} Q \\ H \end{bmatrix} = \begin{bmatrix} -A_{10}H_0 \\ q \end{bmatrix} \tag{3.12}$$

Equation (3.12) equals equation (3.3) when the head losses are given by equation (3.4).

The upper part of the system (3.12) represents the head loss-flow relationships (np non-linear equations in Q) while the lower part corresponds to the nodal flow balances (nn linear equations in H).

3.2.2 Derivation of the gradient method

In order to solve the system of non-linear equations (3.12), the Newton-Raphson technique can be used, provided that matrix A_{11} does not become singular, which happens when the heads at the extremes of a pipe are identical and consequently the flow in the pipe vanishes. This problem can be avoided by defining a lower bound for the elements of matrix A_{11}.

The Newton-Raphson iterative scheme can thus be obtained by differentiating both sides of equation (3.12) with respect to Q and H to get:

$$f_1 = A_{11}Q + A_{12}H + A_{10}H_0 = 0 \tag{3.13}$$

$$f_2 = A_{21}Q - q = 0 \tag{3.14}$$

$$\frac{\partial f_1}{\partial Q} = A_{11}\frac{\partial Q}{\partial Q}dQ + \frac{\partial A_{11}}{\partial Q}QdQ + \frac{\partial (A_{12}H)}{\partial Q}dQ + \frac{\partial (A_{10}H_0)}{\partial Q}dQ \tag{3.15}$$

$$\frac{\partial f_1}{\partial H} = \frac{\partial (A_{11}Q)}{\partial H}dH + A_{12}\frac{\partial H}{\partial H}dH + \frac{\partial (A_{10}H_0)}{\partial H}dH \tag{3.16}$$

$$\frac{\partial f_2}{\partial Q} = A_{21}\frac{\partial Q}{\partial Q}dQ - \frac{\partial q}{\partial Q}dQ \tag{3.17}$$

$$\frac{\partial f_2}{\partial H} = \frac{\partial (A_{21}Q)}{\partial H}dH - \frac{\partial Q}{\partial H}dH \tag{3.18}$$

Equation (3.15) can be rewritten as

$$\frac{\partial f_1}{\partial Q} = R_i|Q_i|^{n_i-1}dQ_i + (n_i-1)\,R_i|Q_i|^{n_i-2}Q_i dQ_i \tag{3.19}$$

$$= (1+n_i-1)\,R_i|Q_i|^{n_i-1}dQ_i = NA_{11}dQ \tag{3.20}$$

and after simplifications the system becomes

$$\begin{bmatrix} NA_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix} \begin{bmatrix} dQ \\ dH \end{bmatrix} = \begin{bmatrix} dE \\ dq \end{bmatrix} \tag{3.21}$$

with $N\,(np, np)$ being the diagonal matrix

$$N = \begin{bmatrix} n_1 & & \\ & n_i & \\ & & n_{np} \end{bmatrix} \tag{3.22}$$

and where

$$dE = A_{11}Q^k + A_{12}H^k + A_{10}H_0 = 0 \tag{3.23}$$

$$dq = A_{21}Q^k - q = 0 \tag{3.24}$$

are the residuals to be iteratively reduced to zero and Q^k e H^k the flows and heads at iteration k. Assuming

$$NA_{11} = D^{-1} \tag{3.25}$$

(and therefore $DA_{11} = N^{-1}$, N, A_{11} and D being diagonal), the inverse of the system matrix can be obtained analytically (Ayres, 1962) by partitioning:

$$\begin{bmatrix} D^{-1} & A_{12} \\ A_{21} & 0 \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{3.26}$$

with

$$B_{11} = D - DA_{12}\,(A_{21}DA_{12})^{-1}\,A_{21}D \tag{3.27}$$

$$B_{12} = DA_{12}\,(A_{21}DA_{12})^{-1} \tag{3.28}$$

$$B_{21} = (A_{21}DA_{12})^{-1}\,A_{21}D \tag{3.29}$$

$$B_{22} = -\,(A_{21}DA_{12})^{-1} \tag{3.30}$$

The solution of equation (3.21) can be found bearing in mind that:

$$dQ = B_{11}dE + B_{12}dQ \tag{3.31}$$

$$dH = B_{21}dE + B_{22}dQ \tag{3.32}$$

By substituting for equations (3.27), (3.28), (3.29) and (3.30) into equations (3.31) and (3.32) to give:

$$\begin{aligned} dH &= \left(A_{21}DA_{12}\right)^{-1}A_{21}D\left[A_{11}Q^k + A_{12}H^k + A_{10}H_0\right] + \\ &\quad - \left(A_{21}DA_{12}\right)^{-1}\left(A_{21}Q^k - q\right) \qquad (3.33) \\ &= \left(A_{21}DA_{12}\right)^{-1}A_{21}DA_{11}Q^k + \left(A_{21}DA_{12}\right)^{-1}A_{21}DA_{12}H^k + \\ &\quad + \left(A_{21}DA_{12}\right)^{-1}A_{21}DA_{10}H_0 - \left(A_{21}DA_{12}\right)^{-1}A_{21}Q^k + \left(A_{21}DA_{12}\right)^{-1}q \qquad (3.34) \\ &= H^k + \left(A_{21}DA_{12}\right)^{-1}\left[A_{21}D\left(A_{11}Q^k + A_{10}H_0\right) + \left(q - A_{21}Q^k\right)\right] \qquad (3.35) \end{aligned}$$

$$\begin{aligned} dQ &= \left[D - DA_{12}\left(A_{21}DA_{12}\right)^{-1}A_{21}D\right]\left[A_{11}Q^k + A_{12}H^k + A_{10}H_0\right] + \\ &\quad + DA_{12}\left(A_{21}DA_{12}\right)^{-1}\left(A_{21}Q^k - q\right) \qquad (3.36) \\ &= DA_{11}Q^k + DA_{12}H^k + DA_{10}H_0 - DA_{12}\left(A_{21}DA_{12}\right)^{-1}A_{21}DA_{11}Q^k + \\ &\quad - DA_{12}\left(A_{21}DA_{12}\right)^{-1}A_{21}DA_{12}H^k - DA_{12}\left(A_{21}DA_{12}\right)^{-1}A_{21}DA_{10}H_0 + \\ &\quad + DA_{12}\left(A_{21}DA_{12}\right)^{-1}A_{21}Q^k - DA_{12}\left(A_{21}DA_{12}\right)^{-1}q \qquad (3.37) \\ &= D\left(A_{11}Q^k + A_{10}H_0\right) + \\ &\quad - DA_{12}\left\{\left(A_{21}DA_{12}\right)^{-1}\left[A_{21}D\left(A_{11}Q^k + A_{10}H_0\right) + \left(q - A_{21}Q^k\right)\right]\right\} \qquad (3.38) \end{aligned}$$

Substituting for equation (3.35) into equation (3.38), bearing in mind the definition of D and that:

$$\begin{aligned} dQ &= Q^k - Q^{k+1} \qquad (3.39) \\ dH &= H^k - H^{k+1} \qquad (3.40) \end{aligned}$$

one finally obtains the recursive Newton-Raphson algorithm:

$$\begin{aligned} H^k - H^{k+1} &= H^k + \left(A_{21}DA_{12}\right)^{-1}\left[A_{21}D\left(A_{11}Q^k + A_{10}H_0\right) + \left(q - A_{21}Q^k\right)\right] \qquad (3.41) \\ H^{k+1} &= -\left(A_{21}N^{-1}A_{11}^{-1}A_{12}\right)^{-1}\left[A_{21}N^{-1}A_{11}^{-1}A_{11}Q^k\right] + \\ &\quad - \left(A_{21}N^{-1}A_{11}^{-1}A_{12}\right)^{-1}\left[A_{21}N^{-1}A_{11}^{-1}A_{10}H_0 + q - A_{21}Q^k\right] \qquad (3.42) \\ &= -\left(A_{21}N^{-1}A_{11}^{-1}A_{12}\right)^{-1}\left[A_{21}N^{-1}\left(Q^k + A_{11}^{-1}A_{10}H_0\right)\right] + \\ &\quad - \left(A_{21}N^{-1}A_{11}^{-1}A_{12}\right)^{-1}\left(q - A_{21}Q^k\right) \qquad (3.43) \end{aligned}$$

$$
\begin{aligned}
Q^k - Q^{k+1} &= -DA_{12}\left(A_{21}DA_{12}\right)^{-1}\left[A_{21}D\left(A_{11}Q^k + A_{10}H_0\right) + \left(q - A_{21}Q^k\right)\right] + \\
&\quad + D\left(A_{11}Q^k + A_{10}H_0\right) \qquad (3.44)\\
&= N^{-1}A_{11}^{-1}A_{11}Q^k + N^{-1}A_{11}^{-1}A_{10}H_0 + \\
&\quad -N^{-1}A_{11}^{-1}A_{12}\left(A_{21}N^{-1}A_{11}^{-1}A_{12}\right)^{-1}\left(A_{21}N^{-1}A_{11}^{-1}A_{11}Q^k\right) + \\
&\quad -N^{-1}A_{11}^{-1}A_{12}\left(A_{21}N^{-1}A_{11}^{-1}A_{12}\right)^{-1}\left(A_{21}N^{-1}A_{11}^{-1}A_{10}H_0\right) + \\
&\quad -N^{-1}A_{11}^{-1}A_{12}\left(A_{21}N^{-1}A_{11}^{-1}A_{12}\right)^{-1}\left(q - A_{21}Q^k\right) \qquad (3.45)\\
Q^{k+1} &= Q^k - N^{-1}Q^k - N^{-1}A_{11}^{-1}A_{10}H_0 - N^{-1}A_{11}^{-1}A_{12}H^{k+1} \qquad (3.46)\\
Q^{k+1} &= \left(I - N^{-1}\right)Q^k - N^{-1}A_{11}^{-1}\left(A_{12}H^{k+1} + A_{10}H_0\right) \qquad (3.47)
\end{aligned}
$$

where A_{11} is computed using Q^k.

Summarizing the network analysis problem can be reconducted to the iterative solution of a system of linear equations of size equal to the number of nodes nn plus a scalar projection and a linear combination of the results, of size equal to the number of pipes np.

The linear equation (3.43) can be conveniently solved by using the Incomplete Cholesky Factorization/Modified Conjugate Gradient algorithm (ICF/MCG) due to Kershaw (1978).

3.3 Derivation of the recursive algorithm extended to include pumps

Salgado et al. (1988) extended the original gradient method to incorporate pumps into the system. A physically based algorithm is introduced for modelling pressure control devices, which fits within the framework of the gradient method.

The Hazen-Williams and Darcy-Weisbach formulae are widely accepted for describing the head loss-flow phenomena in pipes; quadratic (or nearly) functions are used for the same purposes in the case of valves and pumps. Thus, a general relationship between flow and head loss (or gain, in the case of pumps) may be written as:

$$h_i = r_i Q_i^n + v_i \qquad (3.48)$$

with $i = 1, ..., np$. h_i $[m]$ is the head loss in the branch, while n is a formula dependent exponent, typically 1.85-2. r_i and v_i are the characteristic parameters. For pipes and valves we drop the constant term (i.e. $v_i = 0$), for pumps both parameters are needed. These values are usually supplied by the manufacturer, or are determined via laboratory or field head loss-flow measurements, for different operating conditions. np is the number of branches (pipes, pumps, valves).

Equation (3.48) is a set of np non-linear equations. Using the link to node topological matrix A_{12}, we can express the head loss or gain of each link connecting two different nodes as

$$A_{11}Q + A_{12}H = -A_{10}H_0 \tag{3.49}$$

where:

$$A_{11} = \begin{bmatrix} R_1|Q_1|^{n_1-1} + \dfrac{v_1}{Q_1} & & \\ & R_i|Q_i|^{n_i-1} + \dfrac{v_i}{Q_i} & \\ & & R_{np}|Q_{np}|^{n_{np}-1} + \dfrac{v_{np}}{Q_{np}} \end{bmatrix} \tag{3.50}$$

On the other hand, the mass balance at each node can be written as

$$A_{21}Q = q \tag{3.51}$$

with the same meaning of variables.

The system of equations (3.49) and (3.51) can be set in a compact form by the matrix equation (3.12). The lower part caters for mass conservation and the upper part for the flow-head loss relationship. Due to the fact that the relationships (3.48) and (3.49) are non-linear in the flows, the system of simultaneous equations represented by (3.12) is non-linear and a direct solution is not possible. Some form of linear approximation is needed in order to solve the problem, leading to the formulation of the iterative gradient algorithm. On applying the gradient operator to the system of equations (3.12) we get:

$$\begin{bmatrix} NA'_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix} \begin{bmatrix} dQ \\ dH \end{bmatrix} = \begin{bmatrix} dE \\ dq \end{bmatrix} \tag{3.52}$$

where now

$$A'_{11} = \begin{bmatrix} r_1|Q_1|^{n_1-1} & & \\ & r_i|Q_i|^{n_i-1} & \\ & & r_{np}|Q_{np}|^{n_{np}-1} \end{bmatrix} = -\begin{bmatrix} \dfrac{v_1}{Q_1} & & \\ & \dfrac{v_i}{Q_i} & \\ & & \dfrac{v_{np}}{Q_{np}} \end{bmatrix} \tag{3.53}$$

where $N\,(np, np)$ is the diagonal matrix of the exponents n of the head loss-flow relationship.

At an intermediate iteration, an $(np, 1)$ residual vector dE and an $(nn, 1)$ residual vector dq can be computed through equations (3.23) and (3.24), which represent the energy imbalance at each link and the flow imbalance at each node respectively.

We seek the solution of equation (3.52)

$$\begin{bmatrix} dQ \\ dH \end{bmatrix} = \begin{bmatrix} NA'_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix}^{-1} \begin{bmatrix} dE \\ dq \end{bmatrix} \tag{3.54}$$

Following a similar approach to Pilati and Todini (1984), the inverse of the block-triangular matrix in equation (3.54) can be computed as another block matrix:

$$\begin{bmatrix} NA'_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix}^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{3.55}$$

On using:

$$G = NA_{11} \tag{3.56}$$

and computing the blocks of the inverse in equation (3.55), we get:

$$B_{11} = G^{-1} - G^{-1}A_{12}\left(A_{21}G^{-1}A_{12}\right)^{-1}A_{21}G^{-1} \tag{3.57}$$

$$B_{22} = -\left(A_{21}G^{-1}A_{12}\right)^{-1} \tag{3.58}$$

$$B_{12} = G^{-1}A_{12}\left(A_{21}G^{-1}A_{12}\right)^{-1} \tag{3.59}$$

$$B_{21} = \left(A_{21}G^{-1}A_{12}\right)^{-1}A_{21}G^{-1} \tag{3.60}$$

Then rewriting system (3.55) as:

$$dQ = B_{11}dE + B_{12}dq \tag{3.61}$$

$$dH = B_{21}dE + B_{22}dq \tag{3.62}$$

equations (3.23), (3.24), (3.57), (3.58), (3.59) and (3.60) are substituted into equations (3.61) and (3.62), which after some reordering gives:

$$\begin{aligned} dQ = {} & \left[I - G^{-1}A_{12}\left(A_{21}G^{-1}A_{12}\right)^{-1}A_{21}\right]G^{-1}\left(A_{11}Q_i + A_{10}H_0\right) + \\ & + \left[G^{-1}A_{12}\left(A_{21}G^{-1}A_{12}\right)^{-1}\right]\left(A_{21}Q_i - q\right) \end{aligned} \tag{3.63}$$

$$\begin{aligned} dH = {} & \left[\left(A_{21}G^{-1}A_{12}\right)^{-1}A_{21}G^{-1}\right]\left(A_{11}Q_i + A_{12}H_i + A_{10}H_0\right) + \\ & - \left(A_{21}G^{-1}A_{12}\right)^{-1}\left(A_{21}Q_i - q\right) \end{aligned} \tag{3.64}$$

On considering:

$$dQ = Q_i - Q_{i+1} \tag{3.65}$$

$$dH = H_i - H_{i+1} \tag{3.66}$$

and replacing equations (3.65) and (3.66) we obtain:

$$Q_{i+1} = \left[I - \left(NA'_{11}\right)^{-1}A_{11}\right]Q_i - \left(NA'_{11}\right)^{-1}\left(A_{12}H_{i+1} + A_{10}H_0\right) \tag{3.67}$$

$$\left[A_{21}\left(NA'_{11}\right)^{-1}A_{12}\right]H_{i+1} = -\left[A_{21}\left(NA'_{11}\right)^{-1}\left(A_{11}Q_i + A_{10}H_0\right) + \left(-A_{21}Q_i + q\right)\right] \tag{3.68}$$

Equations (3.67) and (3.68) are the coupled system which has to be solved recursively.

3.4 A pressure driven approach to the network analysis problem

For the past few years the need has been felt of solving the network analysis problem, in case the piezometric head on some nodes is insufficient to deliver the required demand.

The methods used so far are based on a demand driven approach (DDA), which considers the demand q_j at a node as a fixed problem constraint, assuming there is always enough power at the node to allow for the entire demand $q_{j,r}$. The implicit hypothesis is that the piezometric head H_j at the node, which is a problem unknown, is always sufficient to satisfy $q_{j,r}$. This assumption is actually realistic and the analysis results are correct, only when the head at the node is greater than or equal to the minimum head required to satisfy the demand ($H_j \geq H_{j,s}$). If the power condition is instead unsatisfied, the analysis highlights the critical nodes for which $H_j \leq H_{j,s}$. In this case the fixed nodal flow rate is incompatible with the power H_j calculated through a DDA and the two quantities are uncorrelated, appearing q_j to be completely independent from H_j. A DDA approach is widely accepted when the goal of the hydraulic analysis is the network design, but it appears to be inadequate when an extended period simulation is performed.

The pressure driven analysis or head driven analysis (PDA) is based on a completely different approach, which calculates the q_j values at the nodes as a function of the available pressure H_j. The goal of the PDA approach is the fulfilment of the flow rate-head relationship $q_j = f(H_j)$, other than the classical continuity and flow equations.

3.5 The proposed three step approach

Todini (2003) shows that it is possible to correctly solving the WDN problem when the head is insufficient, by using a three step technique. The proposed approach stems from three basic considerations.

1. Equation (3.12) derives from the minimization of a convex functional with linear constraints at the nodes.

2. If the head is really insufficient with a negative pressure, no water will be drafted from the taps.

3. If the head is small but the pressure is non-negative, thus insufficient to allow drafting the actual demand, a reduced demand may be satisfied. The logic says that what will

be drafted from the taps is the possible maximum (given that the actual demand is higher) that will make the pressure drop to zero at the node.

With these simple considerations in mind, it is possible to develop an extremely simple procedure that can allow to find correct solutions to the head driven problem. One must:

1. Start by solving the WDN in the conventional manner with fixed demands. If all the constraints are satisfied, namely all the $H_i \geq H_i^*$, the found solution is the correct one.

2. If some of the nodes show an insufficient head $H_i \leq H_i^*$, solve a WDN problem setting $H_i = H_i^*$ in non-satisfied nodes and compute $\widehat{q}_i$, the maximum demand compatible with this constraint. Three possibilities inevitably descend from this solution, either $\widehat{q}_i \geq q_i$, $0 \leq \widehat{q}_i \leq q_i$ or $\widehat{q}_i < 0$.

3. At this point a third step is needed which will require replacing a number of constraints.

   - At the nodes where $\widehat{q}_i \geq q_i$, meaning that there is enough power at that node to allow for the entire demand, the original constraints is set back in terms of demand, where the demand will be again equal to q_i.
   - At the nodes where $0 \leq \widehat{q}_i \leq q_i$, meaning that the power is insufficient to deliver the entire demand, the head constraint is retained, since the users will inevitably try to draw as much water as possible.
   - At the nodes where $\widehat{q}_i < 0$, meaning that the power is insufficient to provide any water, the original constraint is set back in terms of demand but with demand $q_i = 0$.

This procedure is demonstrated to converge to the right solution, successfully dealing with the cases of insufficient head that may occur.

3.6 Introduction to NETAN_HD

NETAN_HD (NETwork ANalysis Head Driven) is a FORTRAN programming language code, which performs the analysis of a WDN through a more realistic pressure driven approach. The code uses the gradient algorithm originally proposed by Pilati and Todini (1984), which is embedded into the three step procedure proposed by Todini (2003) to correctly solve the network analysis problem when the head is insufficient.

The analysis is based on the steady flow hypothesis. The head losses in pipes are computed using the Darcy-Weisbach formula:

$$h = \frac{f_f l}{12.1 d^5} Q^2 \tag{3.69}$$

where h is the head loss, f_f is the friction factor, l is the pipe length, d is the pipe diameter and Q is the pipe flow.

Concerning the Darcy-Weisbach relationship, the friction factor is calculated using the iterative implicit Colebrook-White formula:

$$\frac{1}{\sqrt{f_f}} = -2Log_{10}\left(\frac{e}{3.71d} + \frac{2.52}{Re\sqrt{f_f}}\right) \tag{3.70}$$

where e is the pipe roughness.

Prior to the calculation of the head losses, the Re number is verified in order to define the network links where the laminar flow condition occurs. In this case a linear relationship is used for the head loss calculation:

$$h = 4.153\frac{l}{\nu d^5} Q \tag{3.71}$$

where ν is the kinematic water viscosity.

NETAN_HD is able to verify both looped and not looped pressurized networks, considering both nodal demands and distributed withdrawals. In this case the computation of the head losses is based on the Messina method, which equally subdivides the distributed withdrawals along the pipes between the two extremes of a pipe. The method works then with a fictitious flow rate, which is defined as the equivalent flow rate producing the same head losses as distributed withdrawals do. This technique substitutes the distributed flow rate in the pipe P with two concentrated withdrawals $\frac{P}{2}$ at the extremes. The fictitious flow rate Q^* and the two flow rates entering Q_e and going out Q_u are linked through the relationships:

$$Q^* = Q_e - \frac{P}{2} \tag{3.72}$$

$$Q_u = Q^* - \frac{P}{2} \tag{3.73}$$

The calculation can be developed using and assuming the following relation to hold for the head losses:

$$h = \alpha\beta Q^*|Q^*| \tag{3.74}$$

where α is the pipe resistance factor, while β is the corrective factor due to the distributed withdrawals, which can assume the following values:

$$\beta = 1 + \frac{1}{12}\left(\frac{P}{Q^*}\right)^2 \tag{3.75}$$

when $0 < \dfrac{P}{|Q^*|} < 2$

$$\beta = \frac{2}{3}\frac{|Q^*|}{P} + \frac{1}{2}\frac{P}{|Q^*|} \tag{3.76}$$

when $\dfrac{P}{|Q^*|} > 2$

NETAN_HD also allows for the network analysis when pumps are present, which deliver a flow directly into the system. The delivered flow rate and hydraulic head are computed for each pump. The general relationship between head gain and flow rates is given for pumps as:

$$h_p = H_p - R_p Q^n \tag{3.77}$$

where n is equal to the coefficient used for the head losses. In this way the gain-flow rate relationship for pumps can be directly inserted into the system of equations of the gradient algorithm. Each couple of values H_p and R_p defines a working interval for the pump.

NETAN_HD requires an input file *rete.inp*, whose structure is described later, and produces an output file risultati.out that contains the results of the network flow analysis in terms of hydraulic heads and nodal flow rates for the nodes, and pipe discharges, velocity and head losses for the pipes. The hydraulic heads and flow rates are given for the pump elements.

3.7 Input file rete.inp

The input file *rete.inp* is created from the user. It contains the general network data, the node coordinates and properties, and the characteristics of pipe and pump elements. The variables required in the input file are listed in table (3.3).

Type	Variable	Description	Unit
I	NN	Number of network nodes	
I	NT	Number of network elements (pipes and pumps)	
I	NLN	Number of nodes with fixed demands	
I	NBN	Number of nodes with fixed head (sources)	
R	EPS	Tolerance	
I	NMAX	Maximum number of iterations	
R	CPG	Multiplier for nodal demands (steady state)	
I	ICOOR	Index for the reading of the node characteristics	
		ICOOR=1 given pipe lengths	

Table continues on next page

Table continues from previous page

Type	Variable	Description	Unit
		ICOOR=3 given node coordinates	
I	NNOD	Number of the network node (ordered)	
T	nID	Name of the network node	
R	QT	Z coordinate value	m
R	X1	X coordinate value	
R	Y	Y coordinate value	
I	IS	Index for the imposition of the boundary conditions	
		IS=0 variable head node	
		IS=1 fixed head node (source)	
R	QK	Required nodal demand	ls^{-1}
R	XX	Value of the fixed piezometric head	m
R	Hmin	Minimum required head at the node	m
I	Nel	Number of the network element (pipe or pump)	
T	pID	Name of the pipe or pump element	
I	N1	Starting pipe node	
I	N2	Ending pipe node	
R	D	Pipe diameter	mm
R	L	Pipe length	m
R	e	Pipe roughness	mm
R	q	Distributed withdrawals along the pipes	ls^{-1}
I	IPP	Number of pumps	
T	ppID	Name of the pump element	
I	NPT	Number of points of the characteristic curve	
R	QP	Pump delivered flow rate	ls^{-1}
R	HP	Pump delivered piezometric head	m

Variable types: I = integer, R = real, T = text.

Table 3.3: *Structure of the input file rete.inp.*

3.8 Program NETAN_HD

The principal routines defined in NETAN_HD are listed in table (3.4).

NETAN_HD	Main program
subroutine INPUT	Reading of the input file
subroutine ELAB	Imposition of the boundary conditions
subroutine HDRIV	Head driven analysis
subroutine TODINI	Demand driven analysis
subroutine STIFF	Calculation of the system matrix coefficients
subroutine CW	Friction factor for turbolent flow regime

Table 3.4: *NETAN_HD program structure.*

3.8.1 INPUT

The subroutine INPUT reads the input file *rete.inp*. The main steps are:

1. Reading of the general network data.
2. Reading of the node coordinates.
3. Reading of the node properties and nodal demands.
4. Reading of the pipe properties and distributed demands.
5. Reading of the pump number and hydraulic properties.
6. Calculation of the overall network discharge.

3.8.2 ELAB

The subroutine ELAB reorders the input data and imposes the necessary boundary conditions on fixed head nodes. The nodal withdrawals due to concentrated and distributed demands are calculated (Messina).

3.8.3 HDRIV

The subroutine HDRIV solves the network analysis problem through the three step procedure described in Todini (2003). The main steps are:

1. A conventional DDA is performed from the subroutine TODINI with fixed demands. If all the constraints are satisfied, the found solution is the correct one and the analysis stops.

2. If at least one node shows an insufficient head, a PDA is undergone.

3. Step 1: the minimum required head values are imposed on the critical nodes. If all the nodes besides the source nodes are critical, then the network is insufficient and the analysis stops.

4. Step 2: the subroutine TODINI computes the maximum demands $\widehat{q}_i$ compatible with the given constraints. Three possibilities descend from this solution:

   - $\widehat{q}_i \geq q_i$, the original constraint is set back in terms of demand, again equal to q_i.
   - $0 \leq \widehat{q}_i \leq q_i$, the head constraint is retained.
   - $\widehat{q}_i < 0$, the original constraint is set back in terms of demand but with demand $q_i = 0$.

5. Step 3: the second step is repeated until convergence.

3.8.4 TODINI

The subroutine TODINI performs the conventional DDA with fixed demands. The gradient algorithm implemented allows for the modelling of pressure regulating devices such as pumps or valves. The main steps are:

1. Imposition of the boundary conditions. If the heads at the extremes of a pipe are identical and thus the flow in the pipe vanishes, the imposed head value is lowered for one of the two nodes.

2. Building of the topological matrices.

3. Building of the system matrix calling the subroutines STIFF and CW.

4. Building of the nodal Newton-Raphson iterative scheme.

5. Solving of the coupled system through the ICF/MCG algorithm.

6. Once the solution has been found in terms of pressure heads and flow rates distribution, the corresponding flow velocity and head loss in pipes are computed.

7. The whole result is returned to the main program and written in the output file *risultati.out*.

3.8.5 STIFF

The subroutine STIFF calculates the coefficients of the system matrix according to the flow regime in pipes (laminar or turbolent) and using the Darcy-Weisbach formula for the description of the head losses. The main steps are:

1. Calculation of the Re number.

2. At the first iteration the flow is considered to be laminar.

3. Control and adjustment of the range of flow rates in which the pumps are working.

4. Calculation of the friction factor considering the type of flow (laminar or turbolent).

5. Calculation of the correction coefficients for the Messina method.

3.8.6 CW

The subroutine CW calculates the value of the friction factor to be used in the Darcy-Weisbach formula, assuming the Re number and the pipe roughness are known. The Colebrook-White relationship is used.

3.9 Output file echo.out

The output file *echo.out* summarizes the general network data. The properties of nodes, links and pumps are listed to verify they have been correctly read and processed. The value of the total flow rate delivered to the network analyzed is reported at the end.

3.10 Output file risultati.out

The output file *risultati.out* contains the results of the analysis. The required number of iterations is reported for each step of the method. For each node are reported the calculated nodal flow rate, the piezometric head and pressure head. For each link are reported the flow rate and the head loss. The principal variables are listed in table (3.5).

Type	Variable	Description	Unit
	nID	Name of the network node	
R	QUC	Calculated nodal flow rate	ls^{-1}
R	HN	Calculated piezometric head	m
R	PP	Calculated pressure head	m
	pID	Name of the link element (pipe or pump)	
	N1ID	Name of the starting node	
	N2ID	Name of the ending node	
R	QQ	Calculated pipe discharge	ls^{-1}
R	VV	Calculated flow velocity	ms^{-1}
R	DDH	Calculated hydraulic head loss	m

Variable types: I = integer, R = real, T = text.

Table 3.5: *Structure of the output file risultati.out.*

Two example networks are reported in Appendix A to prove the effectiveness of the implemented pressure driven hydraulic network model.

Chapter 4

EnKF vs. ES for losses identification in WDSs

4.0.1 Introduction

This study proposes a method for the identification of the spatial distribution of water losses in water distribution networks (WDNs) through the use of pressure head measurements (Pudar and Liggett, 1992; Chen and Zhang, 2006). The proper identification of areas most prone to water losses reduces the costs associated with acoustic surveys both in terms of number of pipes to be examined and working time.

To get the best estimate of the water losses spatial distribution, data assimilation techniques based on the Kalman Filter approach, the Ensemble Kalman Filter (EnKF) (Evensen, 1994) and the Ensemble Smoother (ES) (Van Leeuwen and Evensen, 1996) are coupled with the hydraulic network model (EpaNET) (Rossman et al., 2000). The coupled model performances are investigated on the Anytown benchmark system (Walski et al., 1987) with both a known and unknown consumption pattern. Water demand and pipe roughness are assumed as known. A method to identify the most effective network monitoring locations is also proposed, based on a first order approximation analysis of the uncertain parameters (Bush and Uber, 1998; Xu and Goulter, 1998). Despite the fact that the method is tested on a single synthetic network, the result suggests that the tool is promising for water losses identification.

4.1 The coupled inverse model

Two Kalman filter based data assimilation techniques, the EnKF and the ES, are applied to infer the spatial distribution of water losses L through a set of pressure head measurements

Z^m. The transient model EpaNET is used to retrieve the pressure values H during an extended period simulation characterized by a varying nodal demand.

The EnKF si recursively applied on a Monte Carlo (MC) ensemble of system realizations. Each MC realization is characterized by a different set of model parameters L with assigned statistical properties and is propagated in time through EpaNET to compute the pressure distribution in the network. At each time t_i in which a set of pressure head measurements z^m is available, both L and H values are corrected (updated) by means of these measurements. This procedure is repeated till the last measurement collection time t_F and it returns a progressive correction of the parameters L. The equation that describes the EnKF technique is (2.13):

$$X_{t_i}^u = X_{t_i}^f + K_{t_i}\left(z_{t_i}^m - MX_{t_i}^f\right) \tag{4.1}$$

with $t_i = t_1, ..., t_F$, where

$$X_{t_i}^f = \left[L_1, L_2, ..., L_{nn}, H_{1(t_i)}, H_{2(t_i)}, ..., H_{nn(t_i)}\right] \tag{4.2}$$

is the forecast model state estimate (equation (2.11)) with nn the number of network nodes;

$$z_{t_i}^m = \left[z_{1(t_i)}^m, z_{2(t_i)}^m, ..., z_{nmis(t_i)}^m\right] \tag{4.3}$$

is the measurements vector (equation (2.12)), with $nmis$ the number of collected measurements at each time t_i; K is the Kalman gain matrix that operates on the difference between the measured pressure heads $z_{t_i}^m$ and the corresponding values predicted by the model $MX_{t_i}^f$, being M the matrix operator that maps the EpaNET results to the measurements. In particular, the Kalman gain (equation (2.4)) is equal to

$$K = P^f M^T \left(MP^f M^T + R\right)^{-1} \tag{4.4}$$

where P^f is the model forecast error covariance matrix (equation (2.10)) and R the measurement error covariance matrix.

The model state forecast $X_{t_i}^f$ is updated through equation (4.1) to produce the posterior estimate

$$X_{t_i}^u = \left[L_1^u, L_2^u, ..., L_{nn}^u, H_{1(t_i)}^u, H_{2(t_i)}^u, ..., H_{nn(t_i)}^u\right] \tag{4.5}$$

EpaNET is then applied with the updated values of $L = [L_1^u, L_2^u, ..., L_{nn}^u]$ and the new model state forecast $X_{t_i+1}^f$ at the next measurement time t_{i+1} is computed.

The ES differs from the EnKF because all the measurements are processed in one step and the model state $X_{t_F}^{f*}$ that is now equal to

$$X_{t_F}^{f*} = [L, H_{t_1}, ..., H_{t_F}] \tag{4.6}$$

is updated only once at the last measurement time t_F. The matrix $X_{t_F}^{f*}$ in equation (4.6) has then dimension $X_{t_F}^{f*}[nn, (1+t_F)]$. EpaNET is run just once till time t_F and the H values computed at the measurement times $(t_1, t_2, ..., t_F)$ are collected in $X_{t_F}^{f*}$ and updated all together by means of the matrix Z^{m*} holding the perturbed measurements available:

$$Z^{m*} = \left[Z_{t_1}^m, Z_{t_2}^m, ..., Z_{t_F}^m\right] \tag{4.7}$$

The matrix $X_{t_F}^{u*}$ containing the updated model states for all the measurement times is obtained by applying equation (4.1) just once with the new meaning of the matrices.

It has to be stresses that the Kalman techniques are here applied to estimate water losses, included as model parameters in the state matrix to solve an inverse problem. Both the techniques adopted (EnKF and ES) could be able to correct such values depending on the correlation between losses and the measured hydraulic heads (Van Leeuwen and Evensen, 1996).

4.2 The Anytown network

The EnKF and the ES are coupled with the EpaNET software, to investigate their applicability in leakage detection for the Anytown WDN. It represents a benchmark system for the supply of an hypothetical community of about 350000 inhabitants and was originally conceived to compare results obtained by different optimization methods (Walski et al., 1987). The network, a 20 nodes and 39 pipes system, is shown in figure (5.1) according to the pipe configuration obtained from Farmani et al. (2005). The node elevation and the mean nodal demand of the example here analyzed are given in table (B.1), while the pipe data are reported in table (B.2). Tanks are neglected compared to the literature case. Water is pumped into the system from a water treatment works by means of three identical pumps connected in parallel, whose characteristic curve is given in table (B.3). The link and node data, the pump characteristic curve, the average daily water use at each node and the variation of water use throughout the day are available from the Centre for Water Systems (2004).

4.3 Model set up

The capabilities of the EnKF and of the ES to retrieve the water losses spatial distribution through the assimilation of pressure head measurements are investigated, and the performance of the two approaches is compared. Water losses are assumed as the only uncertain

model parameters and are supposed to be about 50% of the overall system consumption that is 620 ls^{-1} on daily average. As the water demands, also the water losses are modelled as nodal withdrawals, although they are distributed along pipe lines.

To prove the capabilities of the model, the synthetic reference system shown in figure (4.1) is considered, where the daily average of the losses is reported. In such system, the overall water loss is distributed on the subset of arbitrarily chosen network nodes (nodes 2, 4, 9 and 10). To provide the measurements used in the assimilation procedure, an extended period simulation of 24 hours is run, according to the known spatial and temporal evolution of water demand. Through the EpaNET software, the pressure values H on each network node are computed and recorded every 3 hours, that is for 8 time instants in the 24 hours. These values provide the measurements used in the assimilation procedure, and are affected by an uncertainty that is defined through a variation coefficient set to 0.01 on the basis of physical considerations.

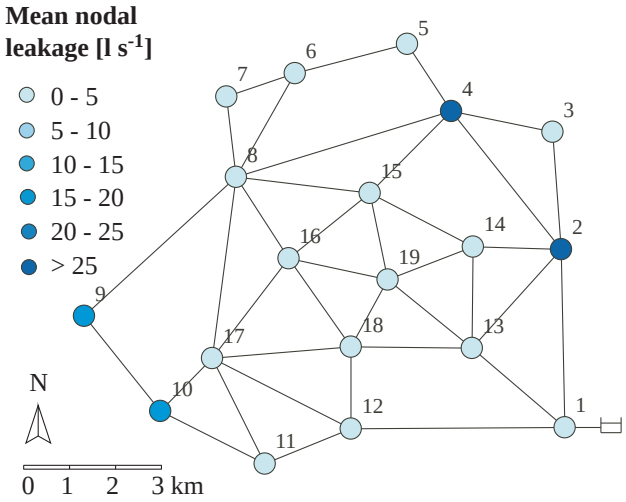

Figure 4.1: *Water losses spatial distribution in the Anytown synthetic reference system.*

The tests are carried out in two different cases. In the first one the temporal evolution of water demand is known: both the EnKF and the ES are applied considering the system state evolution according to the time pattern. In the second case, the temporal evolution of water demand is unknown as usually encountered in real-life problems and for this case just the ES is applied considering a steady state behaviour of the system driven by mean daily water demand, the EnKF application being misleading in the second case. Three different scenarios are analyzed:

1. EnKF application for known consumption pattern;

2. ES application for known consumption pattern;

3. ES application for unknown consumption pattern (the only mean daily values are available).

The model performances are investigated for each scenario with an increasing number of assimilated pressure head measurements: 10, 15 or 19.

In distribution systems a proper sampling design, that is a proper identification of the measurement locations, lets the best possible model calibration. The sampling procedure adopted here selects the network nodes where pressure values are most sensitive to changes in the calibration parameters, that is to changes in the spatial distribution of water losses. By the MC approach, an initial distribution of water losses is generated for each MC simulation, being 500 the proper size of the ensemble deduced from a preliminary sensitivity analysis. The leakage values are sampled from the uniform distribution and properly weighted to respect the known total amount of water losses in the system. A different set of water losses characterized any specific EpaNET simulation, by which the system is propagated forward according to the governing equations. For both the two cases of known or unknown demand time pattern, the cross correlation between the sampled L parameters and the corresponding H variables is calculated. Nodes are ranked according to the mean cross correlation absolute values as shown in figure (4.2). Numerical simulations are developed progressively increasing the number of assimilated H measurements. A proper methodology, which considers the cross correlation relationship between parameters and variables of the model, is suggested to select the H measurement points within the network that are most affected from a change in the spatial distribution of water losses. The result of the model provides the updated average L parameters. In details, the L spatial distribution is obtained and the corresponding H field is computed.

4.4 Numerical experiments

4.4.1 Assimilation of 10 measurements

The results obtained from the assimilation of 10 pressure head measurements are shown in figure (4.3), panel (1a), (1b) and (1c). Different colours correspond to different nodal leakage values L and the network positions that have to be surveyed are those where the L values are higher. The comparison with figure (4.1) shows that the solution nodes characterized by higher cross correlation absolute values are more easily recognized as leakage positions than the nodes characterized by lower cross correlation values. The EnKF (panel (1a)) identifies half of the solution nodes (nodes 2 and 10) while other two nodes (4 and 9) are completely

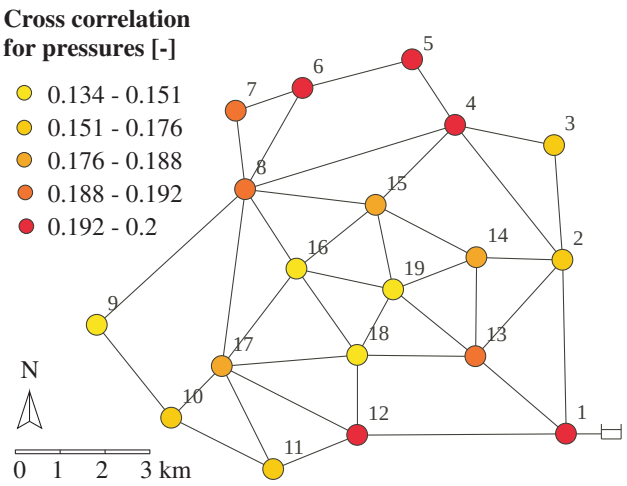

Figure 4.2: *Rank of the nodes for Anytown according to the mean cross correlation absolute value criterion.*

missed. The ES (panel (1b) and (1c)) better defines a network area close to nodes 2 and 4 affected by water losses, while the area surrounding nodes 9 and 10 is not recognized by the model. The former area is better defined from the ES in case the demand pattern is known (panel (1b)) rather than in case of unknown demand pattern (panel (1c)).

4.4.2 Assimilation of 15 measurements

The comparison with panels (1a), (1b) and (1c) of figure (4.3) shows that the uncertainty of the solution is reduced when 15 H values are assimilated (panels (2a), (2b) and (2c)). Moreover, the performance of the EnKF (panel (2a)) is completely overcome by the ES one (panels (2b) and (2c)). The EnKF is not able to well localize the leakages, whereas the ES suggests well defined areas around nodes 2 and 4, and nodes 9 and 10, where water losses are most likely to occur. When the consumption time pattern is available the ES stresses the differences between areas characterized by lower or higher L values, and gives the highest peak value at node 4.

4.4.3 Assimilation of 19 measurements

In this case the EnKF (panel (3a)) gives its best performance, accurately identifying the solution nodes and reducing almost to zero the L values on the remaining network nodes. Also the ES (panels (3b) and (3c)) demonstrates the highest efficiency, displaying higher peaks in correspondence to the solution nodes, with no relevant difference depending on the knowledge of the demand pattern. Anyway, with 19 assimilated measurements the EnKF prevails.

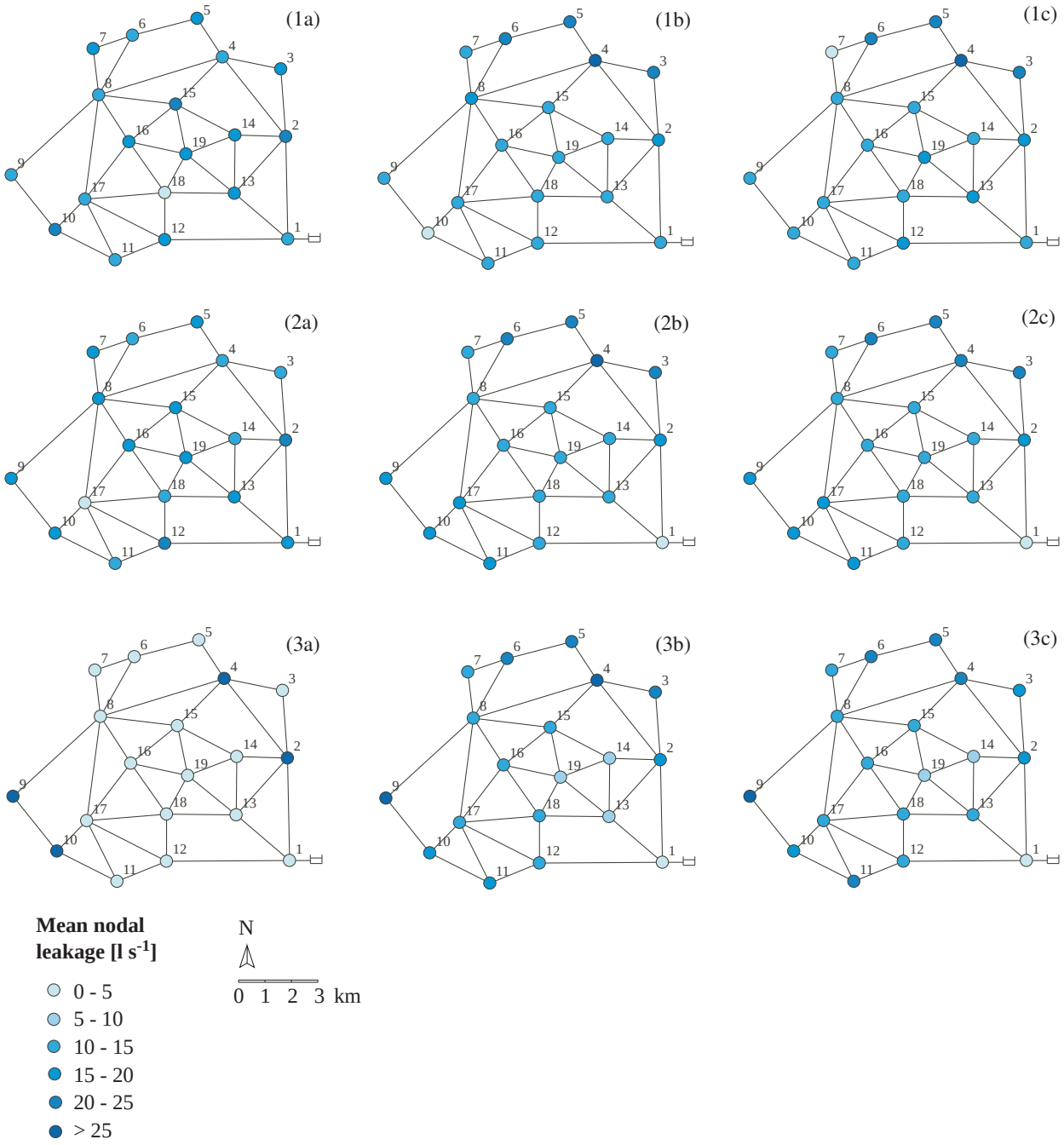

Figure 4.3: *Results of the analysis for the assimilation of 10 measurements: EnKF (1a), ES for known consumtpion pattern (1b) and ES for unknown consumption pattern (1c). Results of the analysis for the assimilation of 15 measurements: EnKF (2a), ES for known consumtpion pattern (2b) and ES for unknown consumption pattern (2c). Results of the analysis for the assimilation of 19 measurements: EnKF (3a), ES for known consumtpion pattern (3b) and ES for unknown consumption pattern (3c).*

4.4.4 Discussion on the results

As expected, both the EnKF and the ES give better performances as the number of assimilated pressure head measurements increases.

In the Anytown system here analyzed, both the EnKF and ES techniques suffer the undifferentiated and small cross correlation values affecting the nodes. The simple arrangement of pipes in Anytown often causes a high number of pipes (compared to the number of the whole system) to join together on the same node. This aspect, together with the high carrying capacity of each pipe, implies that the pressure drop occurring on a node potentially characterized by leakage is easily recovered through a small decrease of the pressure head values that affects most of the network nodes. This justifies the small values of the cross correlation between the L parameters and the H variables, and also the small differences among the cross correlation in different nodes. Both these aspects make difficult the node discrimination. For these reasons, the model result does not rapidly improve with the increase of the assimilated measurement number.

The peculiarity of the network analyzed leads to some unexpected findings.

Even if more than 50% of the network nodes are monitored, the results are not satisfactory in terms of water losses identification. This is due to the topological features of the Anytown network which is characterized by a high resilience (Todini, 2000), thus requiring large efforts in terms of pressure measurements collection to further improve the results in terms of pipeflow distribution. It also implies that, with a limited number of nodes, a threshold number of assimilated measurements above which the model parameters could be considered satisfactorily calibrated cannot be defined in this case.

Moreover, when a limited number of measurements is used in the assimilation procedure, the ES works better than the EnKF. This is probably due to the filter inbreeding problem, that is, the ensemble variance increasingly underestimation over time (Hendricks Franssen and Kinzelbach, 2008), that deeply affects the EnKF assimilation procedure just after few recursive steps since the initial covariance values are small. In this case and in the particular case of the Anytown network, the ES approach seems to give better results.

In the Anytown network, the knowledge of the demand pattern does not lead to any improvement in the delineation of areas prone to leakage, but it increases the results meaning enhancing the peak values of L for both the EnKF and the ES.

Nevertheless, by strongly increasing the number of assimilated measurements, the EnKF outperforms the ES and leads to a right identification of the leakage nodes, as expected. The outperformance of the EnKF depends on the recursive structure of the updated procedure, which allows an effective management of the non-linearity (Crestani et al., 2013) that

characterizes this problem.

4.5 Final remarks

In this work, the capabilities of the EnKF and the ES to retrieve the spatial distribution of water losses in a distribution network through the assimilation of pressure head measurements are analyzed and compared.

The comparison is made in the case of the Antown supply system, a benchmark network characterized by a high resilience, that is, by low values of cross correlation between the uncertain L parameters and the H variables.

Both the EnKF and the ES give better performances as the number of assimilated pressure head measurements increases but, when this number is higher, the EnKF generally outperforms the ES. This is due to the fact that the ES is not recursive, and thus is not capable to manage the non-linearity of the problem.

As the number of measurements reduces, just in one case the results are better for the ES than for the EnKF, probably because of the peculiarity of the network analyzed, where on each nodal pressure H the influence of the flow rates L is limited, and because of the filter inbreeding problem that affects the recursive procedure. For this reason, the ratio between the minimum number of measurements that is needed to obtain a suitable result and the number of the network nodes cannot be defined as it was done in other cases analyzed (Ruzza et al., 2014).

Despite the fact that the network peculiarity may affect the number of measurements needed to reach a satisfactory result, the proposed approach demonstrates to be a promising tool for the calibration of the water losses parameters in a distribution system affected by uncertainty in the flow rate distribution.

It has to be stressed that in some cases a significant advantage can be derived from the application of the ES technique, that is when the system behaviour over time is not sufficiently described from the available information or, when the system knowledge is given only in terms of time averaged measurement data as it happens in Ruzza et al. (2014). In these cases the application of the ES technique is preferred, the EnKF being misleading when no temporal information is available.

Chapter 5

Losses identification in WDSs through NS-EnKF coupled with a pressure driven model

5.0.1 Introduction

Leakage in pressurized water distribution systems (WDSs) is a major issue for water utilities today, because of the huge concern over public health risk and the economic constraints on energy and resources.

As previous works (Ruzza et al., 2014; Ruzza et al., 2015; Ruzza and Salandin, 2015) suggested, data assimilation techniques based on the Kalman Filter coupled with the network hydraulic model EpaNET allow for calibration of model parameters (nodal leakage flow rates) through the assimilation of a suitable number of measurements. The effectiveness of this technique seems to be influenced not only from the hydraulic model features and from the parameters to be calibrated, but also from the specific WDS considered.

In this work, a pressure driven hydraulic model, which properly describes the physical relationship between the available nodal pressure and the system outflows (user consumption as well as leakage outflows), is coupled to the Ensemble Kalman Filter (EnKF) technique and Normal-Score Transform (NST). This approach works as an inverse model to calibrate the emitter coefficients (model parameters) that control the nodal leakage flow rates. Their physical meaning is straightforward, the emitter coefficient value being proportional to the area of the hole responsible for the leakage. The measurements assimilated for the model parameters estimation are pressure heads and flow rates, besides the values of cumulative volume of system inflow as well as of leakage outflow, over an extended period of time. These measurements, with exception of the flow rate measurements, could be easily collected in any

network with a limited effort and no technical troubles. Water demand and pipe roughness are assumed as known.

The capabilities of the NS-EnKF to retrieve the spatial distribution of leakages are analyzed and compared on the two synthetic benchmark systems of Anytown and Net3. The choice of the example networks is justified by considerations on their topological and hydraulic properties, besides their physical characteristics. In fact, the similarity of circulating flow rates and pipe diameters allow for comparison of the effects due to a different topological scheme, which is more complex in Net3 than in Anytown. Moreover, the elevation of the network nodes influences the system behaviour and differently affects the performance of the applied technique, Anytown being much steeper than Net3. The results suggest that the investigated tools are differently effective in helping to identify the leakage positions depending on the WDS physical characteristics.

5.1 The case studies

5.1.1 The Anytown network

The Anytown network is a benchmark system for the supply of an hypothetical community of about 350000 inhabitants and was originally conceived to compare results obtained by different optimization methods (Walski et al., 1987). The network, a 20 nodes and 39 pipes system, is shown in figure (5.1), according to the pipe configuration obtained from Farmani et al. (2005). The node elevation and the mean nodal demand of the example here analyzed are given in table (B.1), while the pipe data are reported in table (B.2). The network shows a significant redundancy in terms of pipe diameters and loops. Tanks are neglected compared to the literature case. Water is pumped into the system from a water treatment works by means of three identical pumps connected in parallel, whose characteristic curve is given in table (B.3). The water use pattern is available over a 24 hour time period with a timestep of 3 hours. The link and node data, the pump characteristic curve, the average daily water use at each node and the variation of water use throughout the day are available from the Centre for Water Systems (2004).

The daily system intake due to user consumption is $47.83 \cdot 10^6\ l day^{-1}$, the average daily flow rate being 620 ls^{-1}. The mean pressure head is 41.53 m in absence of leakage.

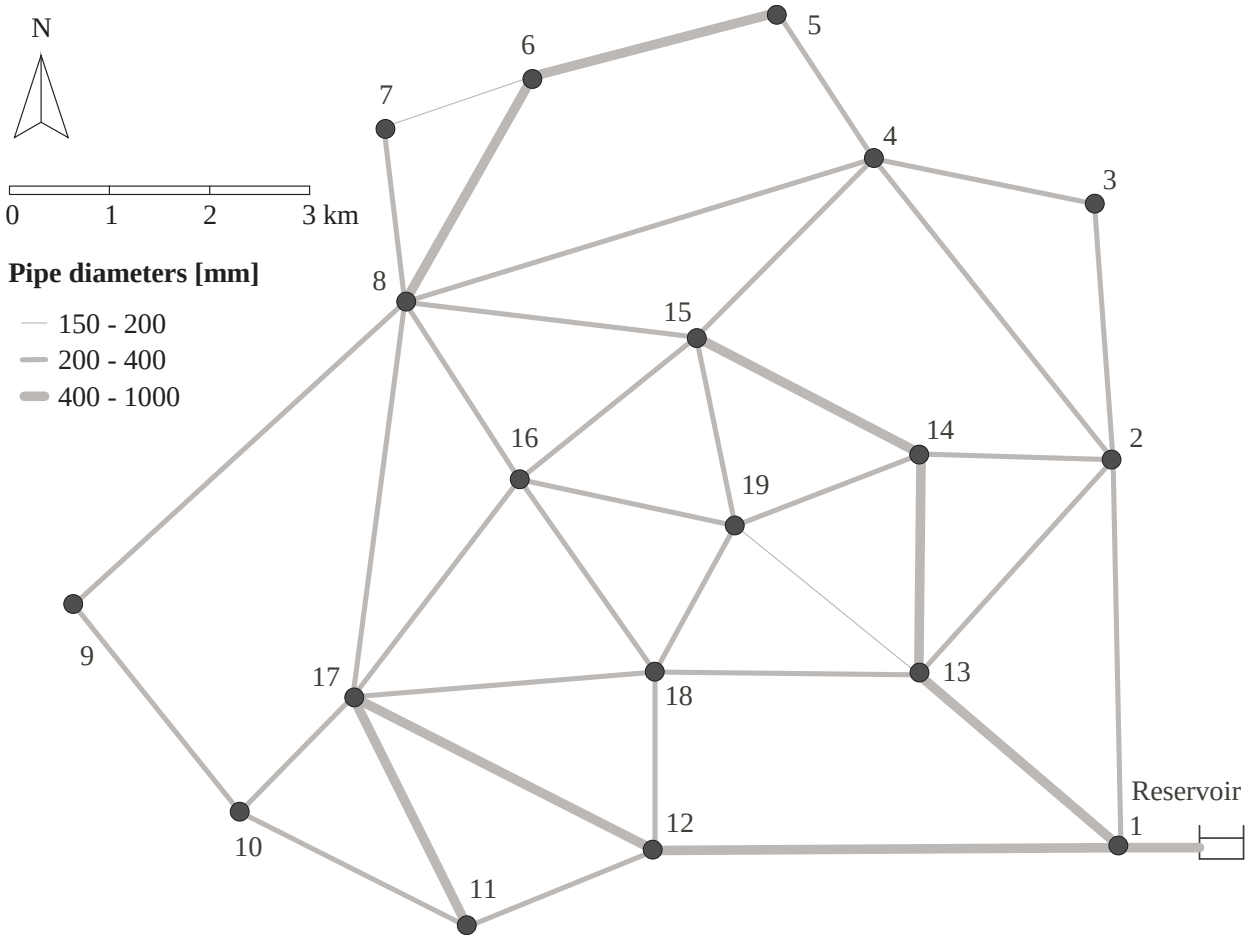

Figure 5.1: *Scheme of Anytown adopted in the developed example.*

5.1.2 The Net3 network

The Net3 network is a benchmark system coming with the EpaNET package. The network is made up by 89 nodes and 111 pipes and is shown in figure (5.2) in the configuration adopted for the developed example. The system is supplied from two sources: the river reservoir satisfies most of the demand, while the lake reservoir contributes for a much smaller part that is less than 5% when no leakage occurs. The node elevation and the mean nodal demand are listed in table (B.4), while table (B.5) shows the values related to pipes. The Net3 network has some branches which depart from a main pipe of large diameter extending throughout most of the system, while the portion of the system closer to the sources is mostly looped. Storage tanks are neglected with respect to the initial configuration. Water is pumped into the system from two water treatment works, both of them with two identical pumps connected in parallel. The pump characteristic curves are given in table (B.6) for the lake source and in table (B.7) for the river source. The water use pattern for Net3 is available over a 24 hour time period with a timestep of 1 hour.

The daily system intake due to user consumption is $59.68 \cdot 10^6\ l day^{-1}$, the average daily flow rate being 690 ls^{-1}. The mean pressure head is 78.45 m in absence of leakage.

5.2 The pressure driven hydraulic network model

Water losses, as well as user consumption, are considered as uniformly distributed along pipelines and modelled as nodal withdrawals. A pressure driven hydraulic network model is used to properly describe the true physical relationship between the available nodal pressure and the system outflows (user consumption as well as leakage outflows).

To overcome the limitations presented by the intrinsically demand driven nature of standard EpaNET, the nodal demands and the water losses are simulated through a nontrivial combination of valves and hydrants, using EpaNET emitters. These are devices modelling the flow through a nozzle or orifice that discharges to the atmosphere. The flow rate through an emitter varies as a function of the pressure available at the node according to a power law (equation (5.1)):

$$q = CH^{\gamma} \tag{5.1}$$

where q is the flow rate, H is the pressure head and γ is a pressure exponent (Rossman et al., 2000). Because the standard EpaNET version only allows for a single value of pressure exponent to be specified, $\gamma = 1$ is used for all the nodal demands and water losses. Figure (5.3) represents the qualitative behaviour of the flow rate-pressure relationship adopted. A similar approach is used by Salandin and Bertola (1996) based on the hypothesis of Gupta

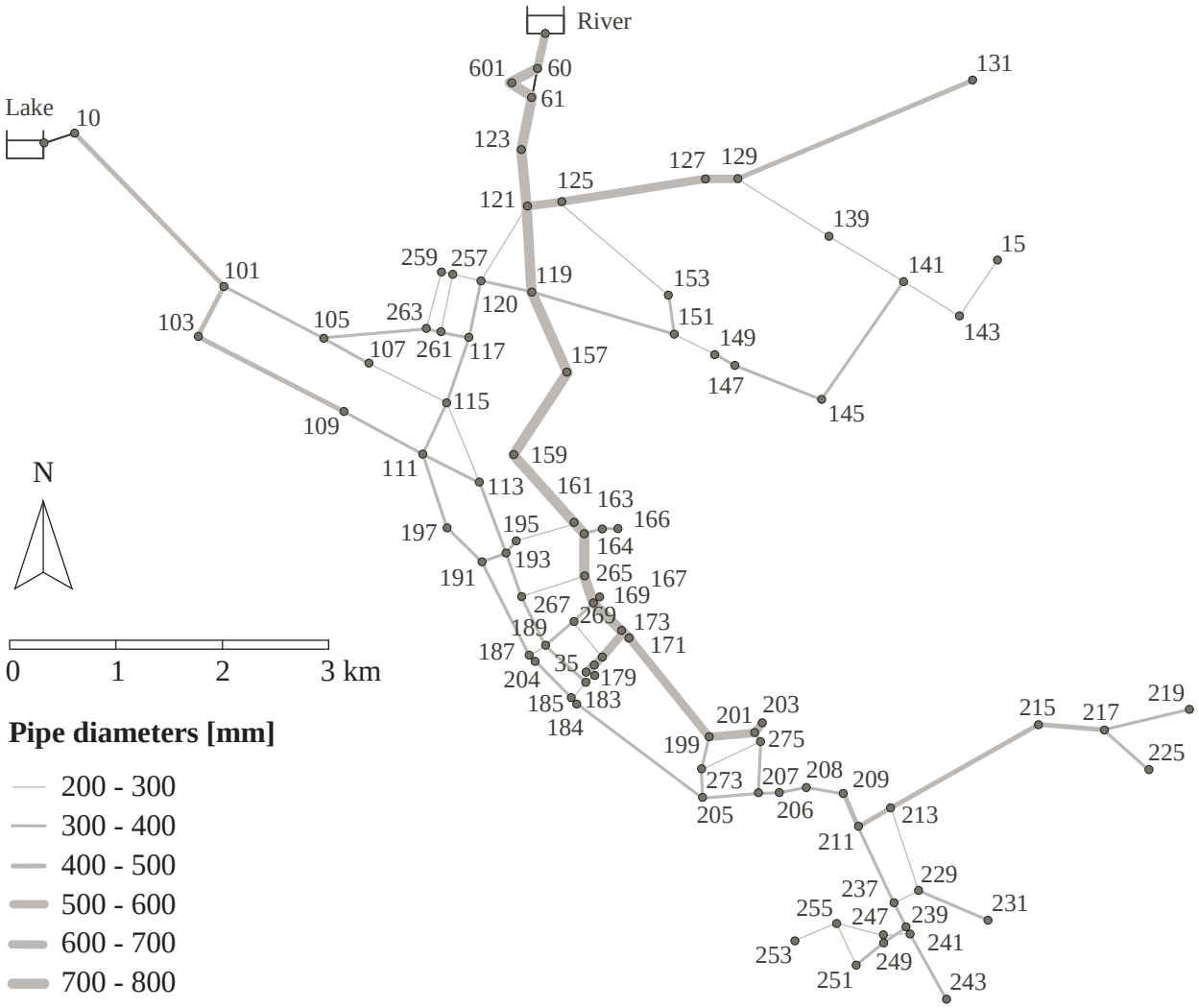

Figure 5.2: *Scheme of Net3 adopted in the developed example.*

and Bhave (1996). Brunone (2007) observes that the relationship between the leakage flow rate and the available pressure tends to be linear due to the variability of the leak areas according to the pressure that is realized. The main steps for building the pressure driven model can be summarized as follows.

- The base demand for all the network nodes is set to zero.
- For each network node three pipes are added, with very short length and very large diameter so as not to cause any significant head loss. These are:

    1. A check valve (CV) that links the network node to a dummy node, where the initial base demand is replaced. The dummy node has the same elevation as the network node. In the same way, the available pattern for water consumption is removed from the network node and assigned to the dummy node. The resulting pressure head or piezometric head can be read on both these nodes, whereas the pressure driven actual demand coincides with the CV flow rate.
    2. A CV that links an emitter node to the dummy node. The elevation of the emitter node exceeds the elevation of the network node of an amount equal to the *service pressure*, that is the pressure head above which the demand is fully satisfied and the node behaviour is demand driven (under the hypothesis that the user does not consume more water than required). The *service pressure* is considered to be 25 m for the two networks analyzed and the outflows behave as in figure (5.3). The emitter coefficient value for the pressure driven demand is calculated as in equation (5.1), considering the *service pressure* of 25 m and the maximum demand that would occur at the node during a standard demand driven extended period simulation. The choice of the maximum demand derives from the need of ensuring that a demand driven analysis is performed when no leakage occurs, thus the model results can be compared with the standard EpaNET output. In the pressure driven scheme, the emitter node acts as a reservoir supplying the dummy node as long as the pressure is below 25 m, decreasing the system outflow when pressure is too low.
    3. A CV that links the network node to an emitter node representing the leakage. The CV prevents the leakage flow rate to become negative when the available nodal pressure is lower than the node elevation. These emitter coefficient values are uncertain model parameters to be calibrated.

The pressure driven scheme proposed does not require the emitter coefficients to be changed

during an extended period simulation, but it is anyhow able to represent the temporal variability of water use thanks to the varying demand imposed on dummy nodes. Such an approach can be easily implemented using the standard EpaNET input file.

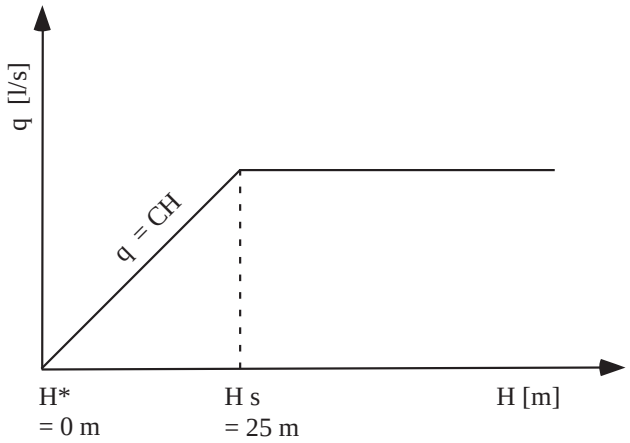

Figure 5.3: *Qualitative behaviour of the flow rate-pressure relationship according to the pressure driven model adopted.*

5.3 Model set up

5.3.1 Synthetic system solution

The work deals with the effectiveness of the EnKF to provide the best possible estimate of the nodal leakage locations when coupled with the proposed pressure driven hydraulic model. The emitter coefficients $C = [C_1, C_2, ..., C_{nn}]$, each of dimension $[ls^{-1}m^{-1}]$, control the nodal leakage flow rates $L = [L_1, L_2, ..., L_{nn}]$, each of dimension $[ls^{-1}]$, where nn is the number of network nodes. The C values are assumed as the only uncertain model parameters, kept constant over time. Their physical meaning is straightforward, the C value being proportional to the area of the hole responsible for the leakage.

The model performances are evaluated through comparison with a synthetic reference system for the two networks analyzed. The synthetic system solution is reported in figure (5.4) for Anytown and in figure (5.5) for Net3, in terms of leakage diameter and in terms of corresponding mean nodal leakage flow rates $\bar{L} = \left[\bar{L}_1, \bar{L}_2, ..., \bar{L}_{nn}\right]$ over a 24 hour simulation period. In such systems, the overall water loss is distributed on a subset of arbitrarily chosen network nodes. These solution systems provide the measurements used in the assimilation procedure.

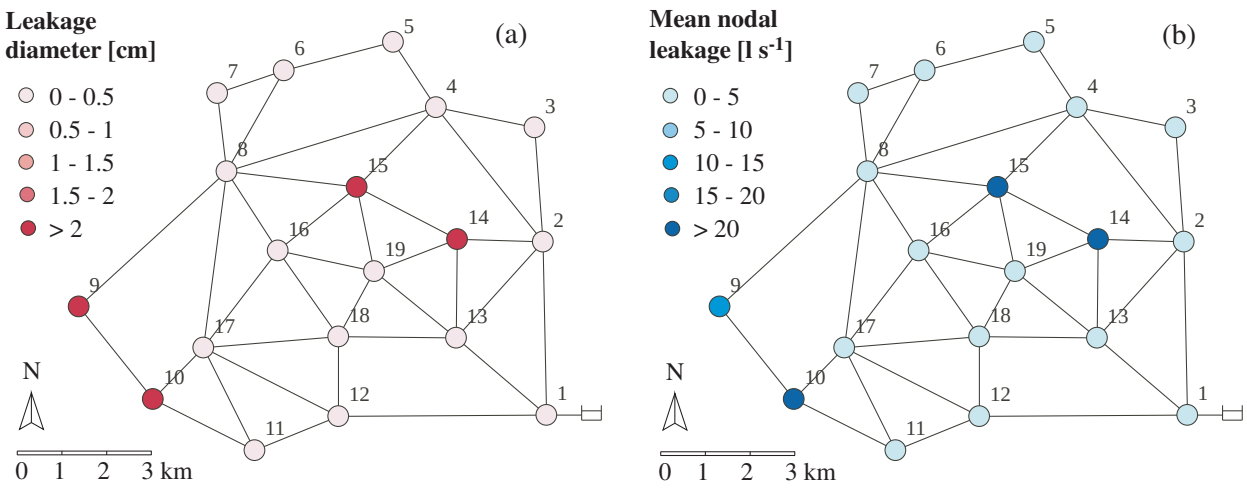

Figure 5.4: *Synthetic solution for Anytown in terms of leakage diameter (panel a) and mean nodal leakage flow rate (panel b).*

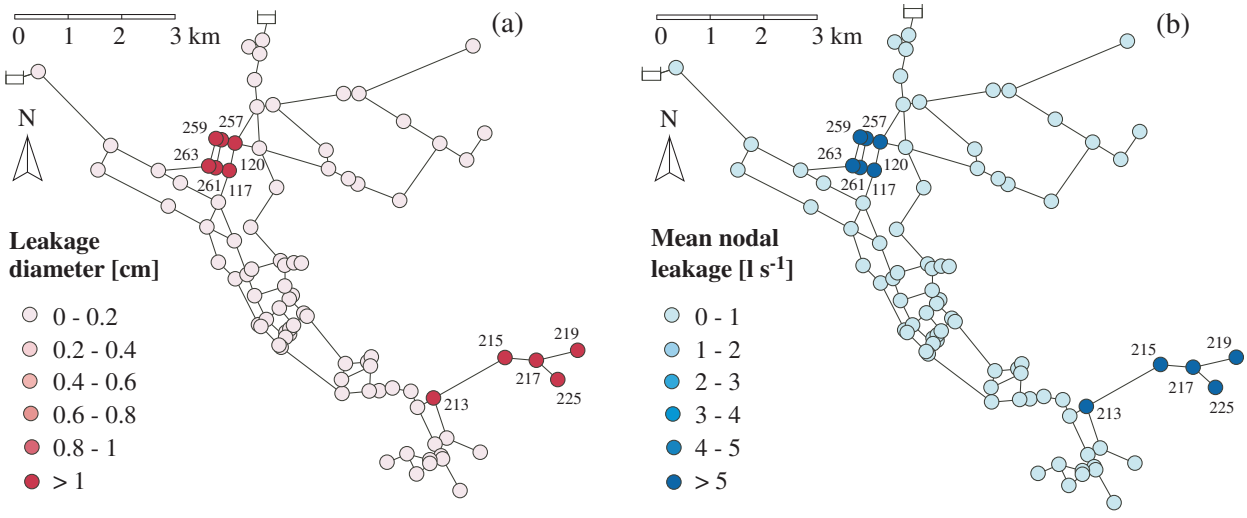

Figure 5.5: *Synthetic solution for Net3 in terms of leakage diameter (panel a) and mean nodal leakage flow rate (panel b).*

5.3.2 Generation of the ensemble

The C field is assumed to follow a bimodal probability density function (pdf) on the basis of physical considerations (figure (5.6)). The bimodal shape gives reason of the fact that only a relative portion of the WDS is affected by losses, that is $C_i > 0$, while most of the network nodes are characterized by leakage of negligible entity, that is $C_i \sim 0$. A different weighting of the pdf masses represents a different percentage of leakage nodes in the WDS.

At time t_0 an initial ensemble of NMC realizations of the random function C is generated. The first mode m_1 of the bimodal pdf is a necessarily positive value close to zero, while the second mode m_2 is derived assuming that the overall leakage volume $(V^m_{out})_{t=24}$, measured over a 24 hour time period, is distributed on a percentage of the nodes. Equation (5.2) considers the mean nodal leakage flow rate $\bar{L}_i$ $[ls^{-1}]$ and a mean pressure $\bar{H}$ $[m]$ that is likely to occur in the leaking system.

$$m_2 = \frac{\bar{L}_i}{\bar{H}} \tag{5.2}$$

The calculated values are $m_2 = 0.48\ ls^{-1}m^{-1}$ for Anytown and $m_2 = 0.39\ ls^{-1}m^{-1}$ for Net3, assuming in first instance that half of the WDS is characterized by leakage and that $\bar{H} = 25$ m is likely to occur.

Each realization of the C field is propagated forward in time through the pressure driven hydraulic model, according to the system governing equations and the known spatial and temporal variability of water demand.

The estimation of the L field is realized by measurements assimilation and using an augmented system in which only the model parameters C are considered (Bailey and Baù, 2010). Measurements are available over discrete time instants $t_i = t_1, t_2, ..., t_F$ $[hours]$ being F the number of observation times. Three types of measurements are used in the assimilation procedure:

- the pressure heads $H^m = [H^m_1, H^m_2, ..., H^m_{nobsH}]$ $[m]$;
- the pipe flow rates $Q^m = \left[Q^m_1, Q^m_2, ..., Q^m_{nobsQ}\right]$ $[ls^{-1}]$;
- the cumulative volume of system inflow V^m_{in} $[l]$ and leakage outflow V^m_{out} $[l]$ over a known time period.

The superscript m indicates the measured values, $nobsH$ and $nobsQ$ are the number of observed variables at each observation time t_i. The state vector is made up of model parameters C and variables: hydraulic heads $H = [H_1, H_2, ..., H_{nn}]$, flow rates $Q = [Q_1, Q_2, ..., Q_{np}]$, the total inflow volume V_{in} and the total leakage outflow V_{out} over a known time period. The

state matrix containing the fixed parameters and the computed variables at each required time instant is built as in equation (5.3)

$$X_{t_i} = [C_1, ..., C_{nn}, H_1, ..., H_{nn}, Q_1, ..., Q_{np}, V_{in}, V_{out}]_{t_i} \tag{5.3}$$

and has dimensions $X\left[(nn + nn + np + 2), NMC\right]$, being nn the number of nodes and np the number of pipes. The observation vector containing the available measurements is built as in equation (5.4)

$$z^m_{t_i} = \left[H^m_1, ..., H^m_{nobsH}, Q^m_1, ..., Q^m_{nobsQ}, V^m_{in}, V^m_{out}\right]_{t_i} \tag{5.4}$$

and has dimension $z^m\left[nobsH + nobsQ + 2\right]$, being the volume values always available as their collection is part of the normal management operations of a WDS.

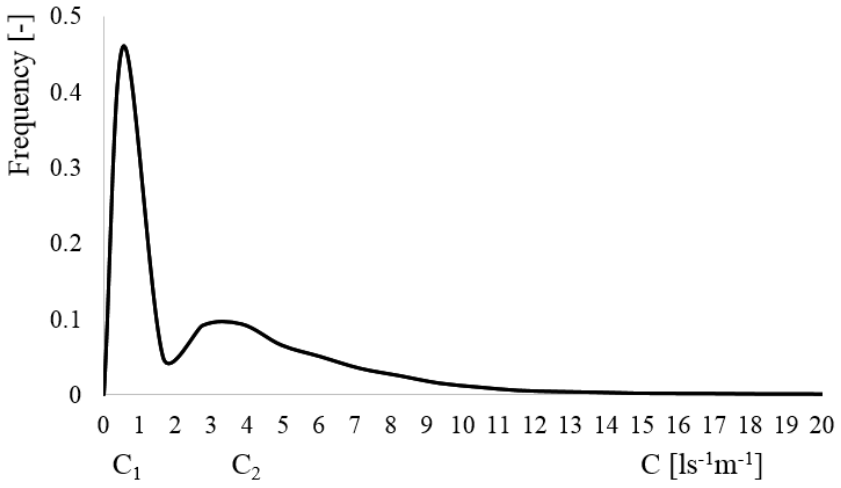

Figure 5.6: *Qualitative behaviour of the bimodal frequency distribution.*

5.3.3 Sampling design

Sampling design is a two-objective optimization problem. The first objective is the minimization of the difference between the model-simulated output and measured data. The second objective is the reduction of total sampling design cost. The effectiveness of the L field estimate is related to the cross correlation between L and the potential measurements H and Q. The sampling procedure here adopted selects the network nodes or the network pipes where the H values or Q values are most sensitive to changes in the spatial distribution of the nodal leakage flow rates L, that is in the field of variables to be estimated. The goal is achieved through a Monte Carlo numerical approach, based on the assumption that the only system uncertainty is related to the random function C, while the remaining parameters

(e.g. pipe roughness, nodal demands) are fixed as deterministic. The only unknown in the sampling design analysis is the location of pressure sensors and flowmeters.

A ensemble of NMC equally likely stochastic realizations of the state vector (equation (5.3)) is generated. Each realization of the C field, then of water losses L at different times, characterizes any specific EpaNET simulation, by which the system is propagated forward according to the governing equations. The mean values of the system variables $\bar{L}$, $\bar{H}$ and $\bar{Q}$ over a 24 hour time period simulation are retrieved. The cross correlation between $\bar{H}$ and $\bar{L}$ is calculated using equation (5.5)).

$$CrCorr\left(\bar{H},\bar{L}\right)=\frac{\left\langle\left(\bar{H}-\left\langle\bar{H}\right\rangle\right)\left(\bar{L}-\left\langle\bar{L}\right\rangle\right)^{T}\right\rangle}{\sqrt{Var\left(\bar{H}\right)Var\left(\bar{L}\right)}} \tag{5.5}$$

where

$$\left\langle\bar{H}\right\rangle=\frac{1}{NMC}\sum_{i=1}^{NMC}\bar{H} \tag{5.6}$$

$$Var\left(\bar{H}\right)=\frac{1}{NMC-1}\sum_{i=1}^{NMC}\left(\bar{H}-\left\langle\bar{H}\right\rangle\right)^{2} \tag{5.7}$$

The result is a cross correlation matrix of dimensions $CrCor\,(nn,nn)$, of which any element is itself a cross correlation that can assume values from -1 to $+1$, where 0 means the considered variables are totally uncorrelated, while 1 means the considered variables are perfectly correlated. The absolute values of cross correlation are used, as both positive and negative fluctuations are equally important. The network nodes are ranked according to the mean cross correlation absolute values. The result is a selection of positions where pressure head measurements are more effective for retrieving the spatial distribution of L. Following the same approach, the cross correlation between $\bar{Q}$ and $\bar{L}$ is calculated, obtaining a matrix of dimensions $CrCor\,(np,nn)$. The result is a rank of pipe positions where flow rate measurements are more convenient to retrieve the L field.

The results are reported in figure (5.7) for Anytown and in figure (5.8) for Net3, in terms of most effective potential measurement locations for pressure heads (panel a) and pipe flow rates (panel b). According to the dimension of the system, cross correlation takes values of order 10^{-1} for Anytown and 10^{-2} for Net3, meaning that the improvement due to the use of a one more measurement would be more effective in Anytown than in Net3. For both Anytown and Net3, it emerges a fairly clear correlation structure in terms of Q, while the distinction of the cross correlation values for H is very weak, with exception of the nodes close to supply points, where cross correlation is definitely low and measurements taken on such nodes would not be much helpful for calibration. More specifically, the cross correlation between $\bar{L}$ and $\bar{Q}$ reveals for Net3 a very strong dependence of the leakage estimation on the

knowledge of some Q values corresponding to the large diameter pipeline departing from the river reservoir. On the other hand a such strong dependence is not recognized in Anytown.

Because cross correlation between different system variables contributes to the system description, the calculated values of cross correlation between $\bar{H}$ and $\bar{L}$ ($CrCorr_{HL}$, column 2), $\bar{H}$ and $\bar{H}$ ($CrCorr_{HH}$, column 3), $\bar{H}$ and $\bar{Q}$ ($CrCorr_{HQ}$, column 4) are reported in table (5.1) for Anytown and in table (5.3) for Net3. The calculated values of cross correlation between $\bar{Q}$ and $\bar{L}$ ($CrCorr_{QL}$, column 2) are reported in table (5.2) for Anytown and in table (5.4) for Net3. $CrCorr_{HH}$ is very high for Net3, implying rather small differences among piezometric heads when generating the ensemble of system states. A similar argument concerns also $CrCorr_{HQ}$. While the $CrCorr_{HQ}$ values are similar in Anytown and Net3, it happens in Net3 that the $\bar{H}$ field is much more dependent on the $\bar{Q}$ field than on the $\bar{L}$ field. In other words, the use of one more pressure measurement in Net3 would be more effective in defining the Q field than the L field.

The use of the mean daily values $\bar{L}$, $\bar{H}$ and $\bar{Q}$ is justified by considerations developed in Darvini et al. (2008). The method demonstrates that the spatial variability of L and the time variability of L can be treated separately, thus considering the mean time values of variables $\bar{L}$, $\bar{H}$ and $\bar{Q}$ is equivalent to observe only the fluctuations due to a different spatial distribution of L. In any WDS, the flow rate Q_j supplied at a node can be expressed as a function f of the head H at the same node and at the adjacent nodes, and of the roughness e (through a head-loss formula)(equation (5.8)).

$$Q_j = f\left(H, e\right) \tag{5.8}$$

The same flow rate Q_j is varying as a function of the temporal behaviour of the user demand q and of the system leakage L, thus it is a function of both space x and time t. In equation (5.9) all the dependencies are located in a function $G = G\left(x, t\right)$.

$$f\left(H, e\right) + q + L = G\left(H, e, q, L\right) = 0 \tag{5.9}$$

By expanding in Taylor series the equation (5.9) around the mean positions $\bar{H}$, $\bar{e}$, $\bar{q}$, $\bar{L}$, and by limiting the expansion to the first-order, one obtains equation (5.10).

$$G\left(\bar{H}, \bar{e}, \bar{q}, \bar{L}\right) + \frac{\partial G}{\partial H}H' + \frac{\partial G}{\partial e}e' + \frac{\partial G}{\partial q}q' + \frac{\partial G}{\partial L}L' \simeq 0 \tag{5.10}$$

The pipe roughness e is a random variable whose spatial variation is unknown, while its time evolution is nearly deterministic. It is generally $e = \bar{e}\left(x\right) + e'\left(x\right)$, but the fluctuation term is here neglected, the roughness values being known in the WDSs analyzed. Regarding the piezometric heads it is $H = \bar{H}\left(x\right) + H'\left(x, t\right)$. Both the spatial and temporal variation derive

from $H = H(q, L)$. For water consumption q the uncertainty regards only the temporal variation, being $q = \bar{q} + q'(t)$, while it is generally possible to reconstruct the spatial distribution of q through meter records. Regarding the water losses it is $L = \bar{L}(x) + L'(x, t)$. The spatial variation remains the only uncertainty, the temporal variation depending on $L = L(H)$. Equation (5.10) becomes

$$\bar{G}(x) + \frac{\partial G}{\partial H}\bar{H}'(x) + \frac{\partial G}{\partial L}\bar{L}'(x) \simeq 0 \tag{5.11}$$

It follows from the assumptions that considering the mean time values of variables $\bar{L}$, $\bar{H}$ and $\bar{Q}$ let us to consider only the fluctuations of H and Q due to a different spatial distribution of L, all the quantities in equation (5.11) being dependent only on x.

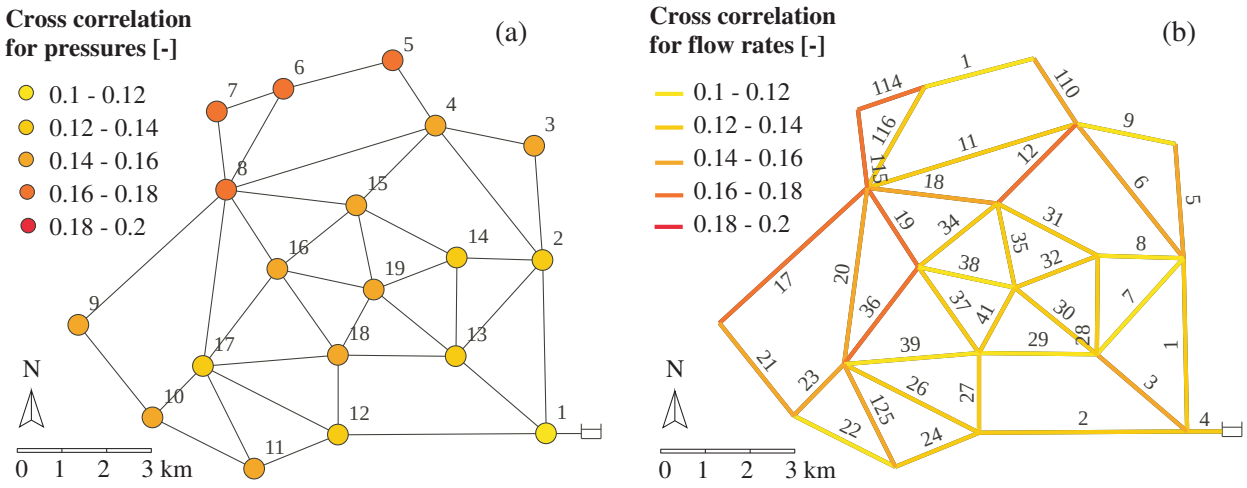

Figure 5.7: *Rank of the most effective potential measurement locations for Anytown, for pressure heads (panel a) and flow rates (panel b) according to the cross correlation criterion.*

Node	$CrCorr_{HL}[-]$	$CrCorr_{HH}[-]$	$CrCorr_{HQ}[-]$
1	0.110	0.703	0.352
2	0.135	0.693	0.350
3	0.146	0.630	0.239
4	0.159	0.669	0.219
5	0.160	0.640	0.194
6	0.161	0.633	0.189
7	0.165	0.710	0.228
8	0.166	0.719	0.233
9	0.146	0.657	0.250
10	0.143	0.693	0.282

Table continues on next page

Table continues from previous page

Node	$CrCorr_{HL}[-]$	$CrCorr_{HH}[-]$	$CrCorr_{HQ}[-]$
11	0.142	0.703	0.297
12	0.128	0.731	0.352
13	0.137	0.774	0.368
14	0.139	0.768	0.364
15	0.143	0.764	0.353
16	0.156	0.763	0.334
17	0.137	0.738	0.335
18	0.142	0.737	0.356
19	0.146	0.695	0.341
20	0.099	0.630	0.320

Table 5.1: *Values of cross correlation for Anytown, between piezometric heads and nodal leakage flow rates ($CrCorr_{HL}$, column 2), piezometric heads and piezometric heads ($CrCorr_{HH}$, column 3), piezometric heads and pipe flow rates ($CrCorr_{HQ}$, column 4).*

Pipe	$CrCorr_{QL}[-]$	**Pipe**	$CrCorr_{QL}[-]$
1	0.132	27	0.132
2	0.128	28	0.137
3	0.140	29	0.116
4	0.123	30	0.127
5	0.151	31	0.138
6	0.152	32	0.128
7	0.094	34	0.136
8	0.093	35	0.130
9	0.101	36	0.168
11	0.136	37	0.130
12	0.161	38	0.092
17	0.167	39	0.118
18	0.157	41	0.125

Table continues on next page

Table continues from previous page

Pipe	$CrCorr_{QL}[-]$	**Pipe**	$CrCorr_{QL}[-]$
19	0.179	110	0.155
20	0.158	113	0.100
21	0.145	114	0.169
22	0.106	115	0.172
23	0.143	116	0.137
24	0.135	125	0.141
26	0.134		

Table 5.2: *Values of cross correlation for Anytown, between pipe flow rates and nodal leakage flow rates ($CrCorr_{QL}$, column 2)*

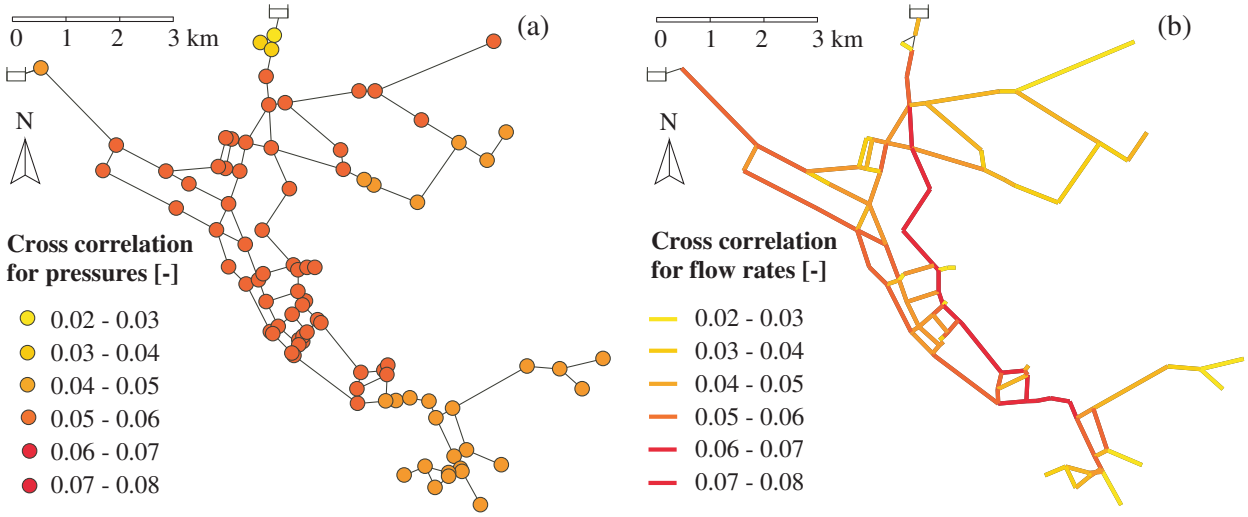

Figure 5.8: *Rank of the most effective potential measurement locations for Net3, for pressure heads (panel a) and flow rates (panel b) according to the cross correlation criterion.*

Node	$CrCorr_{HL}[-]$	$CrCorr_{HH}[-]$	$CrCorr_{HQ}[-]$
10	0.052	0.637	0.180
15	0.055	0.763	0.228
35	0.062	0.930	0.279
60	0.034	0.441	0.120

Table continues on next page

Table continues from previous page

Node	$CrCorr_{HL}[-]$	$CrCorr_{HH}[-]$	$CrCorr_{HQ}[-]$
601	0.038	0.530	0.135
61	0.038	0.530	0.135
101	0.064	0.904	0.263
103	0.064	0.902	0.263
105	0.064	0.910	0.266
107	0.063	0.905	0.265
109	0.064	0.907	0.267
111	0.064	0.920	0.275
113	0.063	0.920	0.278
115	0.064	0.919	0.273
117	0.063	0.917	0.268
119	0.062	0.928	0.267
120	0.063	0.924	0.267
121	0.062	0.926	0.264
123	0.062	0.925	0.263
125	0.062	0.925	0.264
127	0.062	0.924	0.263
129	0.062	0.923	0.263
131	0.061	0.907	0.258
139	0.061	0.873	0.254
141	0.059	0.823	0.245
143	0.056	0.775	0.232
145	0.059	0.827	0.246
147	0.059	0.839	0.248
149	0.060	0.846	0.249
151	0.062	0.911	0.258
153	0.062	0.911	0.258
157	0.062	0.929	0.269
159	0.062	0.930	0.273
161	0.062	0.930	0.275
163	0.062	0.930	0.276
164	0.062	0.930	0.276

Table continues on next page

Table continues from previous page

Node	$CrCorr_{HL}[-]$	$CrCorr_{HH}[-]$	$CrCorr_{HQ}[-]$
166	0.062	0.928	0.276
167	0.062	0.928	0.277
169	0.062	0.930	0.277
171	0.062	0.930	0.277
173	0.062	0.930	0.277
177	0.062	0.929	0.279
179	0.062	0.929	0.279
181	0.062	0.930	0.279
183	0.062	0.929	0.280
184	0.062	0.926	0.280
185	0.063	0.927	0.282
187	0.063	0.927	0.283
189	0.063	0.929	0.281
191	0.063	0.926	0.282
193	0.063	0.927	0.281
195	0.063	0.925	0.282
197	0.063	0.924	0.280
199	0.061	0.931	0.270
201	0.061	0.931	0.270
203	0.061	0.931	0.270
204	0.063	0.927	0.283
205	0.061	0.930	0.268
206	0.057	0.922	0.253
207	0.059	0.928	0.261
208	0.054	0.910	0.245
209	0.052	0.881	0.233
211	0.052	0.869	0.231
213	0.052	0.866	0.231
215	0.052	0.856	0.232
217	0.052	0.852	0.233
219	0.052	0.849	0.233
225	0.052	0.849	0.233

Table continues on next page

Table continues from previous page

Node	$CrCorr_{HL}[-]$	$CrCorr_{HH}[-]$	$CrCorr_{HQ}[-]$
229	0.051	0.840	0.225
231	0.051	0.836	0.225
237	0.051	0.844	0.224
239	0.051	0.836	0.222
241	0.051	0.836	0.222
243	0.051	0.833	0.222
247	0.051	0.835	0.222
249	0.051	0.835	0.222
251	0.051	0.833	0.221
253	0.051	0.828	0.219
255	0.051	0.831	0.220
257	0.063	0.906	0.265
259	0.063	0.898	0.263
261	0.063	0.912	0.267
263	0.063	0.911	0.267
265	0.062	0.930	0.277
267	0.063	0.928	0.281
269	0.063	0.929	0.280
271	0.062	0.930	0.278
273	0.061	0.930	0.269
275	0.061	0.931	0.268

Table 5.3: *Values of cross correlation for Net3, between piezometric heads and nodal leakage flow rates ($CrCross_{HL}$, column 2), piezometric heads and piezometric heads ($CrCross_{HH}$, column 3), piezometric heads and pipe flow rates ($CrCross_{HQ}$, column 4).*

Pipe	$CrCorr_{QL}[-]$	**Pipe**	$CrCorr_{QL}[-]$
60	0.042	211	0.061
101	0.063	213	0.062

Table continues on next page

Table continues from previous page

Pipe	$CrCorr_{QL}[-]$	**Pipe**	$CrCorr_{QL}[-]$
103	0.057	215	0.055
105	0.063	217	0.044
107	0.034	219	0.028
109	0.062	221	0.060
111	0.061	223	0.064
112	0.050	225	0.068
113	0.061	229	0.077
114	0.059	231	0.071
115	0.054	233	0.043
116	0.065	235	0.067
117	0.056	237	0.074
119	0.051	238	0.075
120	0.057	239	0.072
121	0.063	240	0.075
122	0.056	241	0.075
123	0.073	243	0.075
125	0.065	245	0.056
129	0.044	247	0.043
131	0.041	249	0.038
135	0.038	251	0.026
137	0.023	257	0.023
145	0.048	261	0.064
147	0.043	263	0.024
149	0.040	269	0.068
151	0.042	271	0.041
153	0.038	273	0.061
155	0.040	275	0.051
159	0.043	277	0.024
161	0.047	281	0.052
163	0.031	283	0.053
169	0.043	285	0.040
171	0.053	287	0.042

Table continues on next page

Table continues from previous page

Pipe	$CrCorr_{QL}[-]$	**Pipe**	$CrCorr_{QL}[-]$
173	0.076	291	0.031
175	0.077	293	0.039
177	0.077	295	0.042
179	0.077	297	0.050
180	0.029	299	0.033
181	0.023	301	0.037
183	0.079	303	0.044
185	0.024	305	0.036
186	0.054	307	0.046
187	0.076	309	0.053
189	0.074	311	0.055
191	0.064	313	0.058
193	0.033	315	0.061
195	0.050	317	0.040
197	0.047	319	0.062
199	0.050	321	0.078
202	0.061	323	0.070
203	0.051	325	0.059
204	0.069	329	0.064
205	0.052	330	0.000
207	0.055	333	0.021
209	0.051		

Table 5.4: *Values of cross correlation for Net3, between pipe flow rates and nodal leakage flow rates ($CrCorr_{QL}$, column 2)*

5.3.4 The Normal-Score EnKF

Although the EnKF is relatively robust for non-linear model dynamics as in the case of WDSs, it performs not optimally for non-Gaussian parameter distributions, i.e. the emitter coefficients C that follow a bimodal-like distribution. A Normal-Score Transform (NST) is

applied, as reported in literature in Zhou et al. (2011). The method focuses on transforming the non-Gaussian distributed state variables and, most importantly, the non-Gaussian distributed model parameters, into a new vector that follows marginal Gaussian distributions. The NST is applied independently to each parameter and variable at all locations and at all times. Once the NST function is established, the NS-EnKF just follows the same procedure as the standard EnKF. A flow chart of the NS-EnKF consists of the following steps.

1. *Ensemble forecast.* A large number of equally likely stochastic realizations of the state vector $X_{t_i}^f$ is generated (equation (5.3)).

$$X_{t_i}^f = \left[C_1, ..., C_{nn}, H_{1(t_i)}, ..., H_{nn(t_i)}, Q_{1(t_i)}, ..., Q_{np(t_i)}, V_{in(t_i)}, V_{out(t_i)}\right] \tag{5.12}$$

2. *NST.* Establish the local cumulative distribution functions (cdfs) for all the components of $X_{t_i}^f$ from the ensemble of realizations. In our case there will be one such local cdf at each location for the C coefficients and another one for each H, Q, V_{in} and V_{out}. Use these local cdfs to build the NST function and transform $X_{t_i}^f$ into a new vector $X_{nst(t_i)}^f$, with all its components following marginal Gaussian distributions with zero mean and unit variance. The transformation functions need to be recomputed at each time step.

3. *Update.* State data are collected at time t_i (equation (5.4)).

$$z_{t_i}^m = \left[H_{1(t_i)}^m, ..., H_{nobsH(t_i)}^m, Q_{1(t_i)}^m, ..., Q_{nobsQ(t_i)}^m, V_{in(t_i)}^m, V_{out(t_i)}^m\right] \tag{5.13}$$

   These data z_{ti}^m are transformed into $z_{nst(t_i)}^m$ using the NST functions computed in the previous step. Next we apply equation (2.13) to update the state vector $X_{nst(t_i)}^f$, thus obtaining $X_{nst(t_i)}^u$.

4. *Backtransform.* The updated state vector $X_{nst(t_i)}^u$ is back transformed using the previously constructed transformation functions. Time advances one step, from t_i to t_{i+1}. The updated state vector $X_{t_i}^u$ becomes the current vector $X_{t_{i+1}}^f$ and we loop back to the forecast step.

To sum up, the proposed method applies the EnkF always to a state vector all of which components follow a marginal Gaussian distribution. Furthermore, using the NST we ensure that the prior non-Gaussian marginals of the model parameters are kept throughout.

The recursive application of the model provides a progressive correction of the C coefficients for the network nodes. The spatial and temporal distribution of L is then retrieved from the propagation of the system over time through the pressure driven hydraulic model adopted.

5.3.5 Description of the measurement scenarios

The measurements used in the assimilation procedure are provided from the synthetic solution systems of Anytown (figure 5.4) and Net3 (figure 5.5). An extended period simulation is run according to the known C field and the system governing equations. Through the pressure driven hydraulic network model the H, Q, V_{in} and V_{out} values are retrieved every 6 hours, that is for 4 time instants in the 24 hours, for Anytown, and every 4 hours, that is for 6 time instants in the 24 hours, for Net3. These values are affected by an uncertainty that is defined through a variation coefficient (CV) set to 2.5% for H and Q and to 10% for V_{in} and V_{out}, on the basis of physical considerations.

The performances of the NS-EnKF to infer the spatial distribution of the water losses L through the calibration of the C values are evaluated in different scenarios, where the number of collected measurements of each type is progressively increased according to the cross correlation criterion, as described in table (5.5) for Anytown and in table (5.6) for Net3. Two different assimilation schemes are proposed.

The first assimilation scheme (scenarios a) uses a first data assimilation cycle lasting 24 hours, during which only V_{in}^m and V_{out}^m are included. The second part of the data assimilation procedure uses all the measurements available at the observation times, that are transient H_{nobsH}^m and Q_{nobsQ}^m, and the known volumes V_{in}^m and V_{out}^m until the considered assimilation time, the last assimilation time being $t = 24$. The purpose of this scheme is to facilitate the calibration of the C values, that should more quickly reach the correct order of magnitude, by using the information available from the global flow balance, coming from the knowledge of the volumes. Because no information on the spatial distribution of C comes from V_{in}^m and V_{out}^m, the second stage of the procedure allows for a more precise spatial estimation of the C field through transient measurements.

The second assimilation scheme (scenarios b) is equivalent to the second part of the first assimilation scheme, which consists in the update through all the measurements available at each observation time, that are transient H_{nobsH}^m and Q_{nobsQ}^m and the known volumes V_{in}^m and V_{out}^m until the considered time, the last measurement time being $t = 24$.

For Anytown network 8 different scenarios are considered, assuming to have the knowledge of a quarter (scenarios 1a and 1b), a half (scenarios 2a and 2b) or three quarters (scenarios 3a and 3b) of the network variables. The case of complete network knowledge (scenarios 4a and 4b) is analyzed for comparison.

For Net3 network 16 different scenarios are considered, assuming that the knowledge of the system progressively increases from scenario 1a to 7a and in parallel from 1b to 7b. The cases of 5 (scenarios 1a and 1b), 10 (scenarios 3a and 3b) and 15 (scenarios 5a and

5b) assimilated measurements are considered for comparison with the proposed scenarios for Anytown. The case of complete network knowledge (scenarios 8a and 8b) is analyzed for reference.

	Assimilated measurements									
Scenario	**1st stage**			**2nd stage**						
1a	V_{in}	+	V_{out}	5H	+	5Q	+	V_{in}	+	V_{out}
2a	V_{in}	+	V_{out}	10H	+	10Q	+	V_{in}	+	V_{out}
3a	V_{in}	+	V_{out}	15H	+	15Q	+	V_{in}	+	V_{out}
4a	V_{in}	+	V_{out}	20H	+	39Q	+	V_{in}	+	V_{out}
1b		-		5H	+	5Q	+	V_{in}	+	V_{out}
2b		-		10H	+	10Q	+	V_{in}	+	V_{out}
3b		-		15H	+	15Q	+	V_{in}	+	V_{out}
4b		-		20H	+	39Q	+	V_{in}	+	V_{out}

Table 5.5: *Measurements scenarios for Anytown.*

	Assimilated measurements									
Scenario	**1st stage**			**2nd stage**						
1a	V_{in}	+	V_{out}	5H	+	5Q	+	V_{in}	+	V_{out}
2a	V_{in}	+	V_{out}	5H	+	6Q	+	V_{in}	+	V_{out}
3a	V_{in}	+	V_{out}	10H	+	10Q	+	V_{in}	+	V_{out}
4a	V_{in}	+	V_{out}	11H	+	13Q	+	V_{in}	+	V_{out}
5a	V_{in}	+	V_{out}	15H	+	15Q	+	V_{in}	+	V_{out}
6a	V_{in}	+	V_{out}	22H	+	27Q	+	V_{in}	+	V_{out}
7a	V_{in}	+	V_{out}	44H	+	55Q	+	V_{in}	+	V_{out}
8a	V_{in}	+	V_{out}	89H	+	111Q	+	V_{in}	+	V_{out}
1b		-		5H	+	5Q	+	V_{in}	+	V_{out}
2b		-		5H	+	6Q	+	V_{in}	+	V_{out}
3b		-		10H	+	10Q	+	V_{in}	+	V_{out}
4b		-		11H	+	13Q	+	V_{in}	+	V_{out}
5b		-		15H	+	15Q	+	V_{in}	+	V_{out}
6b		-		22H	+	27Q	+	V_{in}	+	V_{out}

Table continues on next page

Table continues from previous page

	Assimilated measurements								
Scenario	**1st stage**	**2nd stage**							
7b	-	44H	+	55Q	+	V_{in}	+	V_{out}	
8b	-	89H	+	111Q	+	V_{in}	+	V_{out}	

Table 5.6: *Measurements scenarios for Net3.*

5.4 Results and discussion

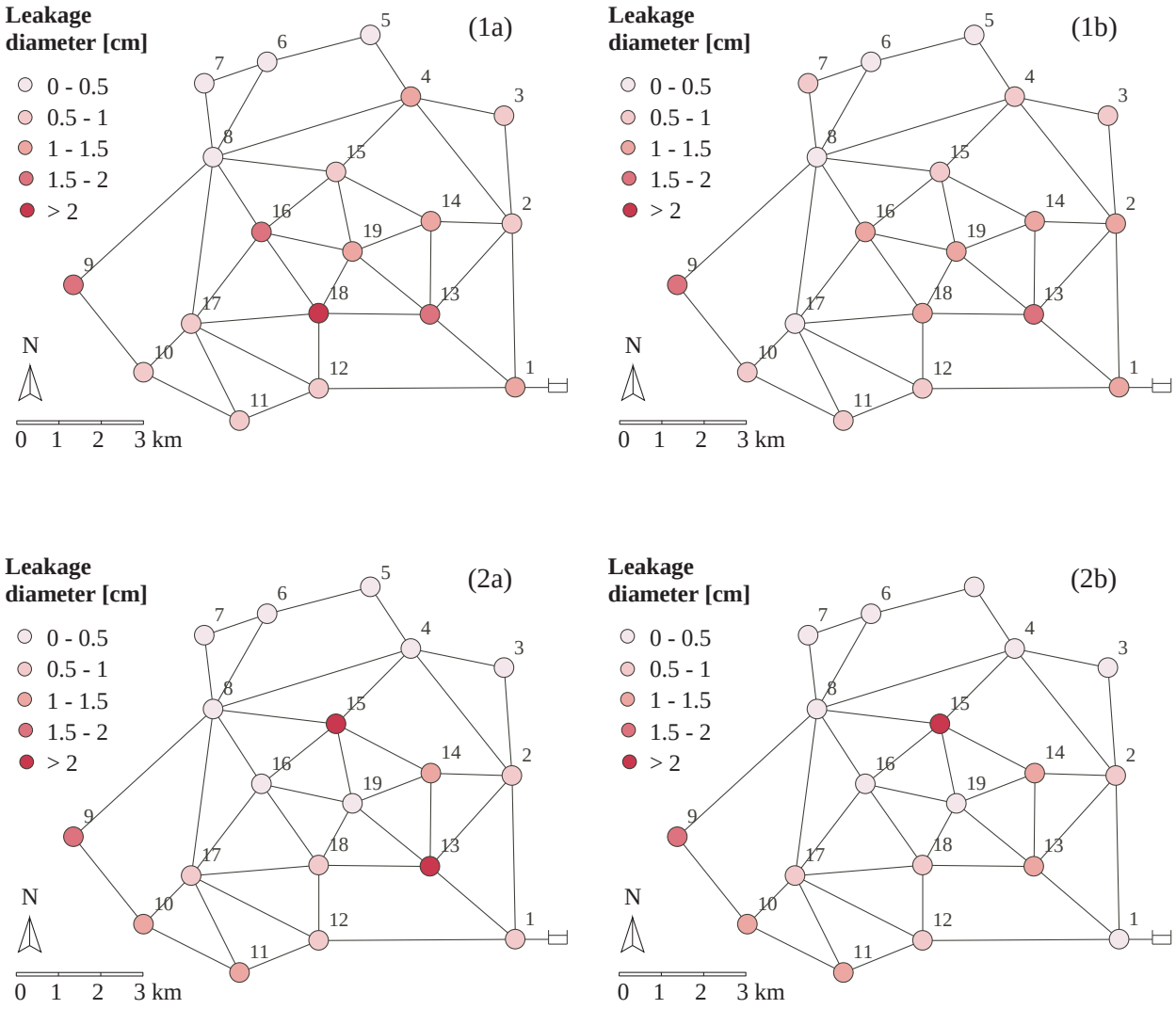

Figure 5.9: *NS-EnKF results for Anytown, in terms of leakage diameter for scenario 1a (panel 1a), scenario 1b (panel 1b), scenario 2a (panel 2a) and scenario 2b (panel 2b).*

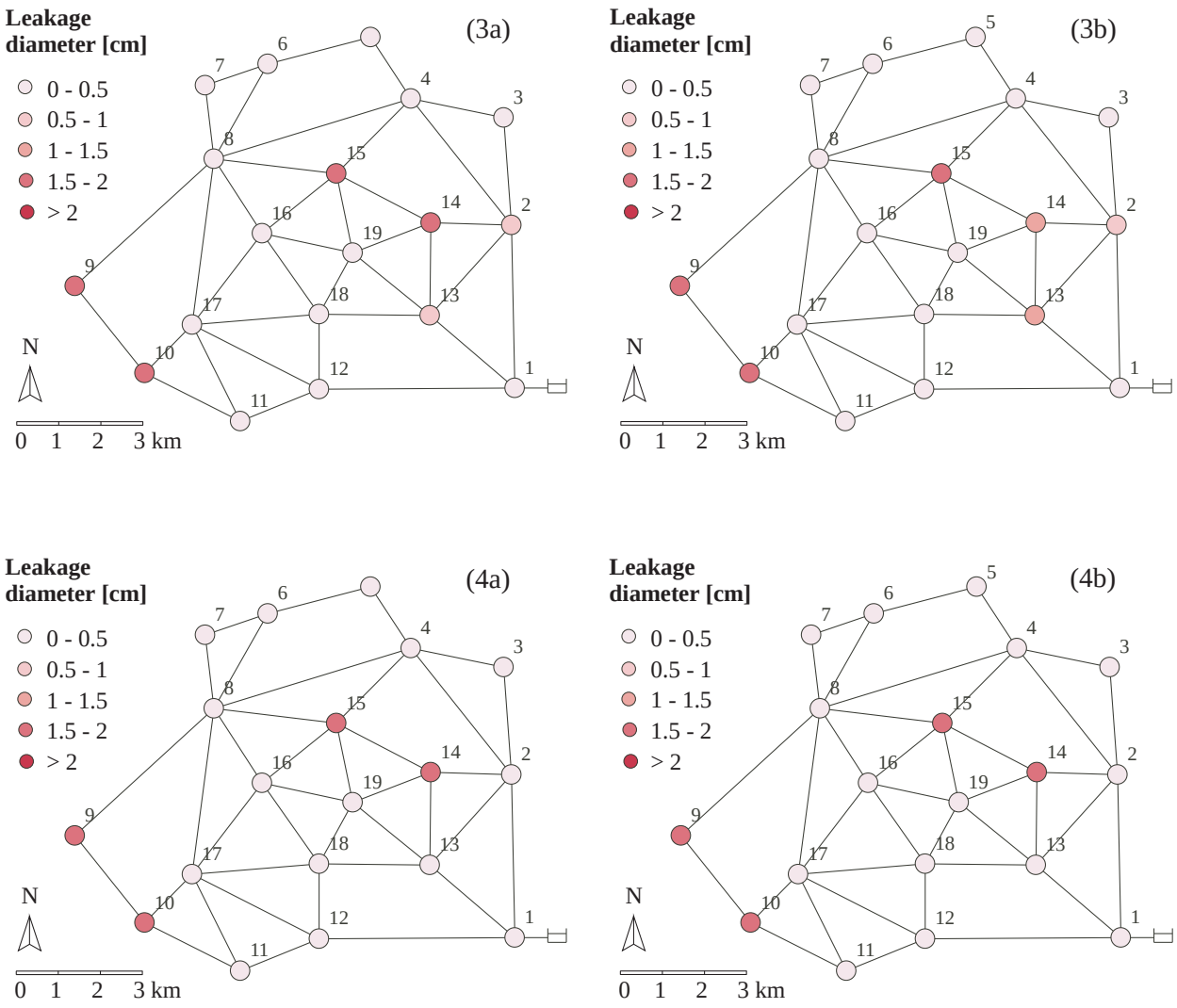

Figure 5.10: *NS-EnKF results for Anytown in terms of leakage diameter for scenario 3a (panel 3a), scenario 3b (panel 3b), scenario 4a (panel 4a) and scenario 4b (panel 4b).*

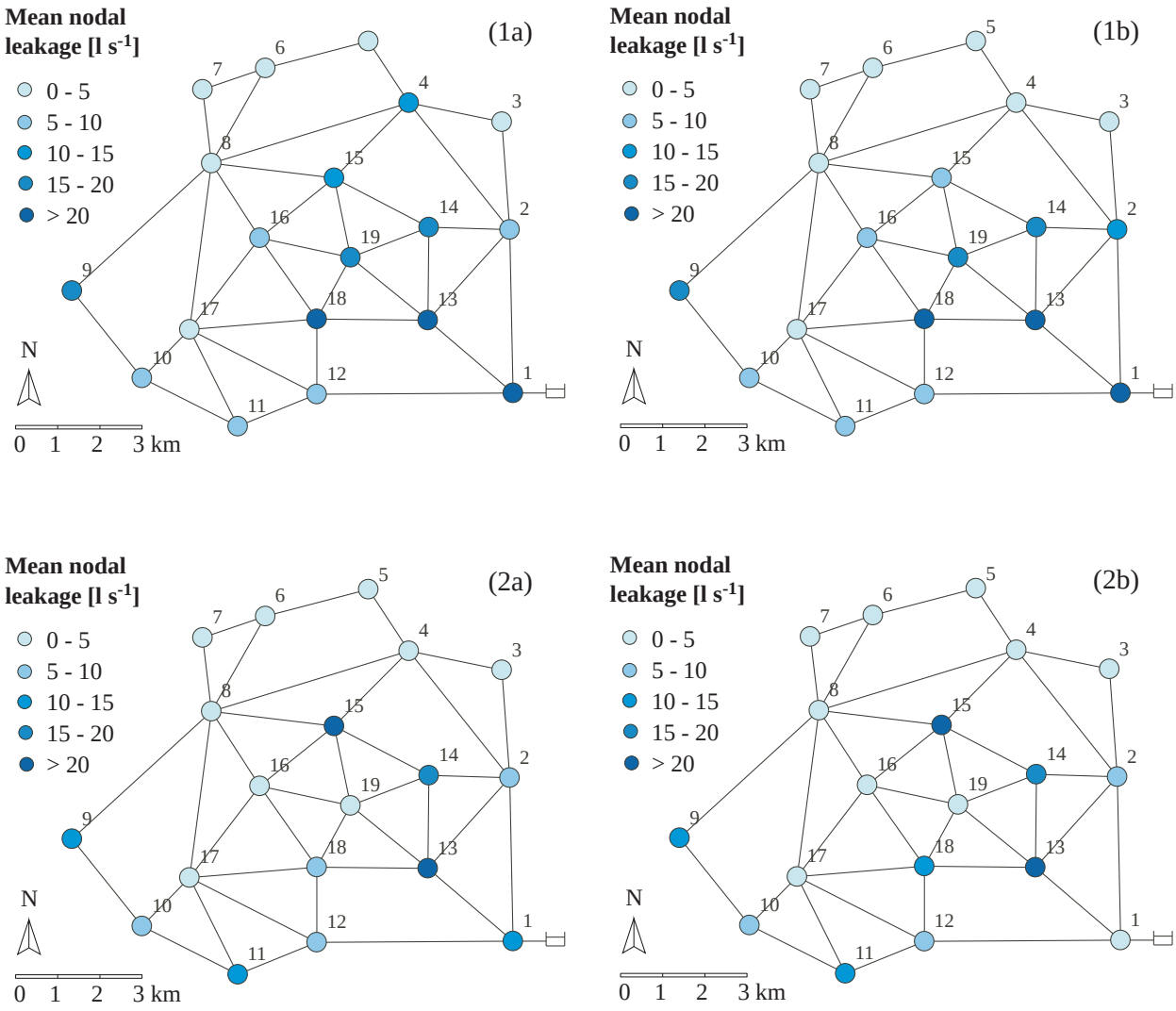

Figure 5.11: *NS-EnKF results for Anytown, in terms of mean nodal leakage flow rates for scenario 1a (panel 1a), scenario 1b (panel 1b), scenario 2a (panel 2a) and scenario 2b (panel 2b).*

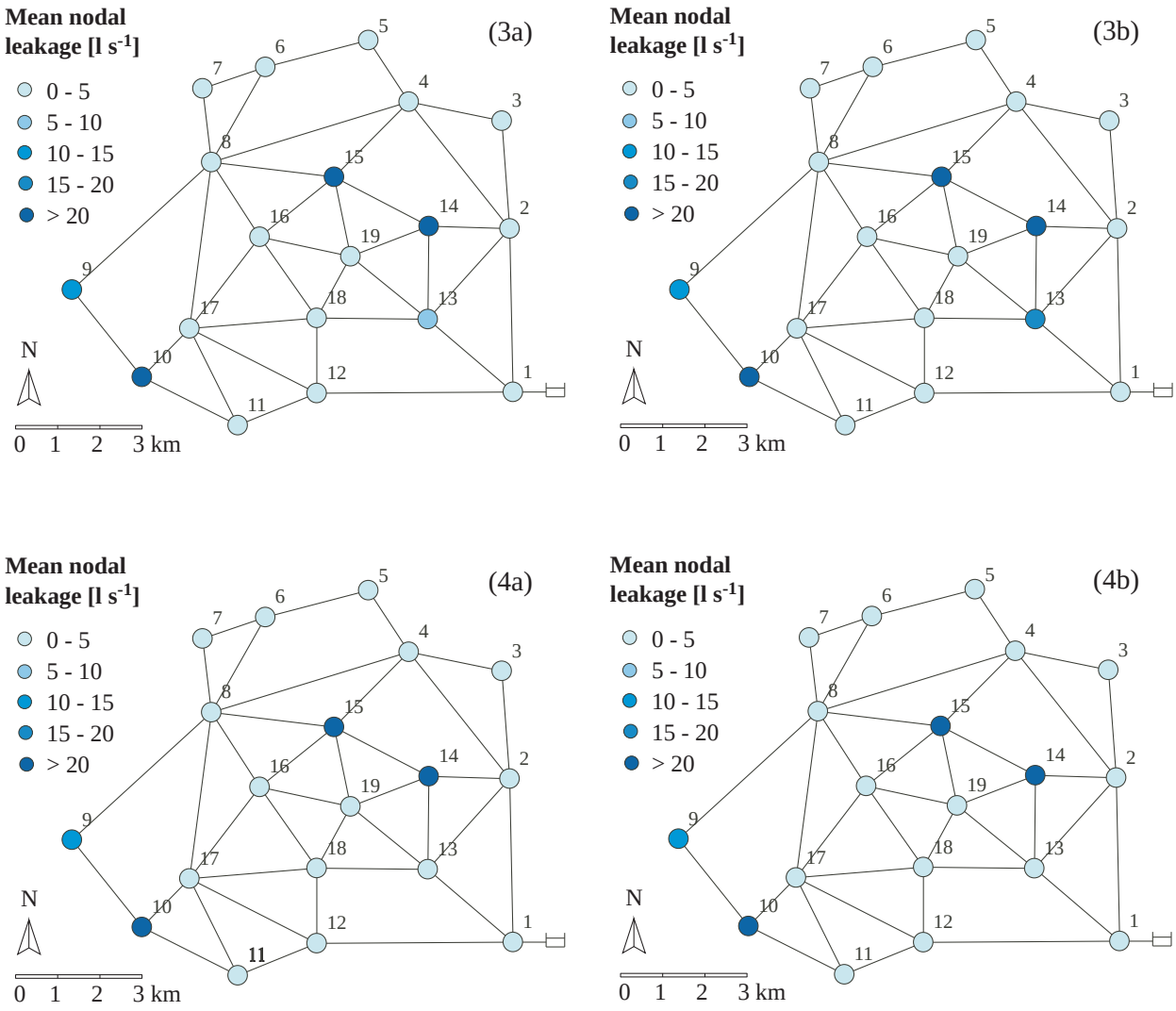

Figure 5.12: *NS-EnKF results for Anytown in terms of mean nodal leakage flow rates for scenario 3a (panel 3a), scenario 3b (panel 3b), scenario 4a (panel 4a) and scenario 4b (panel 4b).*

Scenario	MAE_C $[ls^{-1}m^{-1}]$	MAE_L $[ls^{-1}]$	$RE_{V_{in}}$ $[-]$	$RE_{V_{out}}$ $[-]$	MAE_H $[m]$	MAE_Q $[ls^{-1}]$
1a	0.428	14.721	0.168	1.177	4.802	8.518
2a	0.263	9.121	0.120	0.852	4.065	5.909
3a	0.054	2.105	0.036	0.248	1.180	1.422
4a	0.015	0.421	0.009	0.059	0.268	0.305
1b	0.363	13.074	0.139	0.940	4.023	7.522
2b	0.183	6.344	0.058	0.412	2.013	3.299
3b	0.070	2.800	0.035	0.244	1.221	1.646
4b	0.008	0.230	0.005	0.038	0.211	0.225

Table 5.7: *NS-EnKF results for Anytown. Measurement scenarios (column 1), mean absolute error for the emitter coefficients (MAE_C, column 2), mean absolute error for the nodal leakage flow rates (MAE_P, column 3), relative error for the incoming volume ($RE_{V_{in}}$, column 4), relative error for the leakage outflow ($RE_{V_{out}}$, column 5), mean absolute error for pressure heads (MAE_H, column 6), mean absolute error for pipe flow rates (MAE_Q, column 7).*

Scenario	**Pipe % to survey**	**Scenario**	**Pipe % to survey**
1a	46.1	1b	41.0
2a	31.5	2b	33.7
3a	16.8	3b	24.1
4a	16.8	4b	16.8

Table 5.8: *NS-EnKF results for Anytown in terms of percentage of pipe length to be surveyed.*

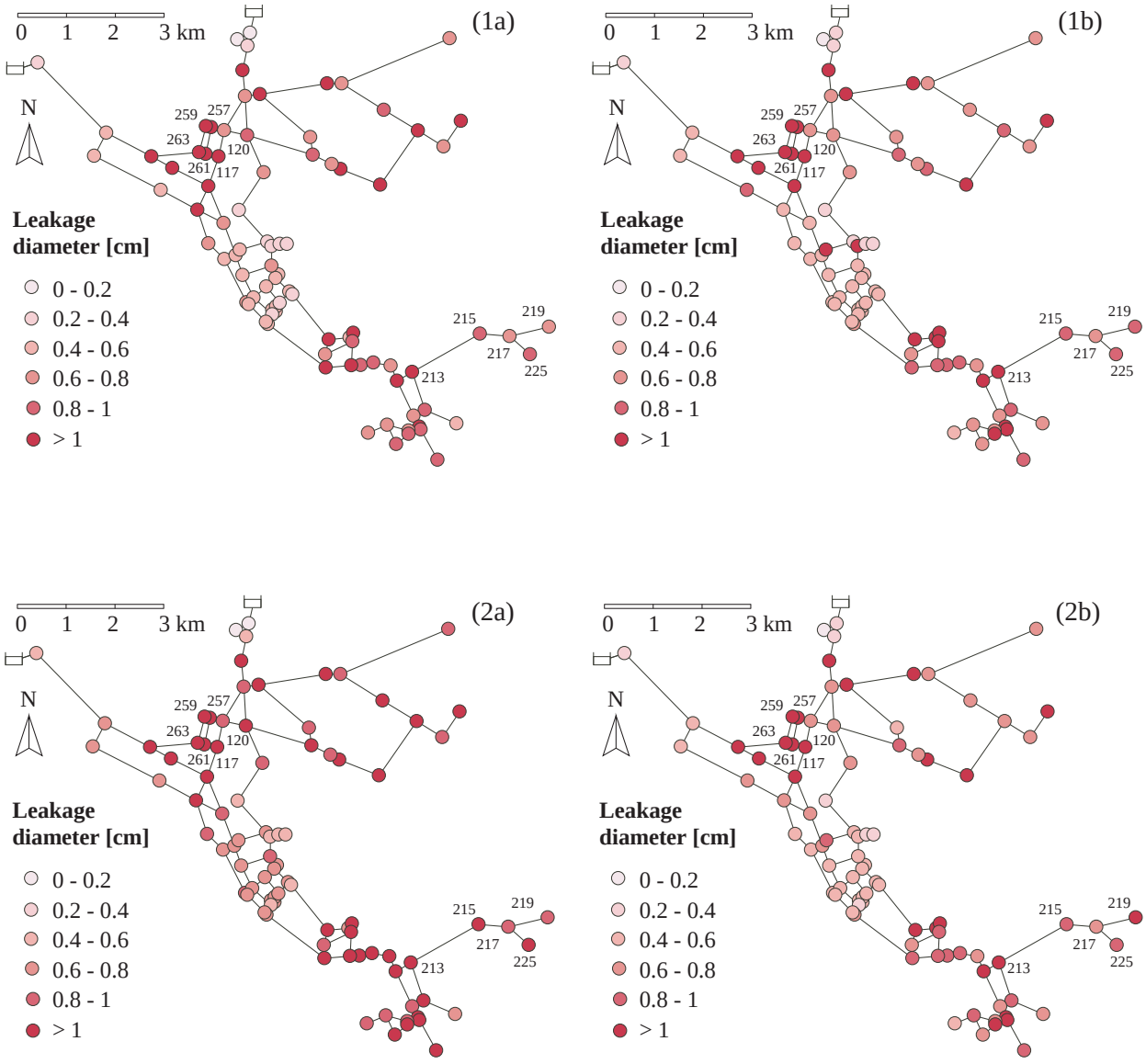

Figure 5.13: *NS-EnKF results for Net3, in terms of leakage diameter for scenario 1a (panel 1a), scenario 1b (panel 1b), scenario 2a (panel 2a) and scenario 2b (panel 2b).*

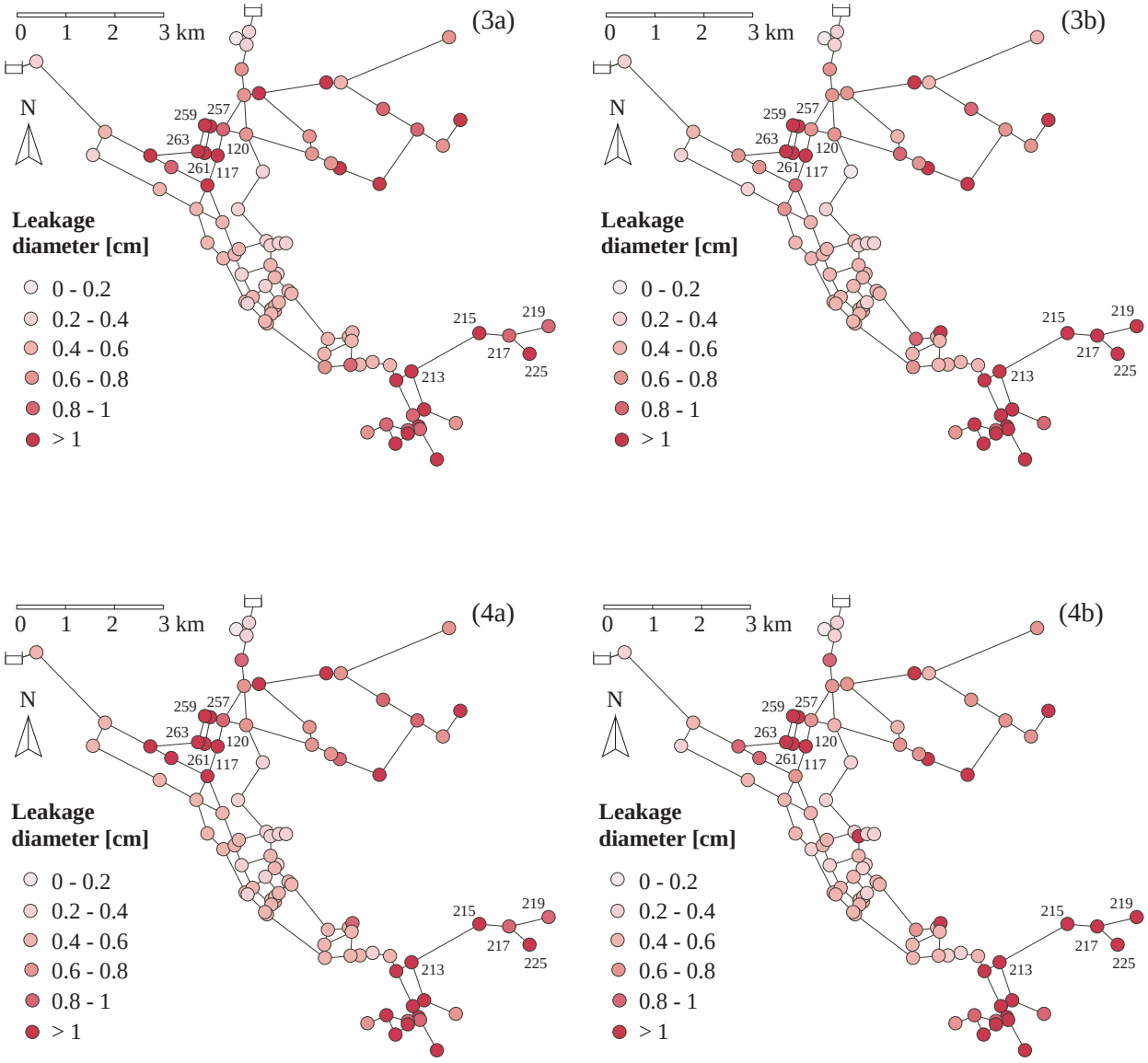

Figure 5.14: *NS-EnKF results for Net3 in terms of leakage diameter for scenario 3a (panel 3a), scenario 3b (panel 3b), scenario 4a (panel 4a) and scenario 4b (panel 4b).*

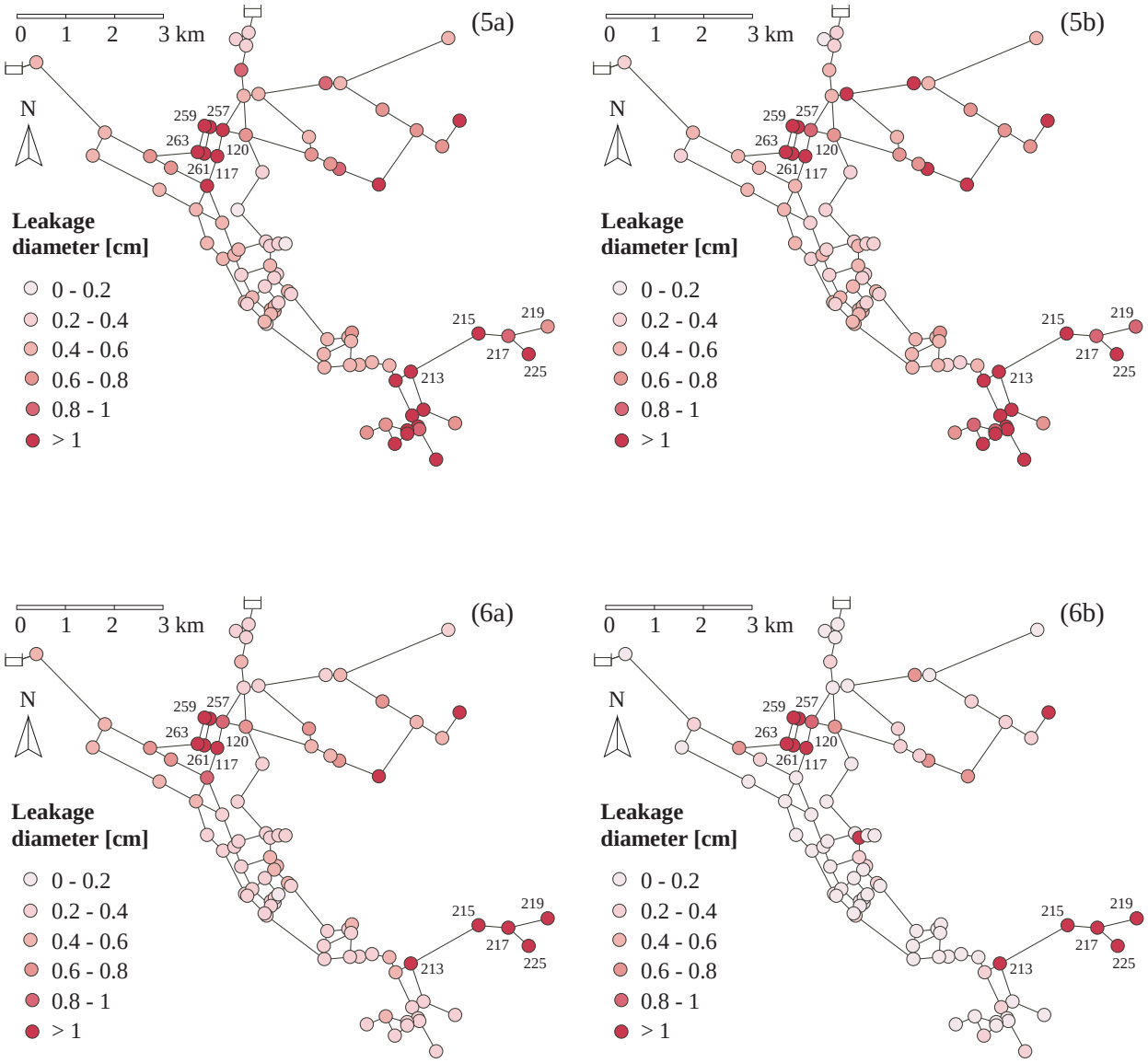

Figure 5.15: *NS-EnKF results for Net3 in terms of leakage diameter for scenario 5a (panel 5a), scenario 5b (panel 5b), scenario 6a (panel 6a) and scenario 6b (panel 6b).*

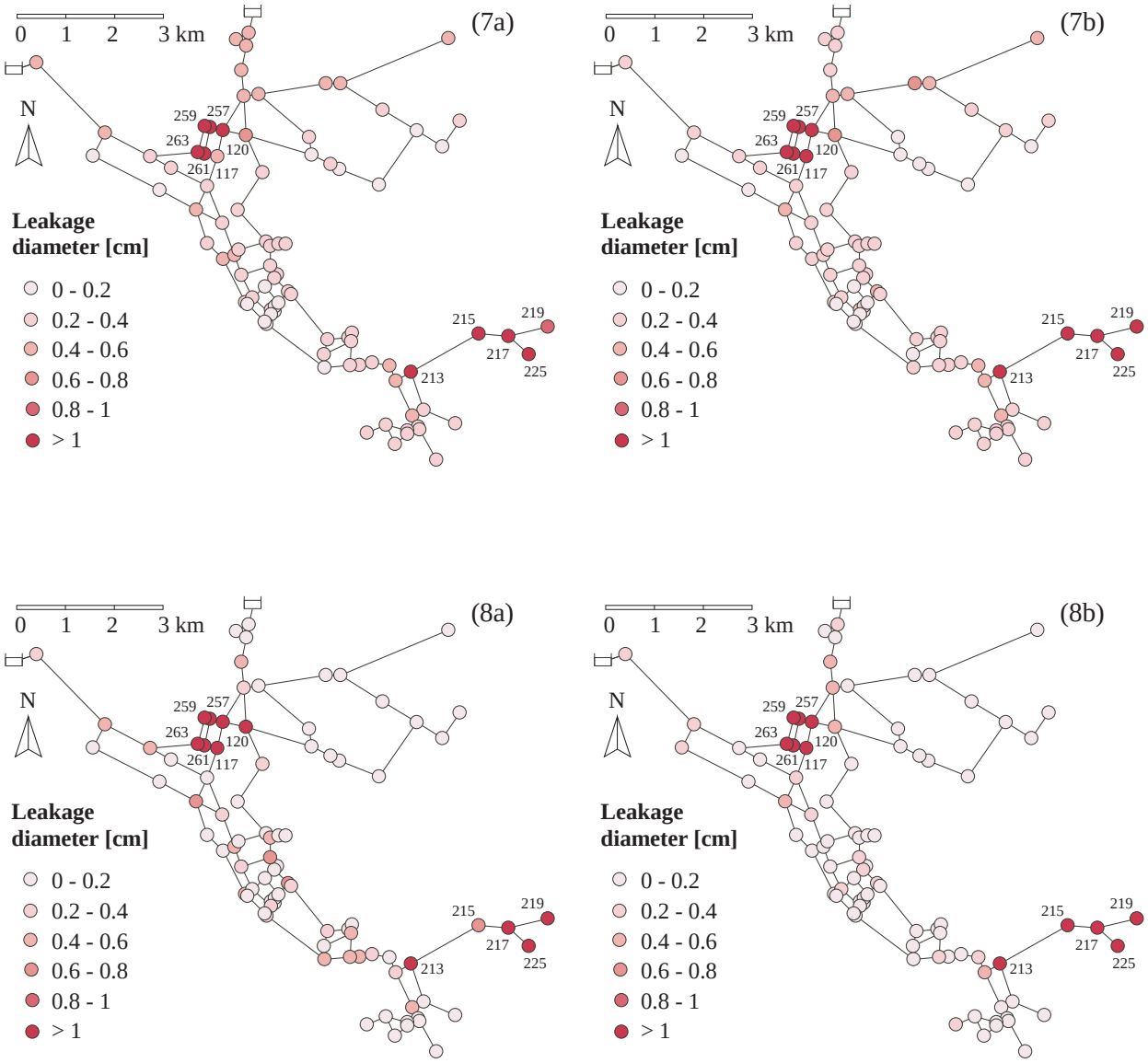

Figure 5.16: *NS-EnKF results for Net3 in terms of leakage diameter for scenario 7a (panel 7a), scenario 7b (panel 7b), scenario 8a (panel 8a) and scenario 8b (panel 8b).*

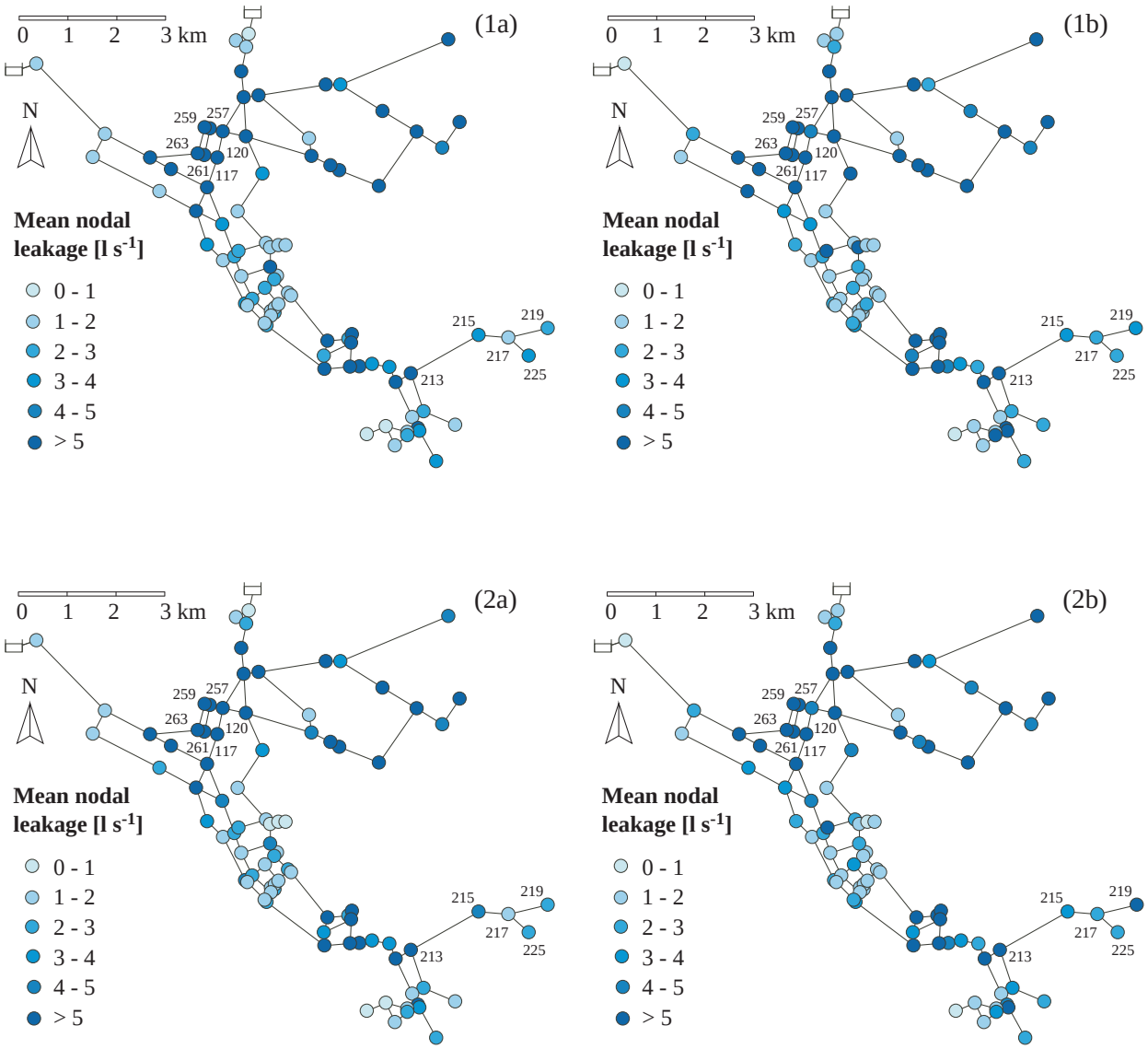

Figure 5.17: *NS-EnKF results for Net3, in terms of mean nodal leakage flow rates for scenario 1a (panel 1a), scenario 1b (panel 1b), scenario 2a (panel 2a) and scenario 2b (panel 2b).*

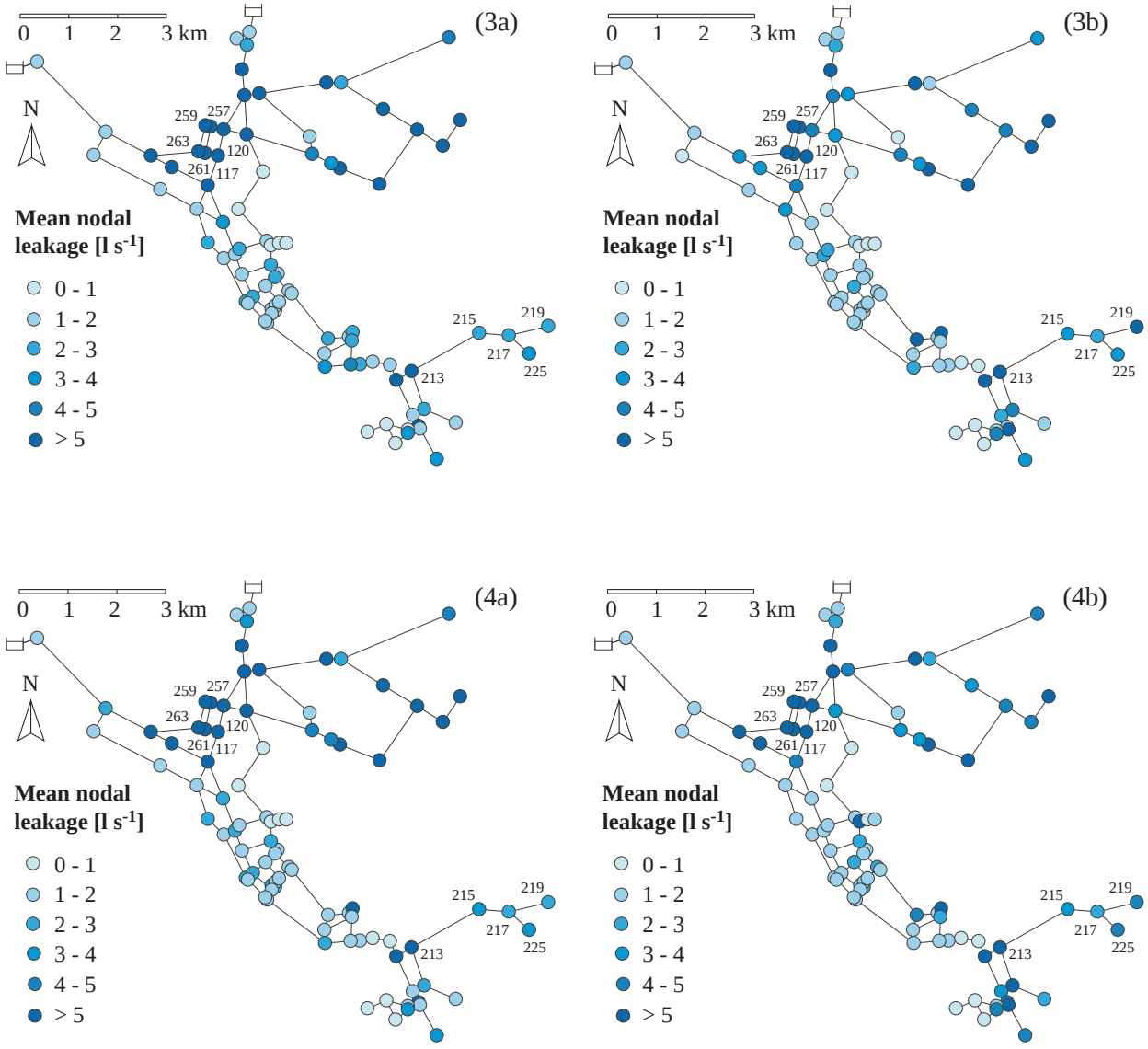

Figure 5.18: *NS-EnKF results for Net3 in terms of mean nodal leakage flow rates for scenario 3a (panel 3a), scenario 3b (panel 3b), scenario 4a (panel 4a) and scenario 4b (panel 4b).*

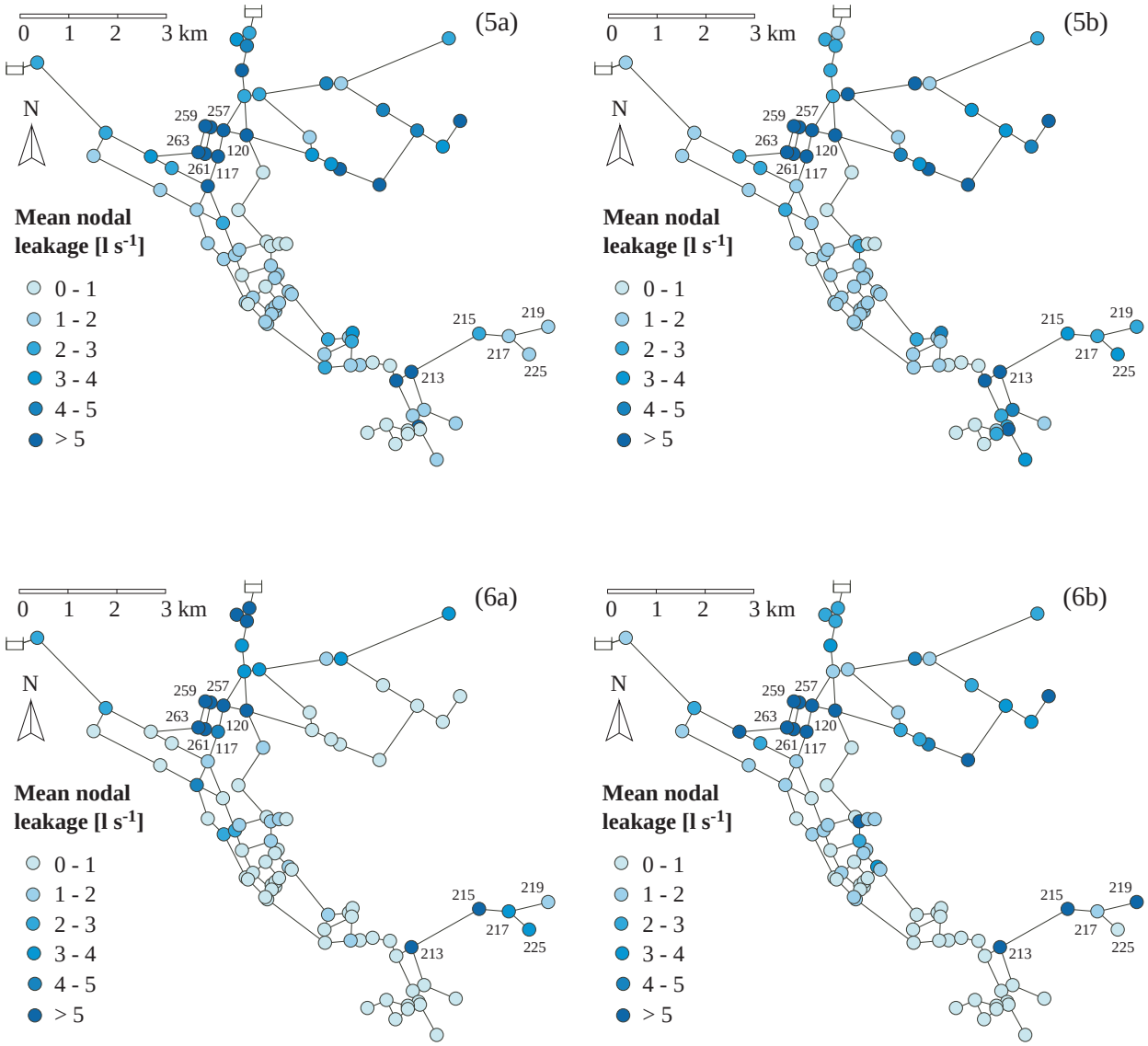

Figure 5.19: *NS-EnKF results for Net3 in terms of mean nodal leakage flow rates for scenario 5a (panel 5a), scenario 5b (panel 5b), scenario 6a (panel 6a) and scenario 6b (panel 6b).*

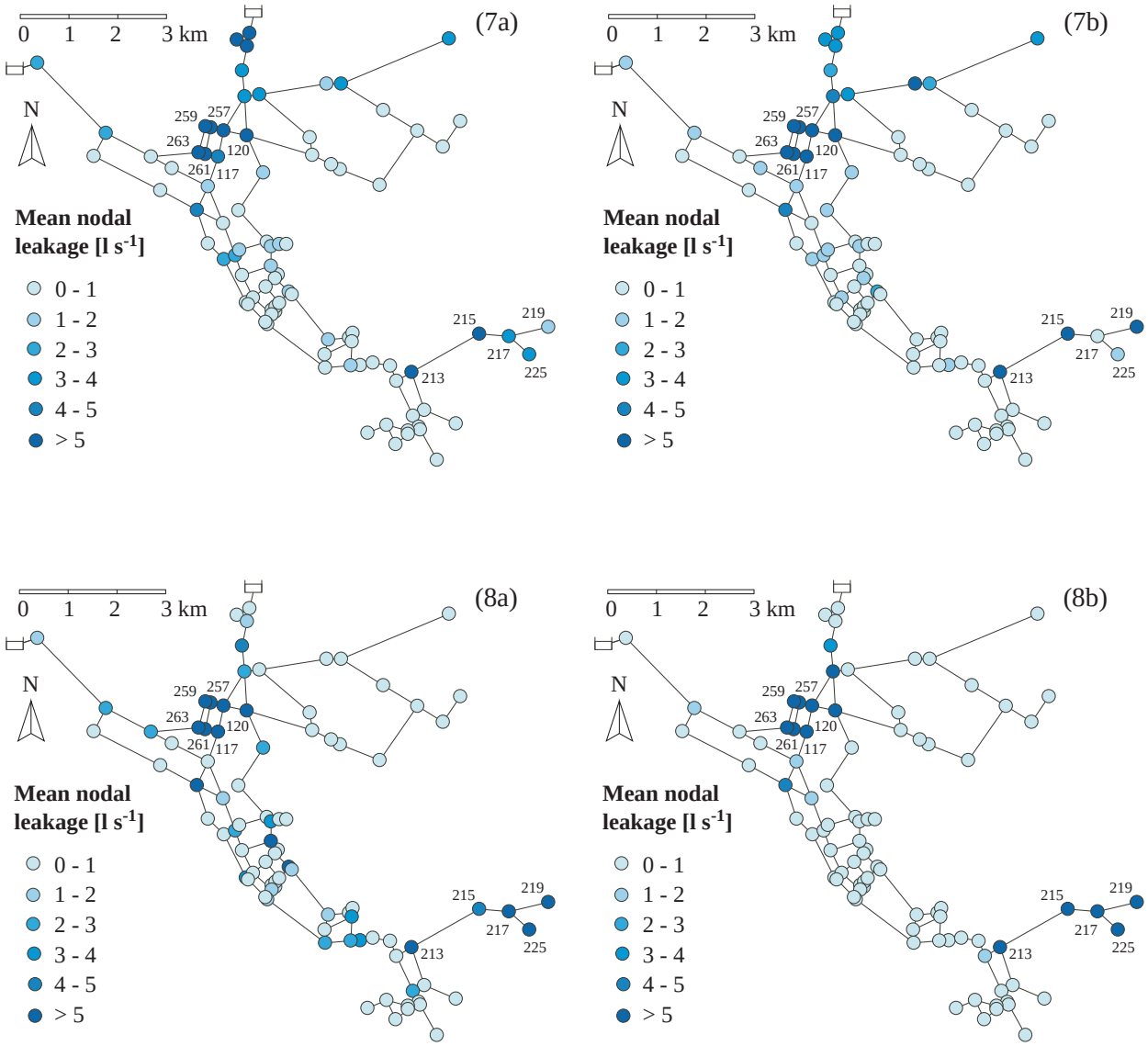

Figure 5.20: *NS-EnKF results for Net3 in terms of mean nodal leakage flow rates for scenario 7a (panel 7a), scenario 7b (panel 7b), scenario 8a (panel 8a) and scenario 8b (panel 8b).*

Scenario	MAE_C $[ls^{-1}m^{-1}]$	MAE_L $[ls^{-1}]$	$RE_{V_{in}}$ $[-]$	$RE_{V_{out}}$ $[-]$	MAE_H $[m]$	MAE_Q $[ls^{-1}]$
1a	0.448	10.820	0.228	0.815	23.073	25.340
2a	0.459	10.779	0.230	0.832	23.313	24.886
3a	0.603	8.821	0.229	0.810	24.429	20.474
4a	0.540	8.653	0.223	0.750	23.394	19.516
5a	0.734	8.121	0.235	0.840	24.948	22.831
6a	2.593	7.558	0.197	0.635	22.167	24.096
7a	0.922	6.472	0.144	0.459	15.858	22.019
8a	0.091	3.554	0.079	0.220	7.346	7.853
1b	0.463	11.016	0.237	0.869	23.472	25.520
2b	0.452	10.609	0.231	0.855	23.536	26.161
3b	1.012	11.773	0.261	1.085	27.301	30.903
4b	0.542	9.454	0.229	0.816	24.235	22.494
5b	0.671	8.133	0.237	0.866	25.126	22.619
6b	1.741	6.438	0.184	0.592	21.205	22.141
7b	0.969	4.836	0.117	0.389	14.829	20.255
8b	0.023	0.983	0.002	0.006	0.191	1.292

Table 5.9: *NS-EnKF results for Net3. Measurement scenarios (column 1), mean absolute error for the emitter coefficients (MAE_C, column 2), mean absolute error for the nodal leakage flow rates (MAE_P, column 3), relative error for the incoming volume ($RE_{V_{in}}$, column 4), relative error for the leakage outflow ($RE_{V_{out}}$, column 5), mean absolute error for pressure heads (MAE_H, column 6), mean absolute error for pipe flow rates (MAE_Q, column 7).*

Scenario	**Pipe % to survey**	**Scenario**	**Pipe % to survey**
1a	45.5	1b	45.3
2a	41.9	2b	41.6
3a	35.6	3b	24.1

Table continues on next page

Table continues from previous page

Scenario	Pipe % to survey	Scenario	Pipe % to survey
4a	35.7	4b	28.4
5a	25.2	5b	15.8
6a	15.8	6b	14.3
7a	20.0	7b	11.5
8a	10.2	8b	13.7

Table 5.10: *NS-EnKF results for Net3 in terms of percentage of pipe length to be surveyed.*

The results of the NS-EnKF application for Anytown are reported in figure (5.9) and figure (5.10) in terms of leakage diameters. The corresponding results in terms of mean daily nodal leakage flow rates $\bar{L}$ are retrieved from the propagation over time of the C parameters estimated from the model, and are reported in figure (5.11) and figure (5.12). Darker colours correspond to higher values. The estimated values are compared with the reference system solution for Anytown (figure 5.4). As expected, the NS-EnKF gives better performances as the number of assimilated measurements increases, progressively reducing the uncertainty of the solution, both in terms of parameters and $\bar{L}$ variables. The mean absolute errors for the estimated emitter coefficients C (MAE_C) and the mean absolute errors for leakages $\bar{L}$ (MAE_L) are reported in column 2 and column 3 of table (5.7) respectively and are consistent with the output of the images. As it is proved from scenarios 1a and 1b for Anytown, the use of 5 measurements of H and Q, besides the volumes V_{in} and V_{out}, is not sufficient to detect any loss of the two leaking areas that characterize the synthetic solution for Anytown (nodes 9 and 10, nodes 14 and 15). As soon as the information available regards half of the system, that is 10 measurements of H and Q (scenarios 2a and 2b), the model is more effective in defining the two leakage areas for Anytown, being rather precise for 15 measurements of H and Q (scenarios 3a and 3b). Scenarios 4a and 4b prove the fairness of the NS-EnKF, which is able to select only nodes 9, 10, 14 and 15 both in terms of parameters and $\bar{L}$, when 20 H measurements and 39 Q measurements are assimilated, besides V_{in} and V_{out}. In general, the adoption of the first assimilation scheme (scenarios a), which plans to run a first assimilation cycle including only the volume measurements, is not able to improve the results obtained in the scenarios of type b, in which all the measurements of H, Q, V_{in} and V_{out} are included all together as soon as available. The major effect of the

former assimilation scheme is providing solution fields with higher peak values, either at the solution nodes or at the adjacent ones. The numerical results suggest that the second assimilation scheme is more successful than the first one. This general trend is shown not only from MAE_C and MAE_L, but also in terms of relative errors on the volumes ($RE_{V_{in}}$ for V_{in} and $RE_{V_{out}}$ for V_{out}). The same holds for the mean absolute errors on piezometric heads (MAE_H) and on pipe flow rates (MAE_Q). All the variable values are retrieved through an extended period simulation of the pressure driven hydraulic model, using the NS-EnKF estimated C coefficients. Table (5.8) gives the model results in terms of percentage of the network pipe length to be examined through in situ techniques. The results for the second type of assimilation scheme (scenarios b) are better than the corresponding results for the same level of system knowledge in scenarios a.

The results of the NS-EnKF application for Net3 network are reported in terms of model estimated leakage diameters in figures (5.13), (5.14), (5.15) and (5.16). The corresponding mean daily leakages $\bar{L}$, calculated from the transient hydraulic model, are represented in figures (5.17), (5.18), (5.19) and (5.20). The synthetic solution for Net3 (figure (5.5)) is used as a reference for the evaluation of the NS-EnKF performances and it is characterized by two leakage areas. The first area, including nodes 117, 120, 257, 259, 261 and 263, is a looped zone close to the system sources, while the second area, including nodes 213, 215, 217, 219 and 225, is a length of branched pipeline. Although scenarios 1a, 1b, 2a and 2b for Net3 demonstrate that about a five percent system knowledge is still insufficient for the identification of any of the leakage areas, the NS-EnKF is better capable of estimating the spatial distribution of $\bar{L}$ as the number of assimilated measurements increases. In scenarios 3a and 3b (10 H and 10 Q measurements), 4a and 4b (11 H and 13 Q measurements) and 5a and 5b (15 H and 15 Q measurements) the gap between the estimated system state and the true solution progressively closes. As the figures show, the selected leakage nodes progressively concentrate on the two solution areas for Net3, both in terms of parameters and $\bar{L}$ variables. In scenarios 6a and 6b (22 H and 27 Q measurements), and 7a and 7b (44 H and 55 Q measurements), corresponding to a quarter and a half of the system knowledge respectively, the two solution areas for Net3 are defined increasingly better. Scenarios 8a and 8b prove the fairness of the NS-EnKF, which is able to select only the network nodes belonging to the solution loop and to the solution pipe branch, when 89 H measurements and 11 Q measurements are assimilated, besides V_{in} and V_{out}. For a small number of assimilated measurements, the model suggest rather high peak values for the selected leakage nodes, spreading such nodes on the whole network. Regarding the type of assimilation scheme adopted, table (5.9) lists the calculated errors for the estimated

parameters (MAE_C) and the model retrieved variables (MAE_L, $RE_{V_{in}}$, $RE_{V_{out}}$, MAE_H and MAE_Q), and shows how the results for scenarios 3a and 4a are slightly better than the corresponding results for scenarios 3b and 4b, meaning that for a relatively small number of measurements the assimilation of the only V_{in} and V_{out} values, before all the available measurements are included in the assimilation procedure, could be helpful. However, as the number of available measurements exceeds a quarter of the system knowledge (scenarios 6, 7 and 8), results for scenarios b are better than results for scenarios a. Table (5.10) gives the results for Net3 in terms of percentage of the network pipe length to be examined through in situ techniques. The results for the second type of assimilation scheme (scenarios b) are always better than the results for the corresponding scenarios of type a, the overall model performance being better for the former assimilation scheme.

5.5 Final remarks

In this work, the capabilities of the NS-EnKF to retrieve the spatial distribution of the water losses through the calibration of the emitter coefficients C were investigated on the two synthetic benchmark systems of Anytown and Net3. The aim was to compare the effectiveness of the NS-EnKF when applied on two different topological schemes, Net3 being more complex than Anytown. Two different assimilation schemes were proposed.

As expected, the NS-EnKF gives better performances as the number of assimilated measurements increases, progressively reducing the gap between the estimated system state and the true solution. In both the systems here analyzed, the NS-EnKF suffer the undifferentiated and small cross correlation values between the pressure heads at the nodes $\bar{H}$ and the leakages $\bar{L}$. This determines the need to assimilate a high number of measurements to define the two leakage areas characterizing the synthetic solution. These requirements in terms of measurements are 10 H and 10 Q measurements (scenarios 3, corresponding to a half of the system knowledge) for Anytown, and 22 H and 27 Q measurements (scenarios 4, corresponding to a quarter of the system knowledge) for Net3, thus the mentioned effect is more noticeable in Anytown than in Net3, due to the higher degree of network scheletonization and the smaller number of elements used to represent the WDS. Besides, the pipe flow rates show a well defined cross correlation structure in both cases. The assimilation of Q measurements is more effective than the assimilation of H values, the flow rate being a direct measurement of the variable L that we are interested to estimate.

The fairly high percentage of network elements that need to monitored in Anytown for a satisfactory calibration, and the progressive closing of the gap between the estimated

system state and the true solution as the measurements number increases, are both due to the structure of the network that is regularly looped. In the case of Net3 a much smaller percentage of elements is required to be monitored in order to effectively detect the leaks. This is justified from the structure of Net3 network which is more complex, that is less looped, than Anytown and shows several branches. On such branches a pressure drop due to an eventual leakage cannot be recovered as it happens in looped systems like Anytown, thus determining the greater sensitivity of the pressure values to the spatial distribution of the flow rates, that is a greater effectiveness of the coupled model here applied. The cross correlation relationship between water losses $\bar{L}$ and potentially measured variables (H and Q), that contributes to determine the coupled model efficiency, is indeed closely linked to the network topological structure: the more complex and least looped the network analyzed, the highest the performance of the coupled model.

A relevant aspect is the weak distinction of the selected solution nodes for Net3, meaning that such nodes are correctly identified by the model, but are poorly differentiated from the lower estimated values. This is due to the specific physical and topological characteristics of Net3, which shows a high cross correlation between the piezometric head values $\bar{H}$. This means that the specific pressure value H that is realized at a network node strongly influences the H values that are realized on the adjacent nodes. Thus when a leakage characterizes a node, the pressure decrease that occurs on such node as a consequence of the loss is immediately transmitted to the adjacent network nodes. For this reason the $\bar{H}$ field does not show a great spatial variability, thus weakening the model capability to clearly select the solution nodes with a small number of assimilated measurements and using a small number of assimilation times. Moreover, the elevation of the network nodes, which is rather constant in Net3 compared to the Anytown network, contributes to stress the mentioned effect, while data assimilation techniques based on the Kalman filter actually work on the differences between the cross correlation values of the involved variables. Moreover, Net3 network shows a high cross correlation between $\bar{H}$ and $\bar{Q}$ values. This feature, coupled with the high carrying capacity of the large diameter pipe extending throughout most of the system, contributes to smooth the spatial variation of the fields of variables H and Q, collected as measurements to be used in the assimilation procedure. A different situation occurs in Anytown, where the cross correlation between piezometric heads $\bar{H}$ is limited, thanks to the redundant looped structure of the network. For the same reason, the Q field in Anytown does not condition heavily the H field (relatively low cross correlation values between $\bar{H}$ and $\bar{Q}$).

The adoption of a different assimilation scheme, which uses a first data assimilation

cycle during which only measurements of volume V_{in} and V_{out} are included (scenarios b), does not lead to better results compared to those obtained with the classical assimilation scheme (scenarios a), in which all the measurements are included all together when available. The results were consistent, even if the errors in scenarios a are generally higher than the corresponding errors for scenarios b. This is due to the anticipated reduction of the ensemble variance in scenarios a, the number of assimilation times for scenarios a being twice the number of assimilation times for scenarios b.

In conclusion, the proposed approach demonstrates to be a promising tool for the spatial identification of leakages $\bar{L}$ in a WDS through the calibration of the C parameters values. The NS-EnKF performances are directly proportional to the topological complexity of the network analyzed and to the cross correlation between $\bar{H}$ and $\bar{L}$, and $\bar{Q}$ and $\bar{L}$. The model effectiveness for the cases here analyzed is inversely proportional to the cross correlation values between the $\bar{H}$ and $\bar{H}$ variables, and between the $\bar{H}$ and $\bar{Q}$ variables.

Chapter 6

Conclusions

In this thesis an inverse model, consisting in coupling Kalman Filter based data assimilation techniques to the network hydraulic model, is originally proposed and systematically applied to some synthetic cases. The aim is the assessment of the spatial distribution of water losses in WDSs through the calibration of the network parameters. The main purpose is to suggest a method to reduce the costs of field surveys currently required from the leakage detection activity on real systems, and to make the calibration of hydraulic network models, usually affected from the lack of knowledge on water losses, pipe roughness and water use, a more reliable procedure. The proposed model relies on the availability of pressure, flow rate and volume measurements data, which can be easily collected in real systems with limited efforts, with exception of the flow rate measurements, for which a relatively small number of monitoring positions is usually available.

The main results can be summarized as follows.

- With the aim of assessing the spatial distribution of water losses through the calibration of the network parameters, the Kalman Filter based techniques prove their effectiveness on some synthetic WDSs.

- In some cases a significant advantage can be derived from the application of the ES technique, that is when the system behaviour over time is not sufficiently described from the available information or, when the system knowledge is given only in terms of time averaged measurement data. Bearing in mind that in these cases the ES technique can be conveniently applied, the numerical experiments developed demonstrate the higher effectiveness of the EnKF compared to the ES, in retrieving the spatial distribution of the leakage parameters, when a fairly high number of measurements is available.

- The filter inbreeding problem, as it is reported in the literature relating to some

applications in the hydrological field, could affect the EnKF technique. However the model performances are not significantly affected and the results are always consistent in the cases here analyzed.

- Despite the fact that a high computational efficiency could be obtained by embedding the hydraulic transient model with the recursive EnKF procedure, the available softwares such as EpaNET are easier to use for the wide range of devices and situations that can be reproduced. A combination of valves and hydrants is conveniently adopted by using standard EpaNET to separately represent the pressure driven behaviour of both user demand and nodal leakage.

- In the cases here analyzed, the application of a Monte Carlo numerical approach, based on the criterion of cross correlation between the model parameters and the potentially measured system variables, leads to the selection of the measurement locations that turn out to be effective for the purpose of model calibration.

- The NS-EnKF procedure applied on the Anytown benchmark system leads to satisfactory results both in terms of leakage detection and variables estimation, even if a fairly high percentage of network elements needs to be monitored to calibrate the model. As the number of assimilated measurements increases, the gap between the estimated system state and the true solution closes progressively. Both these aspects are due to the structure of the Anytown network that is regularly looped. The NS-EnKF application on the Net3 network leads to good results as well, both in terms of parameters and variables. In this case a much smaller percentage of elements is required to be monitored in order to obtain a satisfactory model calibration. This is explained from the structure of the Net3 network which is more complex, that is less looped, than Anytown and shows a number of branches. On such network branches, a pressure drop due to an eventual leakage cannot be recovered as it happens in looped systems, thus determining the greater sensitivity of pressure values to the spatial distribution of the flow rates, that is a greater effectiveness of the coupled model here applied. The cross correlation relationship between water losses and potentially measured variables (pressure heads or pipe flow rates), that contributes to determine the coupled model efficiency, is indeed closely linked to the network topological structure.

Appendix A

Netan_HD examples

Example 1

The Anytown (Walski et al., 1987) literature network is analyzed, which is constituted by 20 nodes and 39 pipes. The source node has a fixed piezometric head of 3.05 m amsl. The system is fed through a pump inserted at element 40, whose head-flow rate curve is available.

The program stops the analysis at the demand driven case, the hydraulic head at each node being sufficient to fully satisfy the requested demand.

Input file input.inp

```
Anytown Todini pompa

                    NN  NT  NLN  NBN  EPS  NMAX  CPG
                    21  40  20   1    2    300   1

  IL PRIMO INDICE DA IL NUMERO DI COORDINATE DA LEGGERE:
  SE I=1 Nn, Z, IS, QK, XX, Hmin
  SE I=3 Nn, Z, X, Y, IS, QK, XX, Hmin
  1
  GRANDEZZE RIFERITE AI NODI

                Nnod  nID  quota  IS  QK       XX    Hmin
                1     n1   6.10   0   31.5451  0     6.10
                2     n2   15.24  0   12.6180  0     15.24
                3     n3   15.24  0   12.6180  0     15.24
                4     n4   15.24  0   37.8541  0     15.24
                5     n5   24.38  0   37.8541  0     24.38
                6     n6   24.38  0   37.8541  0     24.38
                7     n7   24.38  0   37.8541  0     24.38
                8     n8   24.38  0   25.2361  0     24.38
                9     n9   36.58  0   25.2361  0     36.58
                10    n10  36.58  0   25.2361  0     36.58
                11    n11  36.58  0   25.2361  0     36.58
                12    n12  15.24  0   31.5451  0     15.24
                13    n13  15.24  0   31.5451  0     15.24
                14    n14  15.24  0   31.5451  0     15.24
                15    n15  15.24  0   31.5451  0     15.24
```

16	n16	36.58	0	25.2361	0	36.58
17	n17	36.58	0	63.0902	0	36.58
18	n18	15.24	0	31.5451	0	15.24
19	n19	15.24	0	63.0902	0	15.24
20	n20	6.10	0	0.0000	0	6.10
21	n40	3.05	1	0.0000	3.05	3.05

GRANDEZZE RIFERITE AGLI ELEMENTI

Nel	pID	N1	N2	D	L	e	q
1	p1	n1	n2	304.7	3657.6	1.5	0
2	p2	n1	n12	762.9	3657.6	2.0	0
3	p3	n1	n13	699.6	3657.6	2.0	0
4	p4	n1	n20	457.2	30.5	1.0	0
5	p5	n2	n3	253.9	1828.8	1.5	0
6	p6	n2	n4	203.9	2743.2	1.5	0
7	p7	n2	n13	304.7	2743.2	2.0	0
8	p8	n2	n14	253.9	1828.8	1.5	0
9	p9	n3	n4	253.9	1828.8	1.5	0
10	p11	n4	n8	203.1	3657.6	1.5	0
11	p12	n4	n15	253.9	1828.8	1.5	0
12	p17	n8	n9	203.1	3657.6	1.5	0
13	p18	n8	n15	253.9	1828.8	1.5	0
14	p19	n8	n16	203.1	1828.8	1.5	0
15	p20	n8	n17	203.1	1828.8	1.5	0
16	p21	n9	n10	304.9	1828.8	1.5	0
17	p22	n10	n11	394.6	1828.8	1.5	0
18	p23	n10	n17	355.6	1828.8	1.5	0
19	p24	n11	n12	203.1	1828.8	1.5	0
20	p26	n12	n17	606.9	1828.8	1.5	0
21	p27	n12	n18	355.6	1828.8	2.0	0
22	p28	n13	n14	762.2	1828.8	2.0	0
23	p29	n13	n18	304.7	1828.8	2.0	0
24	p30	n13	n19	152.4	1828.8	2.0	0
25	p31	n14	n15	598.0	1828.8	2.0	0
26	p32	n14	n19	253.9	1828.8	2.0	0

27	p34	n15	n16	253.9	1828.8	2.0	0
28	p35	n15	n19	253.9	1828.8	2.0	0
29	p36	n16	n17	203.1	1828.8	1.5	0
30	p37	n16	n18	203.1	1828.8	2.0	0
31	p38	n16	n19	253.9	1828.8	2.0	0
32	p39	n17	n18	203.1	1828.8	1.5	0
33	p41	n18	n19	253.9	1828.8	2.0	0
34	p110	n4	n5	355.6	1828.8	1.0	0
35	p113	n5	n6	406.4	1828.8	1.0	0
36	p114	n6	n7	152.4	1828.8	1.0	0
37	p115	n6	n8	203.9	1828.8	1.0	0
38	p116	n7	n8	609.6	1828.8	1.0	0
39	p125	n11	n17	406.4	2743.2	1.0	0
40	pump	n40	n20	457.2	30.5	1.0	0

POMPE(IPP)

1

	Nel	Npti
	pump	5

Q1	Q2	H1	H2
0.0000	378.5412	91.4400	89.0016
378.5412	757.0824	89.0016	82.2960
757.0824	1135.6236	82.2960	70.1040
1135.6236	1514.1648	70.1040	55.1688

Output file echo.out

DATI DI RETE

NUMERO DI NODI:	21
NUMERO DI ELEMENTI:	40
NUMERO DI NODI UTILIZZATORI:	20
NUMERO DI NODI SORGENTE:	1
PRECISIONE DI CALCOLO:	2.00
MASSIMO NUMERO DI ITERAZIONI:	300

COEFFICIENTE DI PUNTA: 1.00

DATI DEI NODI

NODO	nID	Z[m]	QK[l/s]	IS	XX[m]	Hmin[m]
1	n1	6.1	31.55	0	0	6.10
2	n2	15.2	12.62	0	0	15.24
3	n3	15.2	12.62	0	0	15.24
4	n4	15.2	37.85	0	0	15.24
5	n5	24.4	37.85	0	0	24.38
6	n6	24.4	37.85	0	0	24.38
7	n7	24.4	37.85	0	0	24.38
8	n8	24.4	25.24	0	0	24.38
9	n9	36.6	25.24	0	0	36.58
10	n10	36.6	25.24	0	0	36.58
11	n11	36.6	25.24	0	0	36.58
12	n12	15.2	31.55	0	0	15.24
13	n13	15.2	31.55	0	0	15.24
14	n14	15.2	31.55	0	0	15.24
15	n15	15.2	31.55	0	0	15.24
16	n16	36.6	25.24	0	0	36.58
17	n17	36.6	63.09	0	0	36.58
18	n18	15.2	31.55	0	0	15.24
19	n19	15.2	63.09	0	0	15.24
20	n20	6.1	0.00	0	0	6.10
21	n40	3.1	0.00	1	3.05	3.05

DATI DEGLI ELEMENTI

Elem.	pID	N1ID	N2ID	D[mm]	L[m]	e[mm]	q[l/(sm)]
1	p1	n1	n2	304.7	3657.6	1.5	0
2	p2	n1	n12	762.9	3657.6	2.0	0
3	p3	n1	n13	699.6	3657.6	2.0	0
4	p4	n1	n20	457.2	30.5	1.0	0
5	p5	n2	n3	253.9	1828.8	1.5	0
6	p6	n2	n4	203.9	2743.2	1.5	0
7	p7	n2	n13	304.7	2743.2	2.0	0

8	p8	n2	n14	253.9	1828.8	1.5	0
9	p9	n3	n4	253.9	1828.8	1.5	0
10	p11	n4	n8	203.1	3657.6	1.5	0
11	p12	n4	n15	253.9	1828.8	1.5	0
12	p17	n8	n9	203.1	3657.6	1.5	0
13	p18	n8	n15	253.9	1828.8	1.5	0
14	p19	n8	n16	203.1	1828.8	1.5	0
15	p20	n8	n17	203.1	1828.8	1.5	0
16	p21	n9	n10	304.9	1828.8	1.5	0
17	p22	n10	n11	394.6	1828.8	1.5	0
18	p23	n10	n17	355.6	1828.8	1.5	0
19	p24	n11	n12	203.1	1828.8	1.5	0
20	p26	n12	n17	606.9	1828.8	1.5	0
21	p27	n12	n18	355.6	1828.8	2.0	0
22	p28	n13	n14	762.2	1828.8	2.0	0
23	p29	n13	n18	304.7	1828.8	2.0	0
24	p30	n13	n19	152.4	1828.8	2.0	0
25	p31	n14	n15	598.0	1828.8	2.0	0
26	p32	n14	n19	253.9	1828.8	2.0	0
27	p34	n15	n16	253.9	1828.8	2.0	0
28	p35	n15	n19	253.9	1828.8	2.0	0
29	p36	n16	n17	203.1	1828.8	1.5	0
30	p37	n16	n18	203.1	1828.8	2.0	0
31	p38	n16	n19	253.9	1828.8	2.0	0
32	p39	n17	n18	203.1	1828.8	1.5	0
33	p41	n18	n19	253.9	1828.8	2.0	0
34	p110	n4	n5	355.6	1828.8	1.0	0
35	p113	n5	n6	406.4	1828.8	1.0	0
36	p114	n6	n7	152.4	1828.8	1.0	0
37	p115	n6	n8	203.9	1828.8	1.0	0
38	p116	n7	n8	609.6	1828.8	1.0	0
39	p125	n11	n17	406.4	2743.2	1.0	0
40	pump	n40	n20	457.2	30.5	1.0	0

POMPE

NUMERO DI POMPE IN RETE: 1

Pompa 1 inserita nell elemento pump al tronco 40

Q1[l/s]	Q2[l/s]	H1[m]	H2[m]	Rp[1/(m^3/s)^n]	Hp[m]
0.0	378.5	91.44	89.00	-6.44	91.44
378.5	757.1	89.00	82.30	-17.71	95.71
757.1	1135.6	82.30	70.10	-32.21	106.68
1135.6	1514.2	70.10	55.17	-39.45	114.91

Portata totale richiesta dalla rete: 618.284 [l/s]

Output file risultati.out

Convergenza raggiunta al ciclo: 6

nID	QUC[l/s]	HN[m]	PP[m]
n1	31.55	86.64	80.54
n2	12.62	81.81	66.57
n3	12.62	75.83	60.59
n4	37.85	73.30	58.06
n5	37.85	71.34	46.96
n6	37.85	71.26	46.88
n7	37.85	74.49	50.11
n8	25.24	74.58	50.20
n9	25.24	79.32	42.74
n10	25.24	81.83	45.25
n11	25.24	82.13	45.55
n12	31.55	84.44	69.20
n13	31.55	83.01	67.77
n14	31.55	82.37	67.13
n15	31.55	81.27	66.03
n16	25.24	79.84	43.26
n17	63.09	82.92	46.34
n18	31.55	82.58	67.34
n19	63.09	80.16	64.92

n20	0.00	87.80	81.70
n40	-618.28	3.05	0.00

pID	QQ[l/s]	VV[m/s]	DDH[m]
p1	36.92	0.51	4.84
p2	271.73	0.59	2.20
p3	278.08	0.72	3.64
p4	-618.28	-3.77	-1.16
p5	35.89	0.71	5.98
p6	19.49	0.60	8.51
p7	-20.23	-0.28	-1.20
p8	-10.84	-0.21	-0.56
p9	23.27	0.46	2.53
p11	-6.38	-0.20	-1.28
p12	-41.49	-0.82	-7.98
p17	-12.41	-0.38	-4.74
p18	-37.99	-0.75	-6.69
p19	-18.57	-0.57	-5.26
p20	-23.43	-0.72	-8.34
p21	-37.65	-0.52	-2.50
p22	-25.60	-0.21	-0.30
p23	-37.29	-0.38	-1.10
p24	-12.25	-0.38	-2.31
p26	181.17	0.63	1.52
p27	46.77	0.47	1.86
p28	205.58	0.45	0.64
p29	14.70	0.20	0.43
p30	6.03	0.33	2.84
p31	142.41	0.51	1.10
p32	20.79	0.41	2.21
p34	16.70	0.33	1.43
p35	14.69	0.29	1.11
p36	-14.17	-0.44	-3.08
p37	-12.77	-0.39	-2.74
p38	-0.18	0.00	-0.32

p39	4.61	0.14	0.34
p41	21.77	0.43	2.42
p110	52.77	0.53	1.96
p113	14.92	0.12	0.08
p114	-7.16	-0.39	-3.23
p115	-15.77	-0.48	-3.32
p116	-45.02	-0.15	-0.09
p125	-38.59	-0.30	-0.79
pump	618.28	3.77	-84.75

Example 2

The second example refers to a test case reported from Todini (2003). The network is formed by 8 nodes and 14 elements. The source node has a fixed piezometric head of 140 m amsl. The system is gravity fed.

Since the hydraulic head at each node is insufficient to fully satisfy the requested demand, the program developed a pressure driven analysis.

Input file input.inp

```
Todini test

                    NN  NT  NLN  NBN  EPS  NMAX  CPG
                    8   14  7    1    4    300   1

  IL PRIMO INDICE DA IL NUMERO DI COORDINATE DA LEGGERE:
  SE I=1 Nn, Z, IS, QK, XX, Hmin
  SE I=3 Nn, Z, X, Y, IS, QK, XX, Hmin
  1
  GRANDEZZE RIFERITE AI NODI

                  Nnod  nID  quota  IS  QK      XX   Hmin
                  1     n1   140    1   0.00    140  140
                  2     n2   80     0   16.67   0    80
                  3     n3   90     0   16.67   0    90
                  4     n4   70     0   33.33   0    70
                  5     n5   80     0   250.00  0    80
                  6     n6   90     0   250.00  0    90
                  7     n7   90     0   166.67  0    90
                  8     n8   100    0   83.33   0    100

  GRANDEZZE RIFERITE AGLI ELEMENTI

                   Nel  pID  N1  N2  D    L     e     q
                   1    p1   n1  n2  400  1000  0.05  0
                   2    p2   n1  n3  500  1000  0.05  0
                   3    p3   n2  n3  200  1000  0.05  0
                   4    p4   n2  n4  300  1000  0.05  0
```

5	p5	n2	n5	200	1000	0.05	0
6	p6	n3	n5	200	1000	0.05	0
7	p7	n3	n6	300	1000	0.05	0
8	p8	n4	n5	150	1000	0.05	0
9	p9	n5	n6	150	1000	0.05	0
10	p10	n4	n7	250	1000	0.05	0
11	p11	n5	n7	150	1000	0.05	0
12	p12	n5	n8	150	1000	0.05	0
13	p13	n6	n8	200	1000	0.05	0
14	p14	n7	n8	150	1000	0.05	0

POMPE(IPP)

0

Output file echo.out

DATI DI RETE

NUMERO DI NODI:	8
NUMERO DI ELEMENTI:	14
NUMERO DI NODI UTILIZZATORI:	7
NUMERO DI NODI SORGENTE:	1
PRECISIONE DI CALCOLO:	4.00
MASSIMO NUMERO DI ITERAZIONI:	300
COEFFICIENTE DI PUNTA:	1.00

DATI DEI NODI

NODO	nID	Z[m]	QK[l/s]	IS	XX[m]	Hmin[m]
1	n1	140	0.00	1	140	140
2	n2	80	16.67	0	0	80
3	n3	90	16.67	0	0	90
4	n4	70	33.33	0	0	70
5	n5	80	250.00	0	0	80
6	n6	90	250.00	0	0	90
7	n7	90	166.67	0	0	90
8	n8	100	83.33	0	0	100

DATI DEGLI ELEMENTI

Elem.	pID	N1ID	N2ID	D[mm]	L[m]	e[mm]	q[l/(sm)]
1	p1	n1	n2	400	1000	0.05	0
2	p2	n1	n3	500	1000	0.05	0
3	p3	n2	n3	200	1000	0.05	0
4	p4	n2	n4	300	1000	0.05	0
5	p5	n2	n5	200	1000	0.05	0
6	p6	n3	n5	200	1000	0.05	0
7	p7	n3	n6	300	1000	0.05	0
8	p8	n4	n5	150	1000	0.05	0
9	p9	n5	n6	150	1000	0.05	0
10	p10	n4	n7	250	1000	0.05	0
11	p11	n5	n7	150	1000	0.05	0
12	p12	n5	n8	150	1000	0.05	0
13	p13	n6	n8	200	1000	0.05	0
14	p14	n7	n8	150	1000	0.05	0

POMPE

NUMERO DI POMPE IN RETE: 0

Portata totale richiesta dalla rete: 816.670 [l/s]

Output file risultati.out

Convergenza raggiunta al ciclo: 8

nID	QUC[l/s]	HN[m]	PP[m]
n1	-816.67	140.00	0.00
n2	16.67	127.90	47.90
n3	16.67	131.74	41.74
n4	33.33	102.29	32.29
n5	250.00	76.32	-3.68
n6	250.00	82.08	-7.92
n7	166.67	70.66	-19.34

n8	83.33	69.01	-30.99

pID	QQ[l/s]	VV[m/s]	DDH[m]
p1	330.48	2.63	12.10
p2	486.19	2.48	8.26
p3	-29.09	-0.93	-3.85
p4	229.24	3.24	25.60
p5	113.66	3.62	51.58
p6	117.95	3.75	55.43
p7	322.47	4.56	49.67
p8	37.35	2.11	25.98
p9	-16.84	-0.95	-5.76
p10	158.56	3.23	31.64
p11	16.68	0.94	5.66
p12	19.12	1.08	7.31
p13	55.63	1.77	13.07
p14	8.57	0.49	1.65

Convergenza raggiunta al ciclo: 14

nID	QUC[l/s]	HN[m]	PP[m]
n1	-757.83	140.00	0.00
n2	16.67	129.79	49.79
n3	16.67	132.72	42.72
n4	33.33	110.22	40.22
n5	344.73	80.00	0.00
n6	324.20	90.00	0.00
n7	125.72	90.00	0.00
n8	-103.49	100.00	0.00

pID	QQ[l/s]	VV[m/s]	DDH[m]
p1	302.57	2.41	10.21
p2	455.26	2.32	7.28
p3	-25.18	-0.80	-2.94
p4	199.47	2.82	19.57

p5	111.60	3.55	49.79
p6	114.95	3.66	52.72
p7	298.47	4.22	42.72
p8	40.43	2.29	30.22
p9	-22.59	-1.28	-10.00
p10	125.72	2.56	20.22
p11	-22.59	-1.28	-10.00
p12	-32.57	-1.84	-20.00
p13	-48.33	-1.54	-10.00
p14	-22.59	-1.28	-10.00

Convergenza raggiunta al ciclo: 14

nID	QUC[l/s]	HN[m]	PP[m]
n1	-715.56	140.00	0.00
n2	16.67	130.72	50.72
n3	16.67	133.55	43.55
n4	33.33	111.67	41.67
n5	250.00	91.14	11.14
n6	250.00	94.85	4.85
n7	148.89	90.00	0.00
n8	0.00	92.87	-7.13

pID	QQ[l/s]	VV[m/s]	DDH[m]
p1	287.84	2.29	9.28
p2	427.72	2.18	6.45
p3	-24.66	-0.78	-2.82
p4	196.70	2.78	19.05
p5	99.13	3.16	39.59
p6	102.73	3.27	42.41
p7	283.66	4.01	38.70
p8	33.03	1.87	20.54
p9	-13.30	-0.75	-3.71
p10	130.34	2.66	21.67
p11	6.98	0.40	1.14

p12	-8.80	-0.50	-1.73
p13	20.36	0.65	1.98
p14	-11.57	-0.65	-2.87

Appendix B

Network data

Node number	Elevation [m amsl]	Demand [l/s]	Node number	Elevation [m amsl]	Demand [l/s]
1	6.10	31.54	11	36.58	25.24
2	15.24	12.62	12	15.24	31.54
3	15.24	12.62	13	15.24	31.54
4	15.24	37.85	14	15.24	31.54
5	24.38	37.85	15	15.24	31.54
6	24.38	37.85	16	36.58	25.24
7	24.38	37.85	17	36.58	63.09
8	24.38	25.24	18	15.24	31.54
9	36.58	25.24	19	15.24	63.09
10	36.58	25.24	20	6.10	0.00

Table B.1: *Node elevation and nodal demand data for Anytown illustrative network.*

Pipe number	Starting node	Ending node	Length [m]	Diameter [mm]	Roughness [D-W]
1	1	2	3657.6	304.7	1.5
2	1	12	3657.6	762.9	2.0
3	1	13	3657.6	699.6	2.0
4	1	20	30.5	457.2	1.0
5	2	3	1828.8	253.9	1.5
6	2	4	2743.2	203.9	1.5
7	2	13	2743.2	304.7	2.0
8	2	14	1828.8	253.9	1.5
9	3	4	1828.8	253.9	1.5
11	4	8	3657.6	203.1	1.5
12	4	15	1828.8	253.9	1.5
17	8	9	3657.6	203.1	1.5
18	8	15	1828.8	253.9	1.5
19	8	16	1828.8	203.1	1.5
20	8	17	1828.8	203.1	1.5

Table continues on next page

Table continues from previous page

Pipe number	**Starting node**	**Ending node**	**Length [m]**	**Diameter [mm]**	**Roughness [D-W]**
21	9	10	1828.8	304.9	1.5
22	10	11	1828.8	394.6	1.5
23	10	17	1828.8	355.6	1.5
24	11	12	1828.8	203.1	1.5
26	12	17	1828.8	606.9	1.5
27	12	18	1828.8	355.6	2.0
28	13	14	1828.8	762.2	2.0
29	13	18	1828.8	304.7	2.0
30	13	19	1828.8	152.4	2.0
31	14	15	1828.8	598.0	2.0
32	14	19	1828.8	253.9	2.0
34	15	16	1828.8	253.9	2.0
35	15	19	1828.8	253.9	2.0
36	16	17	1828.8	203.1	1.5
37	16	18	1828.8	203.1	2.0
38	16	19	1828.8	253.9	2.0
39	17	18	1828.8	203.1	1.5
41	18	19	1828.8	253.9	2.0
110	4	5	1828.8	355.6	1.0
113	5	6	1828.8	406.4	1.0
114	6	7	1828.8	152.4	1.0
115	6	8	1828.8	203.9	1.0
116	7	8	1828.8	609.6	1.0
125	11	17	2743.2	406.4	1.0

Table B.2: *Pipe data for Anytown illustrative network.*

$Flow[ls^{-1}]$	$Head[m]$
0.0000	91.440
126.1804	89.002

Table continues on next page

Table continues from previous page

$Flow[ls^{-1}]$	$Head[m]$
252.3608	82.296
378.5412	70.104
504.7216	55.169

Table B.3: *Pump curve for Anytown.*

Node number	Elevation [m amsl]	Demand [l/s]	Node number	Elevation [m amsl]	Demand [l/s]
10	44.81	0.00	185	4.88	1.62
15	9.75	0.06	187	3.81	0.00
35	3.81	0.06	189	1.22	6.81
60	0.00	0.00	191	7.62	5.17
601	0.00	0.00	193	5.49	4.50
61	0.00	0.00	195	4.72	0.00
101	12.80	11.98	197	7.01	1.08
103	13.11	8.40	199	-0.61	7.53
105	8.69	8.54	201	0.03	2.81
107	6.71	3.45	203	0.61	0.06
109	6.19	14.60	204	6.40	0.00
111	3.05	8.96	205	6.40	4.12
113	0.61	1.26	206	0.30	0.00
115	4.27	3.29	207	2.74	4.38
117	4.15	7.43	208	4.88	0.00
119	0.61	11.11	209	-0.61	0.05
120	0.00	0.00	211	2.13	0.55
121	-0.61	2.63	213	2.13	0.88
123	3.35	0.06	215	2.13	5.82
125	3.35	2.88	217	1.83	1.53
127	17.07	1.11	219	1.22	2.61
129	15.54	0.00	225	2.44	1.44
131	1.83	2.70	229	3.20	4.05
139	9.45	0.37	231	1.52	1.04
141	1.22	0.62	237	4.27	0.98
143	-1.37	0.39	239	3.96	2.81
145	0.30	1.74	241	3.96	0.00
147	5.64	0.54	243	4.27	0.27
149	4.88	1.71	247	5.49	4.44
151	10.21	9.12	249	5.49	0.00
153	20.18	2.79	251	9.14	1.52
157	3.99	3.27	253	10.97	3.44

Table continues on next page

Table continues from previous page

Node number	Elevation [m amsl]	Demand [l/s]	Node number	Elevation [m amsl]	Demand [l/s]
159	1.83	2.61	255	8.23	2.55
161	1.22	1.00	257	5.18	0.00
163	1.52	0.59	259	7.62	0.00
164	1.52	0.00	261	0.00	0.00
166	-0.61	0.16	263	0.00	0.00
167	-1.52	0.92	265	0.00	0.00
169	-1.52	0.00	267	6.40	0.00
171	-1.22	2.48	269	0.00	0.00
173	-1.22	0.00	271	1.83	0.00
177	2.44	3.67	273	2.44	0.00
179	2.44	0.00	275	3.05	0.00
181	2.44	0.00	River	67.06	0.00
183	3.35	0.00	Lake	50.90	0.00
184	4.88	0.00			

Table B.4: *Node elevation and nodal demand data for Net3 illustrative network.*

Pipe number	Starting node	Ending node	Length [m]	Diameter [mm]	Roughness [H-W]
60	River	60	375.2	609.6	140
101	10	101	4328.2	457.2	110
103	101	103	411.5	406.4	130
105	101	105	774.2	304.8	130
107	105	107	448.1	304.8	130
109	103	109	1200.9	406.4	130
111	109	111	609.6	304.8	130
112	115	111	353.6	304.8	130
113	111	113	512.1	304.8	130
114	115	113	609.6	203.2	130

Table continues on next page

Table continues from previous page

Pipe number	Starting node	Ending node	Length [m]	Diameter [mm]	Roughness [H-W]
115	107	115	594.4	203.2	130
116	113	193	506.0	304.8	130
117	263	105	830.6	304.8	130
119	115	117	664.5	304.8	130
120	119	120	222.5	304.8	130
121	120	117	570.0	304.8	130
122	121	120	624.8	203.2	130
123	121	119	609.6	762.0	141
125	123	121	457.2	762.0	141
129	121	125	283.5	609.6	130
131	125	127	987.6	609.6	130
135	127	129	274.3	609.6	130
137	129	131	1975.1	406.4	130
145	129	139	838.2	203.2	130
147	139	141	624.8	203.2	130
149	143	141	426.7	203.2	130
151	15	143	502.9	203.2	130
153	145	141	1069.9	304.8	130
155	147	145	670.6	304.8	130
159	147	149	268.2	304.8	130
161	149	151	310.9	203.2	130
163	151	153	356.6	304.8	130
169	125	153	1389.9	203.2	130
171	119	151	1054.6	304.8	130
173	119	157	634.0	762.0	141
175	157	159	887.0	762.0	141
177	159	161	609.6	762.0	141
179	161	163	131.1	762.0	141
180	163	164	45.7	355.6	130
181	164	166	149.4	355.6	130
183	265	169	179.8	762.0	141

Table continues on next page

Table continues from previous page

Pipe number	Starting node	Ending node	Length [m]	Diameter [mm]	Roughness [H-W]
185	167	169	18.3	203.2	130
186	187	204	30.5	203.2	130
187	169	171	387.1	762.0	141
189	171	173	15.2	762.0	141
191	271	171	231.7	609.6	130
193	35	181	9.1	609.6	130
195	181	177	9.1	304.8	130
197	177	179	9.1	304.8	130
199	179	183	64.0	304.8	130
202	185	184	30.5	203.2	130
203	183	185	155.5	203.2	130
204	184	205	1380.7	304.8	130
205	204	185	403.9	304.8	130
207	189	183	411.5	304.8	130
209	189	187	152.4	203.2	130
211	169	269	196.9	304.8	130
213	191	187	780.3	304.8	130
215	267	189	374.9	304.8	130
217	191	193	158.5	304.8	130
219	193	195	109.7	304.8	130
221	161	195	701.0	203.2	130
223	197	191	350.5	304.8	130
225	111	197	850.4	304.8	130
229	173	199	1219.2	609.6	141
231	199	201	192.0	609.6	141
233	201	203	36.6	609.6	130
235	199	273	221.0	304.8	130
237	205	207	365.8	304.8	130
238	207	206	137.2	304.8	130
239	275	207	435.9	304.8	130
240	206	208	155.5	304.8	130

Table continues on next page

Table continues from previous page

Pipe number	Starting node	Ending node	Length [m]	Diameter [mm]	Roughness [H-W]
241	208	209	269.8	304.8	130
243	209	211	368.8	406.4	130
245	211	213	301.8	406.4	130
247	213	215	1306.1	406.4	130
249	215	217	506.0	406.4	130
251	217	219	624.8	355.6	130
257	217	225	475.5	304.8	130
261	213	229	670.6	203.2	130
263	229	231	597.4	304.8	130
269	211	237	634.0	304.8	130
271	237	229	240.8	203.2	130
273	237	239	155.5	304.8	130
275	239	241	10.7	304.8	130
277	241	243	670.6	304.8	130
281	241	247	135.6	254.0	130
283	239	249	131.1	304.8	130
285	247	249	3.1	304.8	130
287	247	255	423.7	254.0	130
291	255	253	335.3	254.0	130
293	255	251	335.3	203.2	130
295	249	251	442.0	304.8	130
297	120	257	196.6	203.2	130
299	257	259	106.7	203.2	130
301	259	263	426.7	203.2	130
303	257	261	426.7	203.2	130
305	117	261	196.6	304.8	130
307	261	263	106.7	304.8	130
309	265	267	481.6	203.2	130
311	193	267	356.6	304.8	130
313	269	189	196.9	304.8	130
315	181	271	79.3	609.6	130

Table continues on next page

Table continues from previous page

Pipe number	**Starting node**	**Ending node**	**Length [m]**	**Diameter [mm]**	**Roughness [H-W]**
317	273	275	679.7	203.2	130
319	273	205	196.6	304.8	130
321	163	265	365.8	762.0	141
323	201	275	91.4	304.8	130
325	269	271	393.2	203.2	130
329	61	123	13868.4	762.0	140
330	60	601	0.3	762.0	140
333	601	61	0.3	762.0	140

Table B.5: *Pipe data for Net3 illustrative network.*

$Flow[ls^{-1}]$	$Head[m]$
0.0000	31.6992
126.1804	28.0416
252.3608	19.2024

Table B.6: *Pump curve for Net3 lake source.*

$Flow[ls^{-1}]$	$Head[m]$
0.0000	60.9600
504.7216	42.0624
883.2628	26.2128

Table B.7: *Pump curve for Net3 river source.*

List of simbols

A	linear transition matrix
A_{10}	fixed head nodes incidence matrix
A_{12}	unknown head nodes incidence matrix
A_o	equivalent orifice area
b	boundary condition vector
c	forcing term vector
C	emitter coefficients vector
C_o	orifice coefficient
d	pipe diameter
e	pipe roughness
$\bar{e}$	mean pipe roughness
e_1	forecast model white noise
e_2	observation white noise
E	energy (level)
$f(H)$	law expressing nodal flow rate
$f(H, e)$	law expressing head losses in pipes
$f(Q)$	law expressing head losses in pipes
f_f	friction factor
F	transition matrix
g	gravity acceleration
h	pipe head loss
H	nodal heads vector
$\bar{H}$	mean pressure heads vector
H^*	critical pressure
H^m	measured pressure head
H_0	fixed nodal heads vector
H_p	pump head

H_s	service head
I	identity matrix
k	stochastic non-linear operator
K	Kalman gain matrix
l	pipe length
L	nodal leakage flow rates vector
$\bar{L}$	mean nodal leakage flow rates vector
m_1	first mode of the bimodal pdf
m_2	second mode of the bimodal pdf
M	observation operator
n	exponent of the head loss formula
N	diagonal matrix of the exponents
NMC	Monte Carlo ensemble dimension
$nmis$	number of collected measurements
nn	number of nodes
nl	number of energy paths
no	number of nodes with fixed head
$nobsH$	number of measured pressure heads
$nobsQ$	number of measured pipe flow rates
np	number of pipes
p	pressure head
P^f	forecast state error covariance matrix
P_e^f	forecast ensemble error covariance matrix
P^u	updated state error covariance matrix
q	nodal demands vector
$\widehat{q}$	maximum compatible nodal demand
q_r	required nodal demand
Q	pipe flow rates vector
$\bar{Q}$	mean pipe flow rates vector
Q^*	fictitious flow rate
Q^m	measured pipe flow rate
Q_e	pipe incoming flow rate
Q_u	pipe outgoing flow rate
r	pump characteristic parameter
R	observation noise covariance matrix

R_p	pump flow rate
Re	Reynolds number
s	element of the sensitivity matrix
S	sensitivity matrix
$\bar{S}$	normalized sensitivity matrix
t	assimilation time
t_F	last measurement collection time
u	deterministic non-linear operator
v	pump characteristic parameter
V_{in}	incoming flow volume
V_{in}^m	measured incoming flow volume
V_{out}	outgoing flow volume
V_{out}^m	measured outgoing flow volume
W	model noise covariance matrix
x	random vector
X	system state matrix
X_0	initial model state
X^f	forecast state estimate
X^{f*}	ES forecast state
X_{nst}^f	normal-score forecast state estimate
X^u	updated state estimate
X^{u*}	ES updated model state
X_{nst}^u	normal-score updated state estimate
z^m	observation vector
z_{nst}^m	normal-score observation vector
Z^m	observation matrix
Z^{m*}	ES observation matrix

α	pipe resistance factor
β	corrective factor due to the distributed withdrawals
Γ	minimization function
γ	pressure exponent
γ_w	water density
Δt	time step
λ	Lagrange multiplier
ν	kinematic water viscosity
σ	maximum element on the row of $\bar{S}$
Φ	probability density function
φ	mathematical model solution

Bibliography

AWWA (2008). *Water Audits and Loss Control Programs: M36*, volume 36. American Water Works Association.

Ayres, F. (1962). Schaum's outline of theory and problems of matrices.

Bailey, R. and Baù, D. (2010). Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution. *Water Resources Research*, 46(12).

Boulos, P. F. and Wood, D. J. (1990). Explicit calculation of pipe-network parameters. *Journal of Hydraulic Engineering*, 116(11):1329–1344.

Bragalli, C., Fortini, M., and Todini, E. (2016). Enhancing knowledge in water distribution networks via data assimilation. *Water Resources Management*, pages 1–18.

Brunone, B. (2007). *La ricerca delle perdite e la gestione delle reti di acquedotto*. Morlacchi Editore.

Bush, C. A. and Uber, J. G. (1998). Sampling design methods for water distribution model calibration. *Journal of Water Resources Planning and Management*, 124(6):334–344.

Casillas, M. V., Puig, V., Garza-Castanón, L. E., and Rosich, A. (2013). Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors*, 13(11):14984–15005.

Centre for Water Systems (2004). *Benchmark networks for design and optimization of water distribution networks*. University of Exeter, Exeter, UK. available from: http://emps.exeter.ac.uk/engineering/research/cws/resources/benchmarks/expansion/.

Chen, Y. and Zhang, D. (2006). Data assimilation for transient flow in geologic formations via ensemble kalman filter. *Advances in Water Resources*, 29(8):1107–1122.

Collins, M., Cooper, L., Helgason, R., Kennington, J., and LeBlanc, L. (1978). Solving the pipe network analysis problem using optimization techniques. *Management science*, 24(7):747–760.

Crestani, E., Camporese, M., Baú, D., and Salandin, P. (2013). Ensemble kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. *Hydrol. Earth Syst. Sci*, 17(4):1517–1531.

Cross, H. (1936). Analysis of flow in networks of conduits or conductors. *University of Illinois. Engineering Experiment Station. Bulletin; no. 286*.

Darvini, G., Salandin, P., and Da Deppo, L. (2008). Coping with uncertainty in the reliability evaluation of water distribution systems. In *Proceedings of the 10th Annual Water Distribution Systems Analysis Conference, WDSA 2008*, pages 483–498.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367.

Evensen, G. (2009a). *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media.

Evensen, G. (2009b). The ensemble kalman filter for combined state and parameter estimation. *IEEE Control Systems*, 29(3):83–104.

Evensen, G. and Van Leeuwen, P. J. (2000). An ensemble kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128(6):1852–1867.

Farmani, R., Walters, G. A., and Savic, D. A. (2005). Trade-off between total cost and reliability for anytown water distribution network. *Journal of Water Resources Planning and Management*, 131(3):161–171.

Gertler, J. (1998). *Fault detection and diagnosis in engineering systems*. CRC press.

Gupta, R. and Bhave, P. R. (1996). Comparison of methods for predicting deficient-network performance. *Journal of water resources planning and management*, 122(3):214–217.

Hendricks Franssen, H. and Kinzelbach, W. (2008). Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, 44(9).

Hunaidi, O., Chu, W., Wang, A., and Guan, W. (2000). Detecting leaks in plastic pipes. *American Water Works Association. Journal*, 92(2):82.

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L., and Savić, D. A. (2012). 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(2):169–183.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2005). Optimal sampling design methodologies for water distribution model calibration. *Journal of Hydraulic Engineering*, 131(3):190–200.

Kershaw, D. S. (1978). The incomplete cholesky—conjugate gradient method for the iterative solution of systems of linear equations. *Journal of Computational Physics*, 26(1):43–65.

Martin, D. and Peters, G. (1963). The application of newton's method to network analysis by digital computer. *Journal of the institute of Water Engineers*, 17(2):115–129.

McLaughlin, D. (2002). An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. *Advances in Water Resources*, 25(8):1275–1286.

Morosini, A. F., Veltri, P., Costanzo, F., and Savić, D. (2014). Identification of leakages by calibration of wds models. *Procedia Engineering*, 70:660–667.

Nowak, W. (2009). Best unbiased ensemble linearization and the quasi-linear kalman ensemble generator. *Water Resources Research*, 45(4).

Okeya, I., Kapelan, Z., Hutton, C., and Naga, D. (2014). Online modelling of water distribution system using data assimilation. *Procedia Engineering*, 70:1261–1270.

Pérez, R., Cugueró, M.-A., Cugueró, J., and Sanz, G. (2014). Accuracy assessment of leak localisation method depending on available measurements. *Procedia Engineering*, 70:1304–1313.

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 and technology: water supply*, 9(6):715–721.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., and Peralta, A. (2011). Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice*, 19(10):1157–1167.

Pilati, S. and Todini, E. (1984). La verifica delle reti idrauliche in pressione. *Istituto di Costruzioni Idrauliche, Facoltà di Ingegneria dell'Università di Bologna*.

Poulakis, Z., Valougeorgis, D., and Papadimitriou, C. (2003). Leakage detection in water pipe networks using a bayesian probabilistic framework. *Probabilistic Engineering Mechanics*, 18(4):315–327.

Pudar, R. S. and Liggett, J. A. (1992). Leaks in pipe networks. *Journal of Hydraulic Engineering*, 118(7):1031–1046.

Quevedo Casín, J. J., Cugueró Escofet, M. À., Pérez Magrané, R., Nejjari Akhi-Elarab, F., Puig Cayuela, V., and Mirats Tur, J. M. (2011). Leakage location in water distribution networks based on correlation measurement of pressure sensors. In *IWA Symposium on System Analysis and Integrated Assessment*, pages 290–297. Internatinal Water Association (IWA).

Ragot, J. and Maquin, D. (2006). Fault measurement detection in an urban water supply network. *Journal of Process Control*, 16(9):887–902.

Rossman, L. A. et al. (2000). Epanet 2: users manual.

Rougier, J. (2005). Probabilistic leak detection in pipelines using the mass imbalance approach. *Journal of Hydraulic Research*, 43(5):556–566.

Ruzza, V., Crestani, E., Darvini, G., and Salandin, P. (2014). Losses identification in uncalibrated water distribution networks: A case study. *Procedia Engineering*, 89:1545–1552. Proceedings of 16th Water Distribution System Analysis Conference, WDSA2014 — Urban Water Hydroinformatics and Strategic Planning.

Ruzza, V., Crestani, E., Darvini, G., and Salandin, P. (2015). Losses identification in water distribution networks through enkf and es. *Journal of Applied Water Engineering and Research*.

Ruzza, V. and Salandin, P. (2015). Localizzazione delle perdite nelle reti di distribuzione. In *Tecniche per la difesa dall'inquinamento. 36° corso*, volume 36, pages 1–26. Edibios.

Salandin, P. and Bertola, P. (1996). Analisi dell'affidabilità delle reti di distribuzione soggette a richiesta aleatoria. *Proceedings of the XXI Convegno di Idraulica e Costruzioni idrauliche*, pages 217–228.

Salgado, R., Todini, E., and O'Connell, P. (1988). Comparison of the gradient method with some traditional methods for the analysisof water supply distribution networks. In *Computer applications in water supply: vol. 1—systems analysis and simulation*, pages 38–62. Research Studies Press Ltd.

Shamir, U. and Howard, C. D. (1968). Water distribution systems analysis. *Journal of Hydraulic Division, vol. 94, No. HY1.*

Todini, E. (1979). Un metodo del gradiente per la verifica delle reti idrauliche. *Bollettino degli Ingegneri della Toscana*, 11:11–14.

Todini, E. (2000). Looped water distribution networks design using a resilience index based heuristic approach. *Urban water*, 2(2):115–122.

Todini, E. (2003). *A more realistic approach to the "extended period simulation" of water distribution networks*. Swets and Zeitlinger Publishers, Balkema, Lisse, The Netherlands.

Van Leeuwen, P. J. (2001). An ensemble smoother with error estimates. *Monthly weather review*, 129(4):709–728.

Van Leeuwen, P. J. and Evensen, G. (1996). Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, 124(12):2898–2913.

Walski, T. M., Brill Jr, E. D., Gessler, J., Goulter, I. C., Jeppson, R. M., Lansey, K., Lee, H.-L., Liebman, J. C., Mays, L., Morgan, D. R., et al. (1987). Battle of the network models: Epilogue. *Journal of Water Resources Planning and Management*, 113(2):191–203.

Warga, J. (1954). Determination of steady state flows and currents in a network. *Proceedings of the Instrument Society of America*, 9(Pt 5).

Wood, D. J. and Charles, C. O. (1972). Hydraulic network analysis using linear theory.

Xu, C. and Goulter, I. C. (1998). Probabilistic model for water distribution reliability. *Journal of Water Resources Planning and Management*, 124(4):218–228.

Zhou, H., Gomez-Hernandez, J. J., Franssen, H.-J. H., and Li, L. (2011). An approach to handling non-gaussianity of parameters and state variables in ensemble kalman filtering. *Advances in Water Resources*, 34(7):844–864.

Acknowledgements

I would like to thank my supervisor prof. Paolo Salandin for his knowledge and support of my Ph.D. studies. Undertaking this experience has been a challenge for me and I am thankful for the guidance and motivation I received from him.

My sincere thanks also goes to Dr. Elena Crestani for her helpfulness, knowledge and patience. Without her precious support it would not be possible to widen my research from various perspectives.

I am grateful to prof. Dragan Savic for the opportunity he provided me to work at the Centre for Water Systems at the University of Exeter. It has been a great experience and an occasion for my personal and professional growth.

Last but not least, I would like to thank prof. Bruno Brunone and prof. Maurizio Righetti for their comments to my thesis.