Chapter 4
Parameter Estimation: Definition and Sampling Design

Gerard Sanz and Ramon Pérez

4.1 Introduction

According to [48], calibration "consists of determining the physical and operational characteristics of an existing system and determining the data that when input to the computer model will yield realistic results". In [2], the authors used the word *verified* in place of *calibrated* but described a process of calibration: "System simulation is considered verified during preliminary analysis for design when calculated pressures are satisfactorily close to observed field gauge readings for given field source send-out and storage conditions. If simulation is not satisfactory, the possibility of local aberrations, such as open boundary valves, is investigated. In the absence of other expected causative factors, the assumed local arterial network loads are adjusted until computed and observed field pressures are within reasonable agreement for various levels and extremes of demand, pumping, and storage". Walski [51] proposed a more precise definition: "Calibration of a water distribution model is a two step process consisting of: (1) Comparison of pressures and flows predicted with observed pressures and flows for known operating conditions (i.e., pump operation, tank levels, pressure reducing valve settings); and (2) adjustment of the input data for the model to improve agreement between observed and predicted valves. A model is considered calibrated for a set of operating conditions and water uses if it can predict flows and pressures with reasonable agreement".

A high degree of interest in this topic has been shown by researchers [47], but it has been considerably less covered by practitioners. A number of questions have to be answered, such as: (1) What parameters can be calibrated with confidence? (2) What is the acceptable level of discretization of calibration parameters and what is the acceptable level of agreement between measurements and model outputs? (3)

G. Sanz · R. Pérez (✉)
Research Center "Supervision, Safety and Automatic Control" (CS2AC-UPC), Terrassa, Spain
e-mail: ramon.perez@upc.edu

How to parameterize the model when insufficient data are available? and (4) What objective function type to use?

Orsmbee [33] suggested a seven-step general calibration procedure as follows: (1) identification of the intended use of the model; (2) determination of initial estimates of the model parameters; (3) collection of calibration data; (4) evaluation of the model results; (5) macro-level calibration; (6) sensitivity analysis; and (7) micro-level calibration.

One of the most important issues in model calibration is the determination of the purpose of the model [53]. Seven possible purposes of a network model were identified as follows: pipe sizing for master planning, extended period simulations for planning studies, subdivision layout, rehabilitation studies, energy usage studies, water quality models and flushing programmes. In [50], a real system is modelled for daily pump scheduling and system expansion design to examine the impact of model purpose on the calibration process.

Battle of the water calibration networks is summarized in [34], the goal of this competition was to objectively compare the solutions of different approaches to the calibration of water distribution systems through application to real water distribution system. Interesting references have been extracted from this work and future work is well pointed:

- Due to the inherently *ill-posed or under-constrained* calibration problem in WDN, the solutions that provide a good match between measured and modelled data have to be validated with extra data.
- Uncertainty has to be included in the model parameters to explore the influence on the calibrated model outputs.
- Calibration size problem reduction is an important factor to be considered to avoid model overfitting, avoid unnecessary simulations or reducing the search space.
- Leakage data may be included in hydraulic calibration efforts because leakage directly affects nodal demand allocation and pump curve characterizations.
- The effect of different field data on model calibration should be investigated (use of flow and/or pressure measurements).

In [52], the author described the importance of good data collection. In [54], the same author classified data into three different degrees of usefulness:

- Good data are collected when there is sufficient head loss to draw valid conclusions about model calibration. It is necessary to have head loss in the system that is significantly greater than the error in measurement to avoid random adjustments [55].
- Bad data contain errors because of misread pressure gauge, incorrectly determined elevation of the pressure gauge or lack of information about which pumps were running when calibration data were collected. This type of data should be discarded.
- Useless data are collected when the head loss in the system is so low that head loss and velocity are of a similar order of magnitude as the errors in measurements. Such data can produce misleading models.

Ahmed et al. [1] developed a heuristic three-step procedure to assist in identifying the conditions under which useful data (good data) should be collected. The issue of data quality and quantity is closely related to that of sampling design, which will be addressed later in this chapter.

Formulas may assist the user in deciding whether to adjust roughness or water use and by how much [51]. They are based on fire flow test. To correct for inaccuracies in input data, it is necessary to first understand the sources of these inaccuracies. These can be grouped into several categories: (1) incorrect estimate of water use; (2) incorrect pipe carrying capacity; (3) incorrect head at constant head points (i.e., pumps, tanks, pressure reducing valves); or (4) poor representation of system in model (e.g., too many pipes removed in skeletonizing the system). The major source of error in simulation of contemporary performance will be in the assumed loadings distributions and their variations. On the other hand, [15] states: "the weakest piece of input information is not the assumed loadings condition, but the pipe friction factor". The certainties of a previous model must be stated so that the effort in calibration is in the good direction.

The most important uncertainty sources are demands and model simplifications [18], but uncertainty also originates from measurement errors, incorrect boundary conditions, inherent model structural errors or unknown status of valves [20, 55]. The calibration in this and next chapters focuses on demands due to their daily variability and continuous evolution depending generally on social and climate factors comparing to the more stable evolution of roughness.

The sensitivity matrix plays an important role in the solution of the direct/inverse problem [58], as well as in many of the methodologies developed in this book. Some of the existing general methods for the calculation of the sensitivity matrix are as follows [22]: (a) influence coefficient method (or perturbation method), (b) sensitivity equation method, (c) variational method (or adjoin method) and (d) automatic differentiation method.

The influence coefficient method uses the concept of parameter perturbation. At each simulation, one of the model parameters is perturbed [4], and the outputs measured. This method can be easily implemented, though computationally slow and relatively inaccurate when compared to other methods. $N + 1$ simulations are required, where N is the number of parameters in the model.

In the sensitivity equation method, a set of sensitivity equations are obtained by taking the partial derivatives with respect to each parameter in the governing equation and initial and boundary conditions. The same number of simulation runs as in the influence coefficient method is required. The method requires a solution of the forward problem (heads and flows) prior to the determination of unknown sensitivities. The calculated sensitivities are quite accurate [22].

The adjoin method computes relevant sensitivities once Lagrange multipliers are determined from a set of adjoin equations, which are derived from the basic WDN hydraulic model equations. This method also has a high accuracy and only requires N_s simulation runs, where N_s is the number of selected model's predicted variables.

The automatic differentiation method [19] is based on the differentiation of algorithms. Despite the good accuracy and computational performance, it produces a

lengthy and complex computer code and requires a large number of changes to the source code of the appropriate hydraulic model [10].

Finally, a matrix analysis of the WDN linearized model where only one simulation is required at each iteration is proposed in [11].

The calculation of sensitivity matrices can be computationally demanding, as each element in the network generates an extra row or column in the matrix.

4.1.1 Identifiability

The calibration problem is often ill-posed. The ill-posedness is generally characterized by the non-uniqueness of the identified parameters. The uniqueness problem in parameter estimation is intimately related to identifiability [58].

Observability and identifiability terms are sometimes confused. System observability determines if the state of a system, i.e., the system variables (head, flow), can be estimated. On the other hand, system identifiability resolves if the parameters of the system (consumptions, roughness coefficients) can be calibrated. In conclusion, observability refers to system state (dynamic variables) while identifiability refers to system parameters (assumed constant in a certain time horizon).

An important contribution to the solution of the observability problem was made by [25], who formulated necessary and sufficient conditions for observability in power system state estimation in terms of meter location and network topology. According to their analysis, a network is observable if and only if it contains a spanning tree of full rank. The same problem for water systems is formulated in [5]. The identifiability can be classified as static and dynamic [36]. In [9], the study of identifiability is performed for the static problem using graph analysis based on [35]. The idea is that some operations in graphs are equivalent to operation on equations.

Conditions of identifiability for nonlinear dynamic systems can be found in the literature. The state-space formulation by means of the dynamic information of the system can be used [56]. For the linear case, the invertibility of the matrix of the equations set was studied by [49].

The complexity of the transient equations in dynamic identifiability makes their use difficult for real networks. The extended period identifiability is based on quasi-static equations, which allows to use simpler equations related from one time step to the next one by tank equations. The extended period identifiability is based on the sensitivity matrix rank in both linear and nonlinear cases [36]. The author stated that if many measurements are taken in the same conditions they will not add any information (without increasing the rank of the sensitivity matrix) but could be useful for filtering the noise in the measurements.

In [46], identifiability of the calibration problem is assured by defining a set of demand components to be calibrated that, considering the available measurements, generates a full-rank sensitivity matrix. This new parameterization is suitable for any element that abounds in a complex system. Both nodal demands and roughness

of pipes are grouped [12, 26], and the hydraulic effects of roughness grouping are thoroughly studied in [30].

4.1.2 Sampling Design

Calibration accuracy should be judged both by the model's ability to reproduce data and by a quantitative measure of the uncertainty in calibrated parameter values. This uncertainty depends on the sampling design, including the measurement type, number, location, frequency and conditions existing at the time of sampling [8].

In the literature, the sampling design is defined as the procedure to determine the following [22]: (a) what WDN model predicted variables (pressures, flows, both, etc.) to observe; (b) where in the WDN to observe them; (c) when to observe (in terms of duration and frequency); and (d) under what conditions to observe.

In general, a sampling design may have one of the several purposes [29]: ambient monitoring, detection, compliance or research. Model calibration is considered research sampling, where the objective is to identify accurately the physical parameters of the system. A sampling design (SD) is a set of specified measurements, **y**, at particular locations and times, along with the experimental conditions under which measurements are made [8].

One of the first sampling designs [51] suggested: (a) monitor pressure near the high demand locations; (b) conduct fire flow tests on the perimeter of the skeletal distribution system, away from water sources; (c) use as large as possible test flows at the fire hydrant; and (d) collect both head and flow measurements.

The importance of sensitivity in inverse problems comes from two primary reasons [28]. First, the need for the measurements to be made at a location where they are sensitive to the desired calibration parameters. Second, the degree of confidence that one has in the result depends on the sensitivity. Different approaches for solving the optimization problem have been developed. Usually, the main objective of finding the best locations for sensors is combined with other objectives (i.e., devices' cost). Genetic algorithms (GA), sensitivity matrix analysis or heuristic methods are some of the methodologies used.

The meter placement problem becomes a multi-objective optimization by seeking the best solution in terms of estimation accuracy and metering cost [59]. In this last reference, the authors developed a method employing dynamic analysis of the covariance matrix of state variables and the decision trees technique.

The potential location of the sensors may be ranked according to their overall relative sensitivity of nodal heads with respect to roughness coefficients [16]. Three general sensitivity-based methods are proposed in [8], and they are derived from the D-optimality criterion to rank the locations and types of measurements for estimating the roughness coefficients of a WDN model using pressure measurements, tracer concentration measurements and a combination of both. The authors outlined that the proposed methods, although suboptimal, may have some advantages over purely statistical methods that lack a physical basis. These three sensitivity-based methods

are compared for selecting the worthwhile pressure and flow sensors' location in WDNs for calibrating roughness coefficients [14].

Pinzinger et al. [42] proposed three algorithms based on integer linear programming and greedy paradigm. The SD in [40] is formulated as an optimization problem which minimizes the influence of measurement errors in the state vector estimation subject to the constraint that the Jacobian matrix is of maximum rank. A greedy algorithm was used, which selected at each iteration the optimal location of the sensors. Some of the mentioned approaches used an iterative selection of the sensors, adding one sensor at each iteration to the set of already located ones. However, [22] demonstrated that the optimal set of locations for n monitoring points is not always a superset of the optimal set for $n - 1$ monitoring points.

Sensitivity-based heuristic sampling design procedure for WDS model calibration to identify preferable conditions for data collection is developed in [27] accounting for uncertainty in measurements and its impact on both model parameters and predictions.

Three sampling design approaches are proposed in [13]. The first two were based on the shortest path algorithm, and set sensors' locations depending on the distance between the source and the set of potential sensors nodes. The third approach solved the optimization problem based on maximization of Shannon's entropy, locating sensors in the nodes with highest pressure sensitivity on roughness changes. The sampling design cost was also taken into account.

GA can find the combination of fire flow test locations that, when analysed collectively, stresses the greatest percentage of the hydraulic network, so the roughness parameters of grouped pipes can be calibrated [31]. Multi-objective sensitivity-based methods for sampling design minimize both uncertainty and SD cost objectives [22, 24]. Model accuracy was maximized and formulated as the D-optimal criterion, the A-optimal criterion and the V-optimal criterion. SOGA/MOGA (single/multi-objective GA) were used and compared, leading to the conclusion that the advantages in MOGA outweigh its disadvantages. The Jacobian matrix used was calculated prior to the optimization model run by assuming the model parameter values. Opposed to this deterministic approach, this latter assumption is handled by introducing parameter uncertainty using some predefined probability density function [6]. Results in studied cases [23, 24] assessed that the calibration accuracy based on prediction uncertainty (V-optimality) is preferred over parameter uncertainty (D-optimal and A-optimal criteria). Similarly, D-optimality is preferred over A-optimality.

The sampling design is posed in [21] as a multi-objective optimization problem, where the objective functions represented demand estimation uncertainty, pressure prediction uncertainty and demand estimation accuracy. The optimization problem was solved using MOGA based on Pareto-optimal solutions.

Not all sampling design approaches are addressed to parameter calibration. The sampling design is often based on the model application, for example a leakage detection methodology [37]. One sensor was located at each iteration of the procedure with the objective of minimizing the maximum number of nodes with the same binary signature (which cannot be isolated separately). The pressure sensitivity matrix analysis and an exhaustive search strategy produce an optimal sensor

placement strategy [32]. Different sensor placement methodologies for demand calibration and leak detection are compared in [38]. Which performance criteria should be considered to place water quality and quantity sensors for both early detection and model calibration are investigated in [41].

4.2 Problem Statement

As explained above, the limited number of sensors together with the huge number of parameters requires a grouping of the parameters to make the calibration viable. In [44], the authors grouped demands depending on the type of user. Although good results were obtained with synthetic data, the analysis presented in [45] encourages the use of the demand components model.

The information extracted from the network depends on the type and location of the sensors. Each new sensor represents an additional equation in the system of equations to be solved. In order to have a determined system of equations, the number of measurements (sensors) has to be at least equal to the number of parameters, guaranteeing the system identifiability in the linear approximation.

In this chapter, a methodology both for the parameterization and the sampling design is pursuit. The questions to be answered are as follows:

- How can a huge number of parameters be grouped so that the system becomes identifiable with a reasonable number of field measurements?
- Where should these measurements be located so that a maximum of information is extracted for the calibration?
- Both questions use the information available in the sensitivity matrix.

4.3 Proposed Approach

The singular value decomposition (SVD) is a matrix decomposition method whereby a general $n_y \times n_x$ system matrix $\mathbf{A}$, relating model $\mathbf{x}$ and data $\mathbf{y}$:

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{y}, \tag{4.1}$$

is factored into

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{V}^T, \tag{4.2}$$

where $\mathbf{U}$ is a set of n_y orthonormal singular vectors that form a basis of the measured data vectorial space, $\mathbf{V}$ is a set of n_x orthonormal vectors that form a basis of the parameter vectorial space and $\mathbf{\Lambda}$ is an $n_y \times n_x$ diagonal matrix of singular values of $\mathbf{A}$, where the additional rows (more measurements than parameters) or columns (more parameters than measurements) are filled with zeros [3].

Fig. 4.1 λ singular values from the SVD of a 10×10 example sensitivity matrix
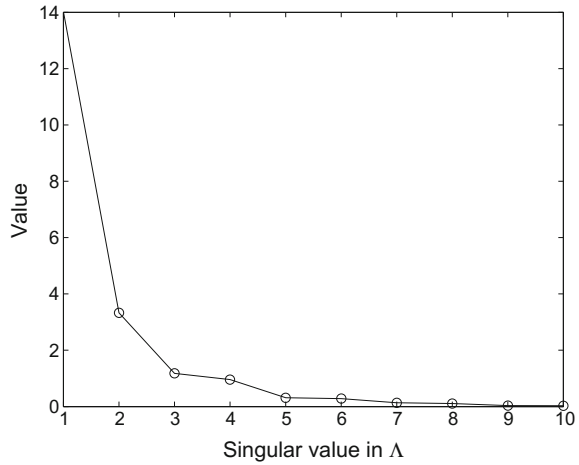

The SVD has many applications that can be useful for the parameter estimation. The key step to ensure the success of the calibration is the grouping of nodal demands into fewer parameters that, in the end, keep the network behaviour as close to the original behaviour as possible. This grouping ensures the identifiability of the system.

When calibrating parameters in nonlinear systems, the system matrix **A** in (4.1) is replaced by the system sensitivity matrix **S**, which relates changes in data with changes in parameters. Explanations from now on focus on the sensitivity matrix **S**.

The SVD allows to compute a reconstructed sensitivity matrix $\mathbf{S_r}$ from a subset of columns of **U** and **V**, ignoring the information from these matrices that correspond to low relevant singular values. The singular values in matrix Λ from a 10×10 example sensitivity matrix are depicted in Fig. 4.1. We can observe quite low singular values

Fig. 4.2 MSE of the reconstructed matrix **S** from a different number of **U** and **V** vectors
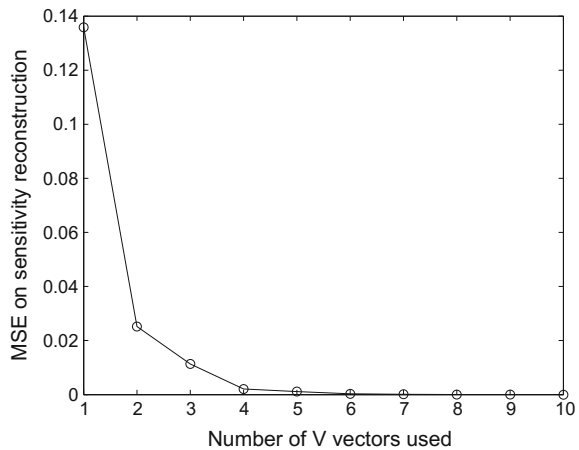
from the fifth position, indicating that the corresponding columns in matrices $\mathbf{U}$ and $\mathbf{V}$ have low importance in the reconstruction of matrix $\mathbf{S}$. Figure 4.2 presents the mean square error (MSE) in the reconstruction of the sensitivity matrix depending on the number of columns used from matrices $\mathbf{U}$ and $\mathbf{V}$, corresponding to the same number of singular values in $\mathbf{\Lambda}$. It can be seen that when considering only the four first columns, the MSE falls to a quite low value.

The reduction of matrices $\mathbf{U}$ and $\mathbf{V}$ is used in the methodology presented to choose which parameters will be calibrated and which sensors will be used in the calibration process.

4.3.1   Parameter Definition

The grouping of parameters can be obtained from the analysis of the SVD of the system sensitivity matrix. "We can think of the eigenvectors $\mathbf{v}_i$, where $i = 1, \ldots, n$, as a new parameterization of the model. These vectors represent a set of n linear combinations of the old parameters that are fixed by the observations" [57]. Similarly, it is possible to reduce matrix $\mathbf{V}$ into $\mathbf{V_r}$, which is formed by the first n_c vectors $\mathbf{v}_i$, where n_c is the number of nonzero singular values of the sensitivity matrix. The new parameterization is obtained by defining a new parameter correction as follows:

$$\mathbf{x}^* = \mathbf{V_r}^T \mathbf{x}. \tag{4.3}$$

In WDNs, very low singular values appear (as seen in (4.1)); thus, n_c is defined in a way that all values below the n_c highest singular values are neglected. Furthermore, the consideration of quite low singular values leads to an increase of uncertainty [3]. The main drawback of this approach is the loss of the physical meaning of the calibrated parameters as they will be generated by a linear combination of the old parameters at each iteration. The sensors' data will be fitted, but the calibrated parameters will not have a direct relation with the WDN.

Consequently, the objective is to define the new parameterization as a static combination of the old parameters. The resolution matrix $\mathbf{R}$, defined as

$$\mathbf{R} = \mathbf{V_r}\mathbf{V_r}^T, \tag{4.4}$$

describes how the generalized inverse solution smears out the original model $\mathbf{x}$ into a recovered model $\hat{\mathbf{x}}$. A perfect resolution is represented by the identity matrix, indicating that each parameter is perfectly resolved. When only n_c parameters corresponding to the highest n_c singular values are considered, the resolution matrix computed with $\mathbf{V_r}$ is not the identity matrix. Compact resolution appears, and parameters with similar sensitivities can be identified.

In the WDN particular case, compact resolution may appear but not being easily observable in the resolution matrix, as the parameter order in the sensitivity matrix $\mathbf{S}$ columns has no geographic order (in meshed networks, it is impossible to establish an

order). The identification can be performed by means of the "*delta vector generation*" process by [57], which is adapted to define the matrix $\mathbf{M}$ with the membership of each individual demand to each demand component. The resulting parameterization is used to calibrate groups of demands.

Algorithm 4.1 presents the whole process to generate the matrix $\mathbf{M}$ from the reduced matrix $\mathbf{V_r}$. In lines 1–7, the delta vector generation process is performed, where the n_c vectors with the highest resolving power in the resolution matrix are obtained and normalized iteratively to generate the delta vectors.

Algorithm 4.1 Computation of nodal demands memberships to demand components

---

Require: $\mathbf{V_r}, n_c, n_d$
1: Compute $\mathbf{R} = \mathbf{V_r}\mathbf{V_r}^T$
2: **for** $z = 1 : n_c$ **do**
3:     Find $j = \max(diag(\mathbf{R}))$
4:     Compute $\mathbf{v}_z^* = \mathbf{R}_{(:,j)}/\sqrt{\mathbf{R}_{(j,j)}}$
5:     Compute $\mathbf{R} = \mathbf{R} - \mathbf{v}_z^* \mathbf{v}_z^{*T}$
6: **end for**
7: Define $\mathbf{V}^* = [\mathbf{v}^*_1 \mid \mathbf{v}^*_2 \mid \cdots \mid \mathbf{v}^*_{n_c}]$
8: **for** $i = 1 : n_d$ **do**
9:     Compute $\mathbf{M}_{(i,:)} = |\mathbf{V}^*_{(i,:)}|/\sum |\mathbf{V}^*_{(i,:)}|$
10: **end for**
11: **return** : $\mathbf{M}$

In lines 8–11, matrix $\mathbf{V}^*$, which is formed by the $\mathbf{v}^*$ delta vectors, is used to generate the matrix $\mathbf{M}$, associating each initial parameter to a new parameter (component) that produces the best resolution if n_c components are considered. The normalization of the rows in $\mathbf{V}^*$ is done so that the weights can be interpreted as memberships of each element parameter to each parameter component.

Three approaches were studied in [46] before reaching the final procedure: binary parameterization, positive hybrid parameterization and free hybrid parameterization

- The first approach assigns a single parameter component to each element parameter. After executing lines 1–7 in Algorithm 4.1, each demand is associated with the parameter component that has the highest value in the corresponding columns of the matrix $\mathbf{V}^*$.
- The second approach assigns a combination of demand components to each nodal demand with positive weights, exactly as presented in Algorithm 4.1.
- The free hybrid parameterization considers a combination of demand components that can include negative weights. For this approach, the absolute value in the numerator of line 9 of Algorithm 4.1 is ignored.

In all the proposed approaches, the solution tends to generate geographical patterns, as the topological information (incidence matrix $\mathbf{B}$) is included in the sensitivity matrix. Results obtained in [46] concluded that the use of positive weights to

perform the calibration of parameter components gave the best results in terms of error minimization.

4.3.2   Sampling Design

The sampling design is performed after the distribution of components, selecting the n_c best sensors. The process for locating the sensors uses matrix $\mathbf{U}$ in the same way as the parameterization process uses matrix $\mathbf{V}$. Initially, the sensitivity matrix $\mathbf{S}^*$ relating head and/or flow variations with demand components variations is computed and decomposed using the SVD. Matrix $\mathbf{U_r}$ is constructed with the first n_c columns of $\mathbf{U}$, as the information from the subsequent columns is negligible (they are multiplied by null rows of the $\mathbf{\Lambda}$ matrix). Then, the information density matrix $\mathbf{I_d}$ is computed as explained in [3], i.e.,

$$\mathbf{I_d} = \mathbf{U_r U_r}^T, \tag{4.5}$$

describes how the generalized inverse solution smears out the original data $\mathbf{y}$ into a predicted data $\hat{\mathbf{y}}$. Since $\mathbf{I_d}$ has been constructed from n_c orthonormal vectors in $\mathbf{U_r}$, a set of n_c orthonormal vectors can be extracted from $\mathbf{I_d}$ in a way that they enhance the delta-like behaviour of the $\mathbf{I_d}$ matrix [57]. This "*delta-like*" vector generation process is presented in Algorithm 4.2 (lines 1–6). This process results in a set of delta-like vectors $\mathbf{u}^*$ that form matrix $\mathbf{U}^*$. Subsequently, the rows of matrix $\mathbf{U}^*$ are normalized (line 7), so that sensors with high sensitivity to multiple parameters are not selected. Finally, the sensor with the highest value in each of the n_c columns is selected as the sensor with highest information density to calibrate a particular parameter (lines 9–11). In the end, $n_s = n_c$ sensors are selected.

Algorithm 4.2 Sensor selection process

  Require: $\mathbf{U_r}, n_s$
  1: Compute $\mathbf{I_d} = \mathbf{U_r U_r}^T$
  2: **for** $z = 1 : n_s$ **do**
  3:    Find $j = \max(diag(\mathbf{I_d}))$
  4:    Compute $\mathbf{u}_z^* = \mathbf{I_d}_{(:,j)}/\sqrt{\mathbf{I_d}_{(j,j)}}$
  5:    Compute $\mathbf{I_d} = \mathbf{I_d} - \mathbf{u}_z^* \mathbf{u}_z^{*T}$
  6: **end for**
  7: Define $\mathbf{U}^* = [\mathbf{u}^*_1 \mid \mathbf{u}^*_2 \mid \cdots \mid \mathbf{u}^*_{n_s}]$
  8: Normalize rows of $\mathbf{U}^*$
  9: **for** $z = 1 : n_s$ **do**
  10:   Find $\mathbf{s}_{(z)} = \max(\mathbf{U}^*_{(:,z)})$
  11: **end for**
  12: **return** : $\mathbf{s}$

4.4 Simulations and Results

Following the idea of Chap. 3, first an academic example is used for illustrating the methodology, and afterwards, it is applied to a real network. Both examples will be used in the next chapter where the calibration problem is solved utilizing the results presented here.

4.4.1 Exemplification

The methodology presented above will be illustrated with the dummy networks presented in Figs. 4.3 and 4.4, which represent a meshed network and a tree-like network, respectively, where demands have to be calibrated. The simplicity of the networks will be useful to exemplify the methodology at each step.

Figures 4.5 and 4.6 show the output of the delta vector, $\boldsymbol{v}^*$, generation process (subfigure a)), and the memberships obtained after the normalization performed in lines 8–11 of Algorithm 4.1 (subfigure b)). Three sensors are considered, and therefore, three components are generated. The memberships represent the modulation of each nodal demand by each component and are produced from the delta vectors' directions.

Figures 4.7 and 4.8 depict in each of their subfigures, the memberships of each demand node to a particular demand component. The darker the colour in the map, the higher the membership to the depicted demand component.
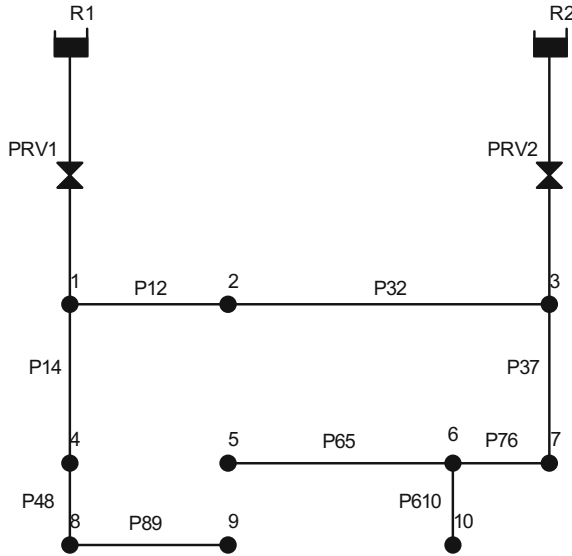

Fig. 4.3 Dummy meshed network
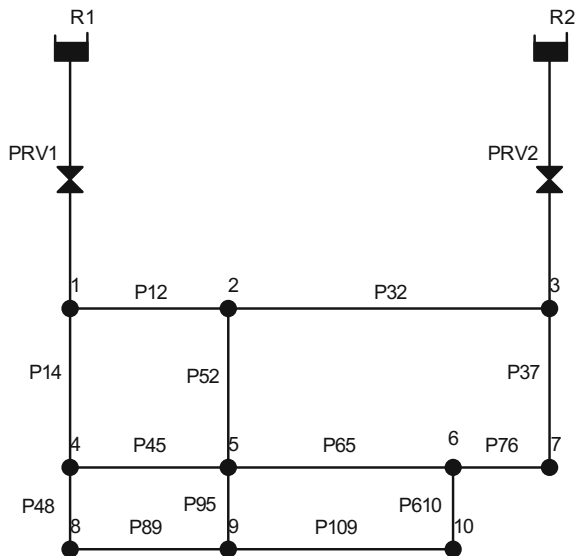
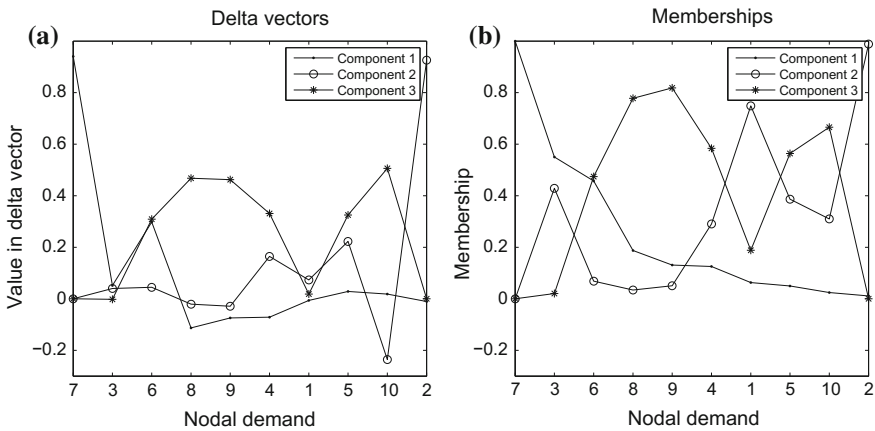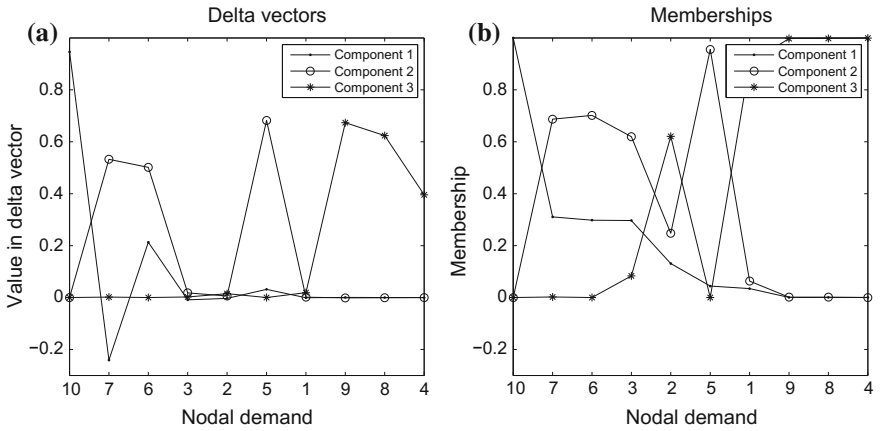**Fig. 4.4** Dummy tree network

Fig. 4.5 Parameterization process applied to a meshed network: **a** Delta vectors and **b** memberships of each nodal demand to each demand component

Algorithm 4.1 uses the sensitivity matrix computed at a particular working point. The procedure can be applied considering multiple boundary conditions to make the membership definition process more robust. However, the static topology of the network is not expected to produce significant changes in the sensitivity matrix. The application of the same process using other working points for the dummy networks generates the same memberships with only ±1% variations in the memberships.
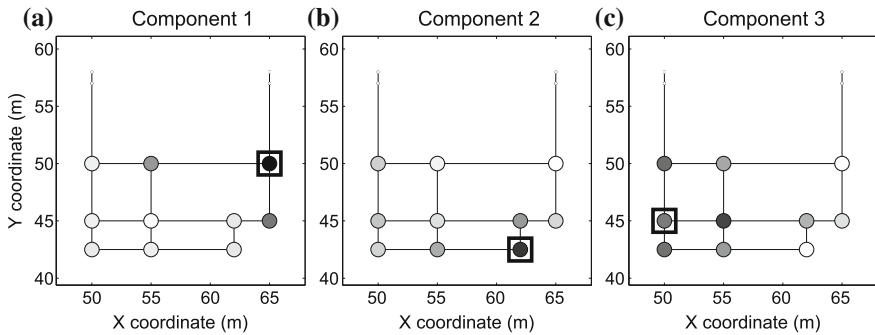
Fig. 4.6 Parameterization process applied to a tree-like network: **a** Delta vectors and **b** memberships of each nodal demand to each demand component
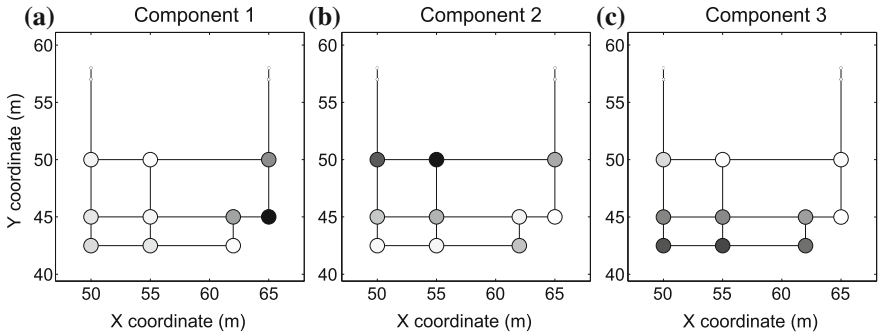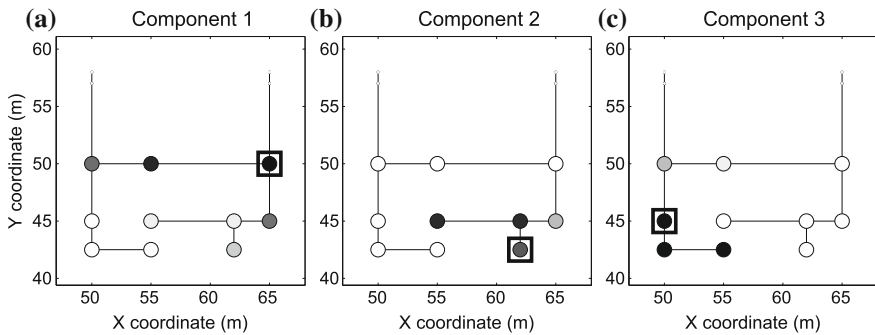

Fig. 4.7 Graphical representation of the nodal memberships to demand components in a meshed network
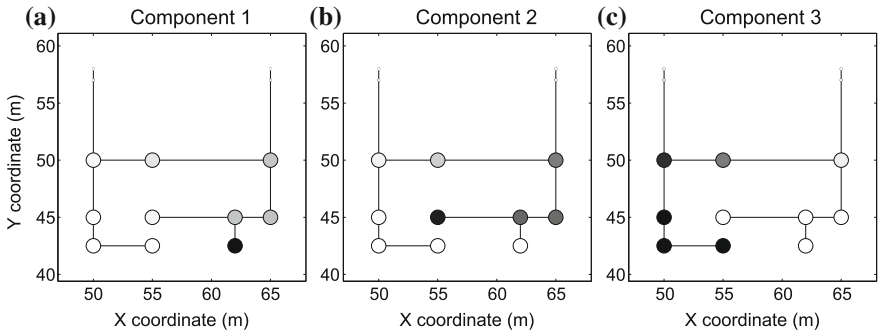
Fig. 4.8 Graphical representation of the nodal memberships to demand components in a tree-like network

Fig. 4.9 Graphical representation of the nodal memberships to demand components in a meshed network considering three installed sensors


Fig. 4.10 Graphical representation of the nodal memberships to demand components in a tree-like network considering three installed sensors

The calibration methodology requires some inner sensors to be distributed through the sampling design. In case the network already has the sensors installed, the **S** matrix introduced in Algorithm 4.1 would be a reduced sensitivity matrix $\mathbf{S_r}$ where only the rows related to the available sensors would be considered. Figures 4.9 and 4.10 depict the parameterization of the two dummy networks considering that three sensors were already installed in the networks. These sensors are marked with a black square.

Figures 4.11 and 4.12 depict the final results of the parameterization and sampling design process. The sensor selection has been performed after the definition of the demand components (results from the previous section).

4.4.2 Demand Components' Model for a Real Network

In Chap. 3, it has been seen that the most widely used demand models are the basic demand model and the demand patterns' model. The basic demand model
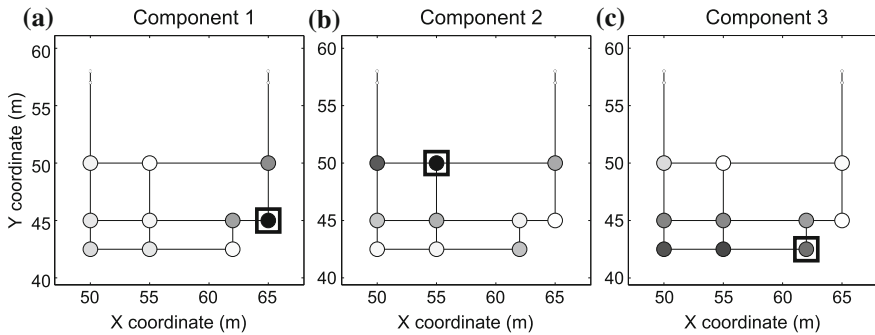
Fig. 4.11 Sensor selection results applied to a meshed network with three demand components
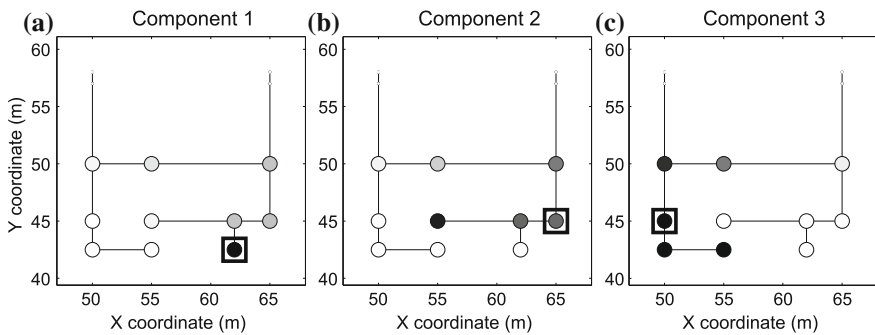
Fig. 4.12 Sensor selection results applied to a tree-like network with three demand components

cannot explain the daily variation of the relative pressure behaviour between two areas in the network, as it fixes the same behaviour to all demands. On the other hand, the demand patterns' model requires a lot of information that is not usually available (users associated with a given node, type of users) or does not fulfils the assumptions (incorrect predetermined diurnal demand patterns' values, users of the same type behaving differently). An example of the latter is presented in Fig. 4.13: automatic metre readings from two different segments (i.e., types of users) from a real network (Nova Icària) presented in Chap. 2 have been analysed. Each reading consists of the daily water consumption of a specific user, metered hourly. The correlation between every pair of readings within the same segment has been computed to assess the distance between their profiles, i.e., the similarity or dissimilarity of the users' behaviours. In each subfigure, the x-axis presents the users' telemetries, and each dot in the y-axis indicates the correlation between the user and all the other users in the same segment: the higher the correlation, the higher the similarity with its own segment's profiles.

Figure 4.13a presents a type of user with no similarity between its members, whereas Fig. 4.13b shows a type of user with more similarity between its members, but not enough to assume that all of them behave in the same way. In conclusion, the
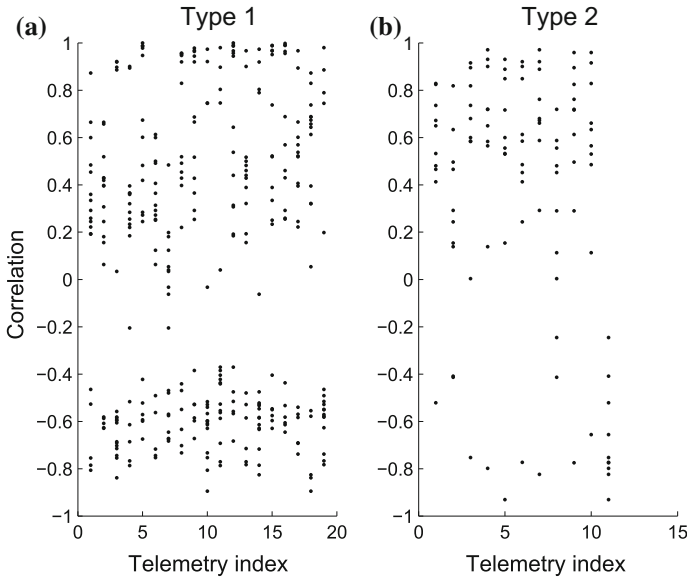
**Fig. 4.13** Cross-correlations of Nova Icària DMA telemetries in users of segments **a** 1 and **b** 2

assumption of considering that all users of the same type behave in the same way
can lead to incorrect results or high uncertainty in the calibrated parameters.

A new approach to model demands depending on their geographical location is
presented, and their sensitivity to hydraulic variables. Initially, nodes in a specific
zone of the network were assigned to a specific behaviour, which from now on will
be called demand component. This produces a new model

$$d_i(t) = \frac{bd_i}{\sum_{k=1}^{n_d} bd_k} \, c_{j \to i}(t) \, q^{in}(t), \tag{4.6}$$

where $c_{j \to i}(t)$ is the value of the demand component j associated with node i depend-
ing on the node location. Demand components are calibrated demand multipliers that
represent the behaviour of nodes in a determined geographical zone, avoiding the
dependency on information of the user type and diurnal pattern behaviour. All nodes
in the same area of node i have the same associated demand component. Conse-
quently, all nodes in the same zone will have the same demand behaviour, weighted
depending on their base demand. This demand model is capable of generating pres-
sure variations in different zones of the network, as it happens in a real situation.
Figure 4.14 presents a network where three demand components have been defined.
Each subplot presents the set of nodes that are modulated by the same demand com-
ponent according to (4.6).

However, the assumption that all nodes in the same area behave exactly in the
same way is not realistic. For example, a node in the limit of the effect zone of two

(a) Component 1   **(b)** Component 2   **(c)** Component 3
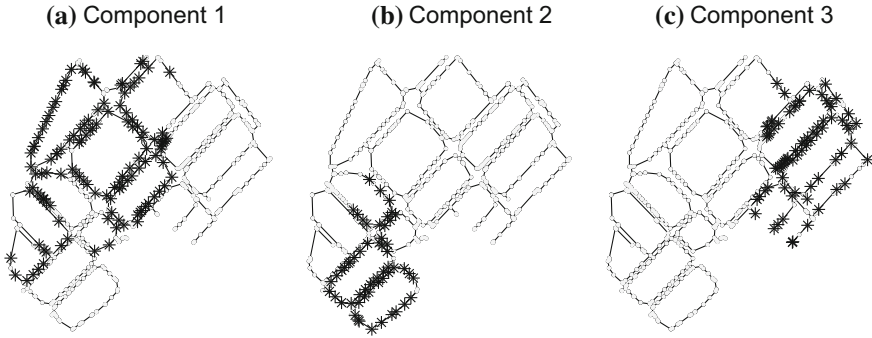
Fig. 4.14 Example of demand components with binary memberships

demand components should probably have a combination of the behaviour of the two demand components, instead of only one. To solve that, it is possible to redefine the demand model in (4.6), so that the degree to which each demand component is associated with each node is given as a membership, which depends on the nodes' geographical location. Thus, (4.7) represents the new demand model which can be written as follows:

$$d_i(t) = \frac{bd_i}{\sum_{k=1}^{n_d} bd_k} \, q^{in}(t) \, (\alpha_{i,1} \, c_1(t) + \alpha_{i,2} \, c_2(t) + \cdots + \alpha_{i,n_c} \, c_{n_c}(t)), \qquad (4.7)$$

with

$$\alpha_{i,1} + \alpha_{i,2} + \cdots + \alpha_{i,n_c} = 1, \quad \forall i,$$

where $\alpha_{i,j}$ is the association of demand component j with node i, and n_c is the number of demand components. The membership $\alpha_{i,j}$ of each node to each demand component depends on the geographical location of the node and is computed by means of the sensitivity analysis presented in Sect. 4.3.1. The model in 4.7 can generate different behaviours in every demand, while only having to calibrate few (n_c) demand components.

This way of calibrating demands incorporates the usually ignored fact that demands depend in some ways of head status of the network [17]. For example, if the pressure in a specific zone of the DMA decreases, the calibration process will estimate demand component values that decrease the consumption of nodes in that zone. Demand components presented in this chapter should not be confused with the ones defined in [17], where demand components were generated with a previous knowledge of the use of water (human-based, volume-based, non-controlled orifice-based, leakage-based).

The calibrated demand components generate individual demands that may not be exactly as the real ones, but the aggregated demand in a zone at a specific sample and the cumulative demand of each individual node during a period of time (similar to the billing) should coincide with the real ones if other parameters (roughness, valve status, etc.) are well calibrated.

Figure 4.15 presents the nodes' memberships to three demand components defined in the network in Fig. 4.14. The first component is located on the north-west side of the DMA; the second component is located on the south-west of the DMA; and the third component is located on the east side of the network. The nodes' memberships are depicted in greyscale: the darker the colour of a node, the higher the membership of that node to the demand component. Table 4.1 contains the memberships of the two nodes highlighted in Fig. 4.15. Demand of node A is modulated (60%) by the value of demand component 1, while component 3 has a lower (35%) effect on it. On the other hand, demand of node B is completely (98%) modulated by demand component 3. Demand component 2 does not have any effect on both demands, as it is far (geographically and hydraulically) from the two example nodes. Note the similarity between binary demand components (Fig. 4.14) and hybrid demand components (Fig. 4.15).

A comparison of the calibration results between type of user-based demand patterns and pressure sensitivity-based demand components is presented in [43], with better results for the latter: the uncertainty in the calibrated parameters is reduced, while the geographical distribution is useful for applications requiring parameters to be related with zones of the network.

(a) Component 1   **(b)** Component 2   **(c)** Component 3
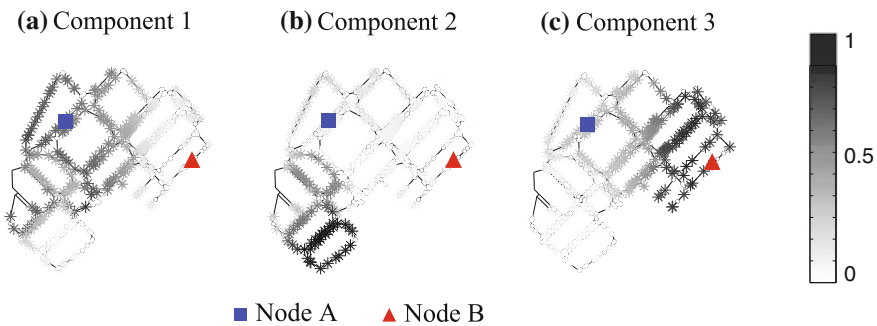
■ Node A   ▲ Node B

Fig. 4.15 Example of demand components and memberships in a network

Table 4.1 Memberships of nodes A and B of the example network

| Node	A	B
Membership to c_1	0.6	0.01
Membership to c_2	0.05	0.01
Membership to c_3	0.35	0.98

4.4.2.1 Sampling Design

The parameterization and sampling design processes were performed to propose the location of three pressure sensors and the parameter definition for demand calibration, as explained in Sects. 4.3.1 and 4.3.2. However, the proposed sensors' locations differ from the final ones, which have been obtained from the methodology in [7] (based on leak detection), developed by Cetaqua (Water Technology Centre of Aguas de Barcelona and the Suez group). The installed sensors can still be used to calibrate demands by defining the demand components depending on the available sensors' locations, thanks to the versatility of the proposed method (Sect. 4.3.1). Figure 4.16 depicts the proposed sensors' locations with circles and the final locations with stars.

The resolution of the sensors is 0.1 mwc (meters of water column), and the sampling times are defined in Table 3.2.

4.4.2.2 Data Analysis

Data from 9 March 2015 to 13 March 2015 (Monday–Friday) are used for the calibration process. Data from the following week, 16 March 2015 to 20 March 2015 (Monday–Friday), are used to validate and analyse the calibrated demand components. Previously, weekdays from 3 March 2015 to 6 March 2015 (Tuesday–Friday) are used to analyse and correct the data coming from the network, and to perform the parameterization process before the calibration starts. 2 March 2015 (Monday) is not used due to missing data. These three weeks will be referred, in current and next chapters, as precalibration week, calibration week and validation week. Weekends

Fig. 4.16 EPANET network model of Canyars sector with highlighted sensor locations. The network water input is signalled with a *triangle*, the installed pressure sensors are signalled with *stars* and the proposed pressure sensors are signalled with *circles*. The flow sensor is installed at the input pressure reduction valve, so that the total flow consumed in the network is known
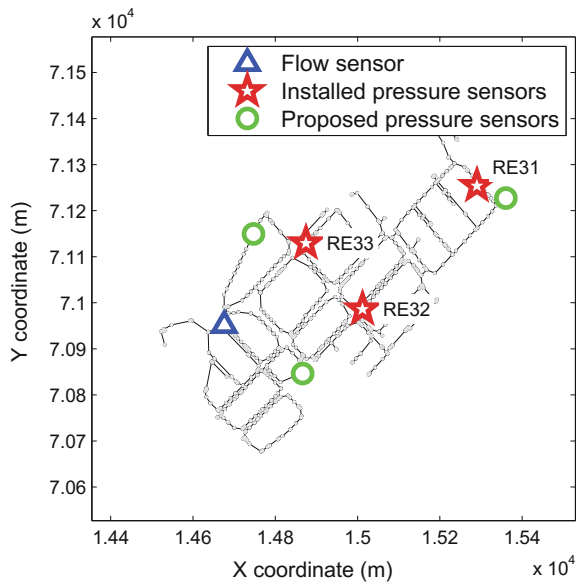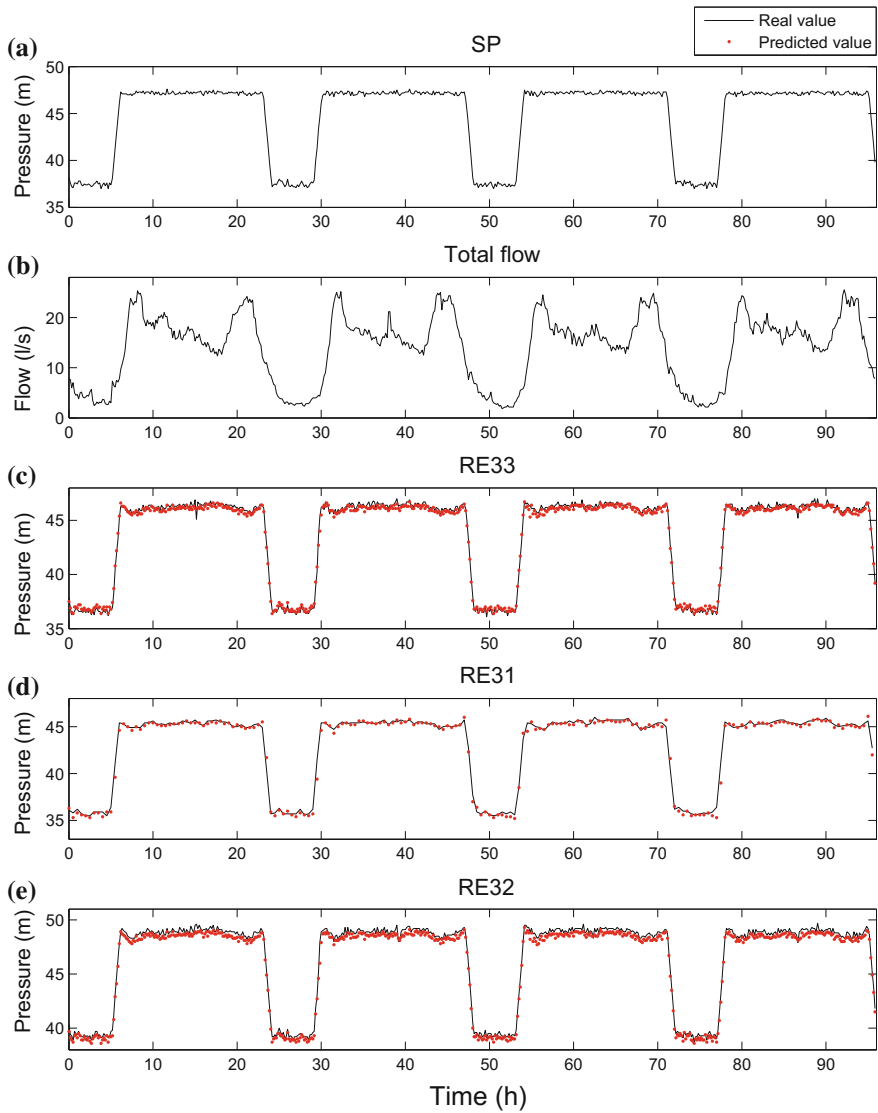
Fig. 4.17 Canyars network real and predicted data from 3 March 2015 to 6 March 2015 (precalibration week). *Black lines* and *red dots* refer to real and predicted data, respectively

are not considered in this case study but would follow the same calibration procedure as weekdays.

Figure 4.17 shows the complete set of data from the precalibration week, including boundary conditions (input valve's pressure set point (SP) and total flow) and the three pressure measurements. Black lines and red dots refer to real and predicted data, respectively. Predicted data have been obtained from simulating the network

model with the given boundary conditions using the basic demand model presented in
Chap. 3. Figure 4.17a shows the aforementioned pressure control at the DMA input.

Figure 3.3 shows the pressure prediction error in the three available sensors when
using the basic demand model. The blue thin line corresponds to the raw error using
all data, and the red thick line represents the smoothed error, which has been com-
puted by means of a smoothing spline. The green dashed line corresponds to the mean
pressure prediction error. This error is treated as an offset that cannot be associated
with the demand model. As suggested in [39], the offset is corrected to eliminate
possible depths errors, model nodes' elevations inaccuracies or badly calibrated sen-
sors' offsets. The same correction in each sensor is also considered when using data
from the calibration and validation weeks. Table 3.2 contains the specific correction
for each sensor.

4.4.2.3   Parameterization

Data from the precalibration week are used to compute the sensitivity matrices to
perform the parameterization process. The memberships of each nodal demand to
three demand components are computed using Algorithm 4.1, considering the three
installed sensors. Figure 4.18 depicts, in each of the network maps, the membership
of each node to a particular demand component: the darker the node, the higher the
membership to that component. Each map in Fig. 4.18 also includes the location of
the sensor with the highest sensitivity to the component drawn.

The average percentage of consumption $\overline{d}_{c_j}$ of demand component j is computed
from the billing information (nodal base demands **BDM**) and the recently computed
memberships (**M**) as

$$\overline{d}_{c_j} = 100 \sum \mathbf{BDM}\,\mathbf{M}_{(:,j)}. \tag{4.8}$$
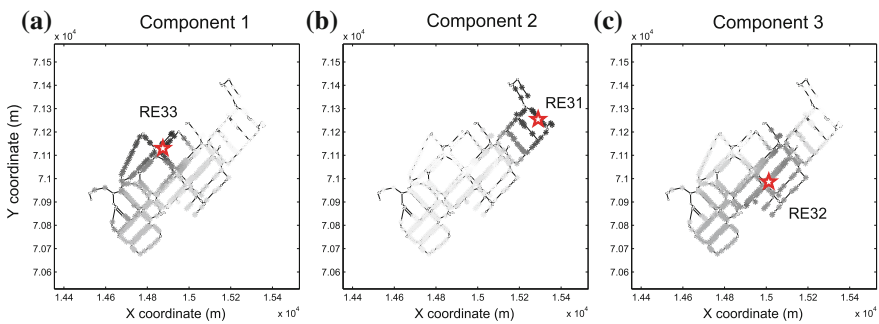
Fig. 4.18 Memberships of nodes to each demand component in Canyars network considering
the three available sensors. Each representation of the network depicts a *greyscale* map with the
membership of each node to a particular demand component: the *darker* the node in the map, the
higher the membership of the node to the demand component. The sensor with the highest sensitivity
to variations in each demand component is also depicted in each map

Table 4.2 Average percentage of demand components' water consumption in Canyars network computed from billing

Demand component	c_1	c_2	c_3
Average percentage of water consumption	40.2%	18.6%	41.2%

Table 4.2 sums up the average percentage of consumption of each demand component: demand component c_2 has the lowest percentage of consumption (18.6%), whereas c_1 and c_3 both have roughly a 40% of consumption. This information will be used to analyse the calibration results: errors in the average percentage of consumption of the calibrated demand components compared to the assumed consumption in Table 4.2 can be assigned to background leakage, burst, fraudulent consumptions, unknown status valves, non-metered users or wrong billing information.

4.5 Conclusions

One of the main issues in a calibration process of a complex system is to assure the identifiability. Here, a redefinition of the parameters to be calibrated in order to reduce its number is used. This new set of parameters is defined by the measurements that are available. If the sensor distribution is a part of the process, an optimal sensor distribution is provided in a straightforward way using the information matrix. Nonetheless, if the sensors are already installed, the parameter definition can adapt optimally to the information available.

The calibration problem is formulated as an optimization. The solution of this problem includes nonlinear equality constraints, and thus, it is not a convex one. Much research in this area is going on and results get seldom to the real applications because of the difficult trade-off between computational effort and reliability of the resulting models. Chapter 5 presents the state of the art and an original approach to this problem.

References

1. Ahmed I, Lansey K, Araujo J (1999) Data collection for water distribution network calibration. In: 2nd International conference on water pipeline systems, pp 271–278, Exeter
2. American Water Works Association Research Committee on Distribution Systems (1974) Water distribution research and applied development needs. J Am Water Works Assoc 66(6):385–390
3. Aster R, Borchers B, Thurber C (2005) Parameter estimation and inverse problems. Elsevier, New York
4. Bard Y (1974) Nonlinear parameter estimation. Academic Press, San Diego, California
5. Bargiela A (1985) An algorithm for observability determination in water-system state estimation. Control Theory Appl 132(6):245–250

6. Behzadian K, Kapelan Z, Savic D, Ardeshir A (2009) Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks. Environ Modell Softw 24(4):530–541
7. Bonada E, Meseguer J, Mirats Tur JM (2014) Practical-oriented pressure sensor placement for model-based leakage location in water distribution networks. In: Piasecki M (ed) Informatics and the environment: data and model integration in a heterogeneous hydro world. New York
8. Bush C, Uber J (1998) Sampling design methods for water distribution model calibration. J Water Resour Plan Manage 124(6):334–344
9. Carpentier P, Cohen G (1991) State estimation and leak detection in water distribution networks. Civil Eng Syst 8(4):247–257
10. Chen LC (1995) Pipe network transient analysis—the forward and inverse problems. PhD thesis, Cornell University
11. Cheng W, He Z (2011) Calibration of nodal demand in water distribution systems. J Water Resour Plan Manage 137(1):31–40
12. Datta R, Sridharan K (1994) Parameter estimation in water distribution systems by least squares. J Water Resour Plan Manage 120(4):405–422
13. de Schaetzen WBF, Walters GA, Savic DA (2000) Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms. Urban Water J 2(2):141–152
14. Del Giudice G, Di Cristo C (2003) Sampling design for water distribution networks. Trans Ecol Environ 61
15. Eggener C, Polwoski L (1976) Network models and the impact of modeling assumptions. J Am Water Works Assoc 68(4):189–196
16. Ferreri G, Napoli E, Tumbiolo A (1994) Calibration of roughness in water distribution systems. In: 2nd International conference on water pipeline systems, pp 379–396
17. Giustolisi O, Walski T (2012) Demand components in water distribution network analysis. J Water Resour Plan Manage 138(4):356–367
18. Goulet J-A, Coutu S, Smith IFC (2013) Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. Adv Eng Inform 27(2):261–269
19. Griewank A, Juedes D, Mitev H, Utke J, Vogel O, Walther A (1998) ADOL-C: a package for the automatic differentiation of algorithms written in C/C++
20. Hutton CJ, Kapelan Z, Vamvakeridou-Lyroudia L, Savić DA (2014) Dealing with uncertainty in water distribution system models: a framework for real-time modeling and data assimilation. J Water Resour Plan Manage 140(2):169–183
21. Kang D, Lansey K (2010) Optimal meter placement for water distribution system state estimation. J Water Resour Plan Manage 136(3):337–347
22. Kapelan Z, Savic D, Walters G (2003) Multiobjective sampling design for water distribution model calibration. J Water Resour Plan Manage 129(6):466–479
23. Kapelan Z, Savic D, Walters G (2003) A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks. J Hydraul Res 41(5):481–492
24. Kapelan Z, Savic D, Walters G (2005) Optimal sampling design methodologies for water distribution model calibration. J Hydraul Eng 131(3):190–200
25. Krumpholz G, Clements K, Davis P (1980) Power system observability: a practical algorithm using network topology. IEEE Trans Power Apparatus Syst PAS-99(4):1534–1542
26. Lansey K, Basnet C (1991) Parameter estimation for water distribution networks. J Water Resour Plan Manage 117(1):126
27. Lansey K, El-Shorbagy W, Ahmed I, Araujo J, Haan C (2001) Calibration assessment and data collection for water distribution networks. J Hydraul Eng 127(4):270–279
28. Liggett J, Chen L (1994) Inverse transient analysis in pipe networks. J Hydraul Eng 120(8):934–955
29. Loaiciga H, Charbeneau R, Everett L, Fogg G, Hobbs B, Rouhani S (1992) Review of ground water quality monitoring network design. J Hydraul Eng 118(1):11–37
30. Mallick K, Ahmed I, Tickle K, Lansey K (2002) Determining pipe groupings for water distribution networks. J Water Resour Plan Manage 128(2):130–139

31. Meier R, Barkdoll B (2000) Sampling design for network model calibration using genetic algorithms. J Water Resour Plan Manage 126(4):245–250
32. Nejjari F, Sarrate R, Blesa J (2015) Optimal pressure sensor placement in water distribution networks minimizing leak location uncertainty. Procedia Eng 119:953–962
33. Ormsbee L (1989) Implicit network calibration. J Water Resour Plan Manage 115(2):243–257
34. Ostfeld A, Salomons E, Ormsbee L, Uber J, Bros C, Kalungi P, Burd R, Zazula-Coetzee B, Belrain T, Kang D, Lansey K, Shen H, McBean E, Wu ZY, Walski T, Alvisi S, Franchini M, Johnson J, Ghimire S, Barkdoll B, Koppel T, Vassiljev A, Kim JH, Chung G, Yoo DG, Diao K, Zhou Y, Li J, Liu Z, Chang K, Gao J, Qu S, Yuan Y, Devi Prasad T, Laucelli D, Vamvakeridou Lyroudia L, Kapelan Z, Savic D, Berardi L, Barbaro G, Giustolisi O, Asadzadeh M, Tolson B, McKillop R (2012) Battle of the water calibration networks. J Water Resour Plan Manage 138(5):523–532
35. Ozawa T (1987) The principal partition of a pair of graphs and its applications. Discrete Appl Math 17(1–2):163–186
36. Pérez R (2003) Identifiability and calibration of water network models. PhD thesis, Universitat de Catalunya
37. Pérez R, Puig V, Pascual J, Peralta A, Landeros E, Ll Jordanas (2009) Pressure sensor distribution for leak detection in Barcelona water distribution network. Water Sci Technol: Water Supply 9(6):715
38. Pérez R, Sanz G (2014) Optimal placement of metering devices for multiple purposes. In: 11th International conference on hydroinformatics, New York
39. Pérez R, Sanz G, Puig V, Quevedo J, Escofet MAC, Nejjari F, Meseguer J, Cembrano G, Mirats Tur JM, Sarrate R (2014) Leak localization in water networks: a model-based methodology using pressure sensors applied to a real network in Barcelona [applications of control]. IEEE Control Syst 34(4):24–36
40. Piller O, Bremond B, Morel P (1999) A spatial sampling procedure for physical diagnosis in a drinking water supply network. In: Savic DA, Walters GA (eds) Water industry systems: modelling and optimization applications, pp 309–316. Exceter
41. Piller O, Deuerlein J, Gilbert D, Weber J-M (2015) Installing fixed sensors for double calibration and early-warning detection purposes. Procedia Eng 119:564–572
42. Pinzinger R, Deuerlein J, Wolters A, Simpson A (2011) Alternative approaches for solving the sensor placement problem in large networks. In: Water distribution systems analysis 2011
43. Sanz G, Pérez R (2014) Comparison of Demand Pattern Calibration in Water Distribution Network with Geographic and Non-Geographic Parameterization. In: 11th International conference on hydroinformatics, New York
44. Sanz G, Pérez R (2014) Demand pattern calibration in water distribution networks. Procedia Eng 70:1495–1504
45. Sanz G, Pérez R (2014) Parameterization and sampling design for water networks demand calibration using the singular value decomposition: application to a real network. In: 11th International conference on hydroinformatics, New York
46. Sanz G, Pérez R (2015) Sensitivity analysis for sampling design and demand calibration in water distribution networks using the singular value decomposition. J Water Resour Plan Manage 04015020
47. Savic D, Kapelan Z, Jonkergouw P (2009) Quo vadis water distribution model calibration? Urban Water J 6(1):3–22
48. Shamir U, Howard C (1977) Engineering analysis of water-distribution systems. J Am Water Works Assoc 69(9):510–514
49. Sorenson H (1980) Parameter estimation: principles and problems. Marcel Dekker, New York
50. Sumer D, Lansey K (2009) WDS calibration and assessment for alternative modelling objectives. Urban Water J 6(4):265–277
51. Walski T (1983) Technique for calibrating network models. J Water Resour Plan Manage 109(4):360
52. Walski T (1985) Assuring accurate model calibration. J Am Water Works Assoc 77(12):38–41

53. Walski T (1995) Standards for model calibration. In: American water works association computer conference, Norfolk
54. Walski T (2000) Model calibration data: the good, the bad, and the useless. J Am Water Works Assoc 92(1):94–99
55. Walski T, Sage P, Zheng W (2014) What does it take to make automated calibration find closed valves and leaks? World Environ Water Resour Congr 2014:555–565
56. Walter E, Pronzato L (1996) On the identifiability and distinguishability of nonlinear parametric models. Math Comput Simul 42(2–3):125–134
57. Wiggins R (1972) The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure. Rev Geophys 10(1):251–285
58. William Y (1986) Review of parameter identification procedures in groundwater hydrology: the inverse problem. Water Resour Res 2:95–108
59. Yu G, Powell R (1994) Optimal design of meter placement in water distribution systems. Int J Syst Sci 25(12):2155–2166