PARAMETERIZATION AND SAMPLING DESIGN FOR WATER NETWORKS DEMAND CALIBRATION USING THE SINGULAR VALUE DECOMPOSITION: APPLICATION TO A REAL NETWORK

GERARD SANZ ESTAPÉ, RAMON PÉREZ MAGRANÉ (1)
(1): Department of Automatic Control, Polytechnic University of Catalonia, Terrassa, Spain

The availability of a good hydraulic model increases the reliability of the results of methodologies using it. Thus, the calibration of the model is a previous step that has to be done. The most uncertain parameters of the model are demands due to their constant variability. However, calibrating these demands requires a high computational cost that can be reduced by redefining the unknown parameters from nodal demands to demand patterns. Besides, the number and location of the used sensors is highly correlated with the definition of such patterns. This paper presents a methodology for parameterizing the network and selecting sensors using the information from the singular value decomposition of the water distribution network sensitivity matrix. The application of this methodology on a real network is presented.

INTRODUCTION

Water distribution network models are used by water companies in a wide range of applications. The reliability of the results when using these models depends on their good calibration. Savic *et. al* [5] thoroughly reviewed the state of the art in water distribution network model calibration.

Water distribution networks are generally formed by thousands of pipes and nodes. However, the number of measurements taken is reduced to a few selected locations. This fact makes unfeasible the problem of calibrating thousands of individual demands. Sanz and Pérez [3] proposed a calibration methodology based on the iterative resolution of the sensitivity matrix inverse using the Singular Value Decomposition (SVD) for calibrating demand patterns. The used criterion for aggregating nodal demands was based in the nodes type of contract. This gave as a result a new parameterization of the model.

Recently, telemetries from a District Metered Area (DMA) called Nova Icaria within Barcelona network have been received from the company. Fig. 1 represents the cross correlation of 24 hours telemetries within the same type of contract: (a) type 1; and (b) type 2. It can be seen that there is no relation between the behaviors of the telemetries that share the same type of contract. This work proposes a methodology for re-parameterizing the nodal demands into demand patterns that will be calibrated together based on the information contained in the sensitivity matrix SVD, and the subsequent selection of the most sensitive sensors.
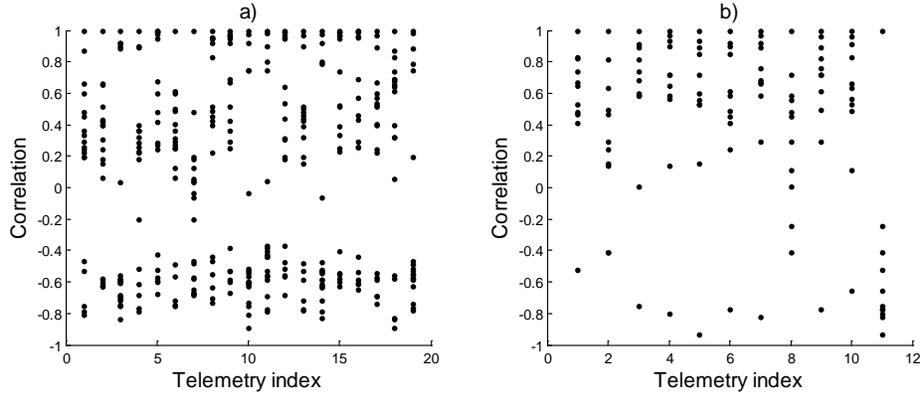
Figure 1. Cross correlations of Nova Icaria DMA telemetries in contracts (a) 1; and (b) 2

SENSITIVITY CALCULATION

The SVD of the sensitivity matrix S is used in [3] for the resolution of the inverse problem applied to water networks demand calibration. There is a lot of information contained in the SVD of matrix S, which can be used for the network parameterization and sensor selection.

Yeh [8] reviewed three methodologies for generating the sensitivity matrix in groundwater hydrology: (a) Influence Coefficient Method; (b) Sensitivity Equation Method; and (c) Variational Method. All three methods require $n+1$ simulations to be run in order to compute the complete sensitivity matrix, where n is the number of initial parameters in the model.

Cheng and He [2] proposed a matrix analysis of the water distribution network linearized model in order to obtain the sensitivity matrix, where only one simulation is required. The matrix model of the water network is

$$BCB^T H = D \tag{1}$$

where B is the incidence matrix of the network; C is the non-linear matrix depending on the pipes roughness, lengths, diameters and hydraulic gradient; H is the head vector; and D is the nodal demand vector. Considering an error ΔD in predicted demands D_p that produces an error ΔH in predicted heads H_p, the perturbation equation can be computed as

$$BCB^T(H_p + \Delta H) = (D_p + \Delta D)$$
$$BCB^T \Delta H = \Delta D \tag{2}$$

Defining $S = (BCB^T)^{-1}$, the formulation of the generalized inverse problem is obtained:

$$S \cdot \Delta D = \Delta H \tag{3}$$

Null head loss problem
When the hydraulic gradient between two nodes is null due to low flow or high conductivity of the pipe, the calculation of the C matrix is unfeasible due to the presence of indeterminate forms. Skeletonization based on [6] is considered to avoid this problem.

SINGULAR VALUE DECOMPOSITION

The SVD is capable of solving under-, over-, even- or mixed-determined problems with no rank conditions in S. The SVD of matrix S with dimensions $m \times n$ is:

$$S = U \cdot \Lambda \cdot V^T \tag{4}$$

where U is an $m \times m$ matrix of orthonormal singular vectors associated with the m observed data; V is an $n \times n$ matrix of orthonormal singular vectors associated with the n system parameters; and Λ is an $m \times n$ matrix of the singular values of S. The resolution of the inverse problem in Eq. 3 is solved by manipulating the SVD matrices:

$$\Delta D = V\Lambda^{-1}U^T \cdot \Delta H \tag{5}$$

Matrices V and U give additional information about the parameter resolution $R=VV^T$ and the information density $Id=UU^T$. The first describes how the generalized inverse solution smears out the original model x into a recovered model $\hat{x}$; while the second describes how the generalized inverse solution smears out the original data b into a predicted data $\hat{b}$.

PARAMETER DEFINITION

Nodal demands in water distribution network models are defined as

$$d_i(t) = bd_i \cdot p_i(t) \cdot q_{in}(t) \tag{6}$$

where $d_i(t)$ is the demand of node i at sample t; bd_i is the base demand of node i; $p_i(t)$ is the behavior (pattern) of node i at sample t; and $q_{in}(t)$ is the total inflow at sample t. The base demand represents the weight of each node over the whole network, and is obtained from billing. Calibrating the thousands of p_i nodal demand behaviors is unfeasible due to the low number of available measurements. However, a new parameterization can be done by considering the same pattern for groups of nodal demands. Subsequently, these nodal demands are estimated through the calibration of the patterns (Eq. 7).

$$D(t) = BDM \cdot TPM \cdot P(t) \cdot q_{in}(t) \tag{7}$$

where $D(t)$ is a vector containing n nodal demands at sample t; BDM is the Base Demand Matrix, a diagonal $n \times n$ matrix containing the base demand values of each node; TPM is the Type of Pattern Matrix, an $n \times k$ matrix associating each initial parameter (nodal demand) to a unique new parameter (demand pattern); and $P(t)$ is a vector containing k patterns at sample t.

The aggrupation of nodal demands into patterns can be obtained from the analysis of the SVD. Wiggins [7]: "We can think of the eigenvectors v_j where $j=1...k$ as a new parameterization of the model. These vectors represent a set of k specific linear combinations of the old parameters that are fixed by the observations". V_r is a matrix formed by k vectors v_j, where k is the number of non-zero singular values of the sensitivity matrix, and hence, the rank of the matrix. The formalization of the new parameterization is done defining a new parameter correction $\Delta D^*=V_r^T\Delta D$. The new k parameters are determined uniquely by the simultaneous equations. In water distribution systems very low singular values appear, thus k is defined in a way that the

values below the k highest singular value are neglected. The use of very low singular values leads to the increment of uncertainty [1].

Ideally, we would like $v_{ji} = \delta_{ij}$ (v_{ji} are the components of V and δ_{ij} is the Kronecker delta), thus each parameter would be individually resolved. In fact, a more idealistic case would have all singular values with the same value. Hence, the objective is to find linear combinations of v_j that generate new v_j^* of delta type:

$$\min \sum_{l=1}^{n}\left(v_{lj}^* - \delta_{lj}\right)^2$$
$$\text{subject to } v_{lj}^* = \sum_{i=1}^{k}(b_{ij} \cdot V_{li}) \tag{8}$$

Since v_j are orthonormal, the solution is $b_{ij} = v_{ji}$. The new V^* matrix is then the resolution matrix R_r computed with V_r. Each column r_j from R_r is the least squares solution that maximizes parameter j. If two parameters are highly correlated, both will be maximized at the same time.

Decision of pattern distribution using delta vectors

Wiggins in [7] presented an approach for extracting k orthonormal vectors from the resolution matrix R_r that enhance the delta-like behavior of that matrix. A similar approach is used in order to classify the consumers of the network in groups depending on their sensitivity. The procedure is as follows:

1. Generate the sensitivity matrix S of the system, and perform the SVD.
2. Generate V_r, a reduced form of the V matrix where only the first k columns corresponding to the highest k eigenvalues of Λ are selected.
3. Compute the resolution matrix $R_r = V_r \cdot (V_r)^T$.
4. Find the column vector r_j with the highest resolving power (generally associated with the highest diagonal element in R_r).
5. Normalize the vector as $v_j^* = r_j/(r_{jj})^{1/2}$.
6. Compute the new resolution matrix $R_r = R_r - v_j^* \cdot (v_j^*)^T$, where the row and column corresponding to v_j are now null.
7. Repeat 4-6 until k orthonormal vectors are obtained.

These called "delta vectors" are used to generate the Type of Pattern Matrix (TPM), which associates each nodal demand to the pattern with highest value in the delta vectors. The solution tends to generate geographical patterns, as the topological information is included in the sensitivity matrix through the incidence matrix B, as seen in Eq. 1.

Number of patterns decision

When facing a water distribution network with no installed sensors, one of the critical decisions is to define the number of patterns k that will be generated. The generated delta vectors lead to different distributions of patterns depending on the number of columns k used in the V_r matrix. Some criteria to define k are:

- A number k of patterns implies a minimum of k sensors to be installed. The economic cost of these sensors has to be considered.
- A high number of patterns leads to higher variances on their estimation. Contrary, a low number of patterns generates higher errors on model predictions.
- The application. Depending on the intended use of the parameterization, the number of patterns can vary.

SENSOR SELECTION

Once the new parameterization is done, the k most sensitive sensors have to be selected. First, the sensitivity matrix S^* has to be computed as in Eq. 9. This new sensitivity relates changes in heads with changes in the new parameters (groups of demands).

$$S^* = S \cdot TPM \tag{9}$$

The same process as in previous section is followed, but this time the objective is to generate delta vectors from the information density matrix, computed as

$$I_d = U_r (U_r)^T \tag{10}$$

where U_r is a reduced form of U, where only the first k columns have been considered. From each generated delta vector, the node with the highest value is chosen as sensor. Ideally, each chosen sensor would have the highest sensitive to one of the new k parameters while being little sensitive to the rest.

Having preinstalled sensors

If M sensors are available before performing the network parameterization, this information can be included in the process in order to get the best pattern definition. Thus, the demand groups are computed from the S matrix with only M rows. The resulting patterns are those that will be better resolved considering the available measurements. In this case the number of patterns depends on the number of sensors.

RESULTS

The explained methodology is applied to a real case study: a DMA called Nova Icaria (Fig. 2a) within Barcelona water network. It is composed by 3455 pipes and 3377 junctions, 1226 of which are consumers. The water is provided from a transport network through two pressure reduction valves (PRV), which are monitored with pressure and flow measurements with a sample time of 10 minutes. These measurements are used to fix the boundary conditions when the model is simulated: the measured pressures fix the set point of the PRVs, and the sum of the measured flows is used to calculate the nodal demands.
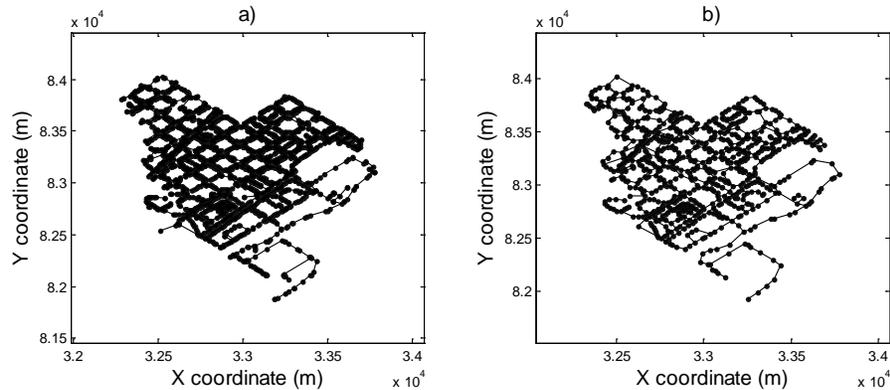

Figure 2. Nova Icaria Water Distribution Network: a) Original Model and b) Reduced Model
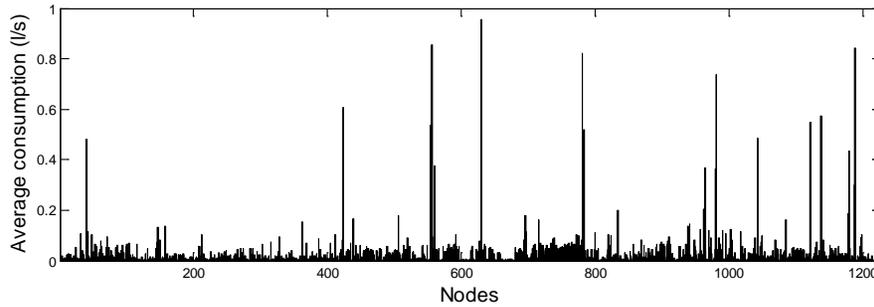
Figure 3. Nodal base demands calculated from quarterly billing

The methodology in this work prepares the network for the calibration, so there are no calibrated patterns yet. The demand model used is based on Eq. 7. Quarterly billing from November 2012 has been used, summing for each network junction the three months average consumption of contracts connected to it (Fig. 3).

Skeletonization of the network
The existence of pipes with high conductivity leads to the problem of obtaining null head losses. This has been solved by skeletonizing the network using the formulas from [6]. The reduced network is depicted in Fig. 2b, with 911 pipes (26% of the original network) and 791 junctions (23%). Demands from reduced nodes joined the remaining ones.

Parameter definition
The parameter definition process has been carried out as explained in previous sections. However, data from 24 hours have been used, leading to 24 parameter distributions. Most of the nodes were assigned to the same parameter in all of the samples. When this did not happen, the nodal demand was assigned to its highest repeated parameter.

Four, five and six pattern distributions have been considered taking into account the cost of the sensors and the final use. This use consists in calibrating demand patterns while detecting anomalies in their expected values in order to find leakages. Fig. 4 depicts the six patterns case. Table 1 presents the combination of parameters when considering four and five patterns. Notice the geographical distribution due to the topological information included in the sensitivity matrix.

Sensor selection
The sensor selection process is performed once the parameterization is completed. In this case, the parameterization with six patterns is used because compared to the four patterns one it separates 2 zones in 4 smaller ones. That is useful for the final application, as more differentiated zones are generated.

Again, the methodology explained previously is applied, and the six most sensitive sensors are suggested. 24 hours of data are used, and the more repeated sensors are chosen. Fig. 4 depicts with a star the position of the selected sensors. It should be noted that each sensor has been located in a different zone (corresponding to a defined parameter). This is not an imposed constraint, but a logic result as the SVD based methodology selects the sensors that produce the best information density matrix.
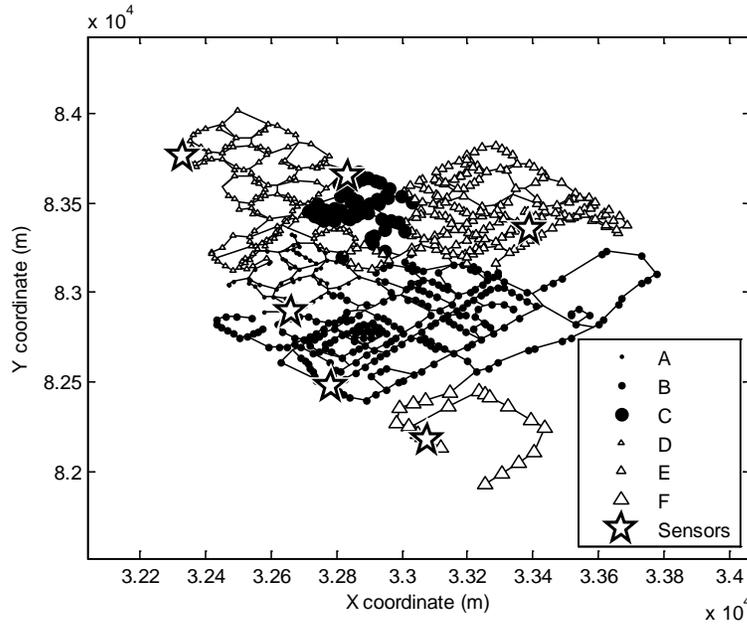
Figure 4. Pattern distribution and sensor selection using 6 parameters

Table 1. Pattern combination depending on number of parameters

6 Parameters	5 Parameters	4 Parameters
A	A	A & D
B	B	B
C	C & E	C & E
D	D	-
E	-	-
F	F	F

CONCLUSIONS AND FUTURE WORK

This work presents two pre-calibration methodologies based on the analysis of the SVD of the water distribution system sensitivity matrix. The first one classifies the nodal demands depending on their pressure sensitivity. The number of parameters (patterns) is defined by the user depending on the final use and economic budget for installing sensors. In the current work a final application for leakage detection and isolation based on demand pattern calibration is assumed, and six parameters have been defined. The solution tends to generate geographical patterns as the topological information is included in the sensitivity matrix.

The second proposed methodology selects the sensors that provide the most information from the system. Six sensors have been distributed, each one located (not as a constraint) in a differentiated zone. Both methodologies are used in [4] to calibrate geographical patterns.

In future works, nodal demands will not be classified in one unique group, but as a combination of multiple patterns. This will generate a more realistic demand model, as each nodal demand would have a different behavior obtained from the mixing of patterns. The membership degree of a node to each pattern would be produced by the same process presented in this work.

ACKNOWLEDGMENTS

REFERENCES

[1] Aster R., Borchers B. and Thurber C., "Parameter Estimation and Inverse Problems", New York: Elsevier, (2005).

[2] Cheng W. and He Z., "Calibration of Nodal Demand in Water Distribution Systems", Journal of Water Resources Planning and Management, Vol. 137, No. 1 (2011), pp. 31–40.

[3] Sanz G. and Pérez R., "Demand Pattern Calibration in Water Distribution Networks", Proc. 12th International Conference on Computing and Control for the Water Industry, Perugia (2013).

[4] Sanz G. and Pérez R., "Comparison of Demand Pattern Calibration in Water Distribution Networks with Geographic and Non-Geographic Parameterization", Proc. 11th International Conference on Hydroinformatics, New York (2014).

[5] Savic D., Kapelan Z. and Jonkergouw P., "Quo vadis water distribution model calibration?", Urban Water Journal, Vol. 6, No. 1 (2009), pp 3–22.

[6] Walski T., Chase D., Savic D., Grayman W., Beckwith S. and Koelle E., "Advanced Water Distribution Modeling and Management", Haestad Press, (2003).

[7] Wiggins R., "The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure", Reviews of Geophysics, Vol. 10, No. 1 (1972), pp 251-285.

[8] Yeh W., "Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem", Water Resources Research, Vol. 22, No. 2 (1986), pp. 95-108.