International Conference on Hydroinformatics

8-1-2014

Experimental Design For Model Discrimination And Precise Parameter Estimation In WDS Analysis

Giovanna Darvini

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

EXPERIMENTAL DESIGN FOR MODEL DISCRIMINATION AND PRECISE PARAMETER ESTIMATION IN WDS ANALYSIS

GIOVANNA DARVINI (1)

(1): Department ICEA, Università Politecnica delle Marche, via Brecce Bianche, Ancona, 60131,Italy

In water distribution system (WDS) analysis, hydraulic models may be used for design, optimization and control purposes. The hydraulic model of an existing network is built on the basis of the quality and quantity of the collected data. During this first stage the correction of input data errors should be completed, since different models of the network could provide results consistent with available experimental data. After choosing the best model among those available, one must calibrate the selected one. Very frequently, distinct procedures are applied for discrimination of rival models and estimation of precise model parameters, leading to different sample designs. To conciliate the objectives of both experimental design procedures, the present paper proposes the use of a multi-objective optimization method. Different models of a water distribution network have been compared and calibrated by using available measurements of nodal pressure. As observed in the analyzed examples, the analysis of the Pareto set provides the identification of the optimal location where install additional pressure sensors for adequate model discrimination and parameter estimation.

INTRODUCTION

In WDS analysis, the initial model building stage is significantly complicated by the uncertainties in the collected data (network topology, informations about pipes, asset data, demand allocation, etc.). As a consequence, more than one model may be taken to simulate the WDS, being each of these different models capable of representing the available experimental data with similar performances. Consequently, additional experiments must be designed to select the best model among the available alternatives. In addition, further experiments may be needed to improve the precision of the parameter estimates, which will improve also model predictions and performance [1]

Experimental design procedures for model discrimination and for estimation of precise model parameters are usually treated as independent techniques and generally lead to different experimental designs, although model discrimination and reduction of variances of parameter estimates are closely related to each other [2]. In the present paper the use of a multi-objective optimization method, derived by the field of the chemical engineering, is proposed for WDS analysis [3]. Different models of an existing network are juxtaposed with one another. During calibration, available measurements of nodal pressure are compared with results obtained by each hydraulic model to determine the values of the pipe roughness, assumed as model parameters. Starting from an initial number of pressure data, the procedure allows to identify optimal locations for additional measures. In turn, these new data are added to ones previously available and used to perform new simulations though which it is possible simultaneously obtain a suitable model discrimination and precise parameter estimation. The results are presented as Pareto fronts where the two different objectives can be compared with each other to find the optimal solution of the sampling design problem.

EXPERIMENTAL DESIGN

Experimental designs are almost always performed iteratively. Based on previous experimental observations, new experiments are designed and performed in order to optimize the performance index. Then, the new observations are included in the experimental data set and the desired objectives are analyzed. If the obtained results do not meet the required performance indexes, new experiments are designed and realized [1]. In the following a brief summary of the sequential experimental design for precise parameter estimation and rival models discrimination is presented separately.

Precise parameter estimation

The sequential experimental design for precise parameter estimation is usually performed through the optimization of a norm of the posterior covariance matrix of parameter estimates, described as the expected covariance matrix of parameter estimates after the inclusion of the new set of observations in the experimental data set. The posterior covariance matrix of parameter estimates can be defined as follows [2,4]:

$$\tilde{\mathbf{V}}_{\theta,m} = \left[\sum_{i=N+1}^{N+K} \mathbf{B}_{i,m}^T \mathbf{V}_i^{-1} \mathbf{B}_{i,m} + \mathbf{V}_{\theta,m}^{-1} \right]^{-1} \tag{1}$$

where $\mathbf{V}_{\theta,m}$ is the current covariance matrix of parameter estimates of model m (obtained with the available N experiments), $\tilde{\mathbf{V}}_{\theta,m}$ is the posterior covariance matrix of parameter estimates (after inclusion of the new K observations), $\mathbf{V}_i$ is the covariance matrix of experimental uncertainties at experimental condition i and $\mathbf{B}_{i,m}$ is the matrix of sensitivities of model m which contains the first derivatives of the responses of model m with respect to the model parameters at the ith experimental condition, defined as

$$b_{r,p} = \frac{\partial y_r}{\partial \theta_p} \tag{2}$$

that is, the derivative of the response r with respect to the parameter p (for a specific model m and experimental condition i).

Model discrimination

To perform the discrimination of rival models, experimental conditions are normally designed for maximization of some measure of the difference of the responses obtained with the distinct probable models. Following Schwaab *et al.* [5], after performing N preliminary experiments, a new experimental condition should be selected in order to maximize the model discrimination function, defined as

$$D_{m,n} = (P_m P_n)^Z (\mathbf{y}_m - \mathbf{y}_n)^T \mathbf{V}_{m,n}^{-1} (\mathbf{y}_m - \mathbf{y}_n) \tag{3}$$

where $\boldsymbol{y}_m$ is a vector of responses of model m at experimental condition $\mathbf{x}$ with model parameters $\boldsymbol{\theta}_m$ estimated from the available N experiments. In the case of M models, the discriminant can be computed considering all pairs of models m and n.

The matrix $\mathbf{V}_{m,n}$ is defined as

$$\mathbf{V}_{m,n} = 2\mathbf{V} + \mathbf{V}_m + \mathbf{V}_n \tag{4}$$

where $\mathbf{V}$ is the covariance matrix of experimental deviations and $\mathbf{V}_m$ is the covariance matrix of model prediction deviations calculated from model m. In turn, the covariance matrix of model prediction for model m can be calculated as [2,4]:

$$\mathbf{V}_m = \mathbf{B}_m \widetilde{\mathbf{V}}_{\theta,m} \mathbf{B}_m^T . \tag{5}$$

In Eq. (3) Z is a parameter used to modulate the relative importance of the rival models: if Z is greater than 1, model prediction differences are magnified; if Z is smaller than 1, model prediction differences are minimized. The model probabilities P_m and P_n used in Eq.(3) can be calculated with the help of standard statistical tools and are not sensitive to the ordering of available experiments [5].

MULTI-OBJECTIVE PROCEDURE

The multi-objective optimization problem consists in the simultaneous optimization of S objective functions ($f_1(\mathbf{x})$, $f_2(\mathbf{x})$,…, $f_S(\mathbf{x})$, $S\geq 2$), where $\mathbf{x}$ is the feasible set of decision vectors. In multi-objective optimization, there does not typically exist a feasible solution that minimizes all objective functions simultaneously. Therefore, attention is paid to Pareto optimal solutions, constituted by solutions that cannot be improved in any of the objectives without degrading at least one of the other objectives.

By following Alberton *et al.* [1], the multi-objective optimization procedure is defined here as the simultaneous maximization of design criteria used for estimation of precise model parameters and discrimination of rival models, assuming that a single experiment must be designed each time. The experimental design criterion used for precise parameter estimation is the minimization of the determinant of the posterior covariance matrix of parameter estimates. To improve the quality of the result presentation, the objective function is normalized to give values in the range [0,1], as follows:

$$F_{Em} = \det\left(\widetilde{\mathbf{V}}_{\theta,m}^{-1}\right)_{\min} \Big/ \det\left(\widetilde{\mathbf{V}}_{\theta,m}^{-1}\right). \tag{6}$$

In analogous manner, Eq. (3) is rewritten as relative objective functions, as

$$F_D = D_{m,n} \big/ D_{\max} . \tag{7}$$

As a consequence, functions expressed by Eqs. (6) and (7) must be maximized in the search region.

ILLUSTRATIVE EXAMPLE

The hypothetical network proposed by Greco and Del Giudice [6] was used to test the procedure for discrimination among three different models. The WDS scheme is illustrated in Figure 1.

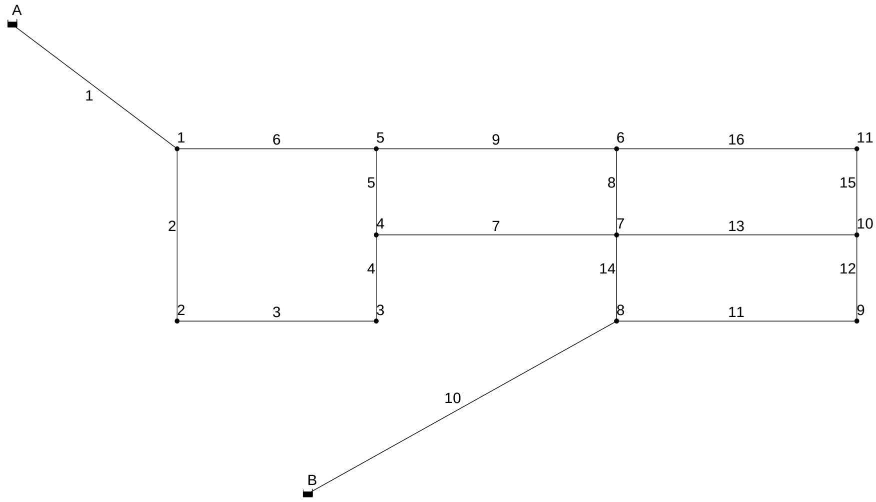

Figure 1. Network scheme [6]

The pipe characteristics and the nodal demand for the model 1 are reported in Tables 1 and 2 respectively. Model 2 and 3 differs from the first one just for the diameter of one pipe. In the model 2 the diameter of pipe 5 is changed from 80 mm to 100 mm; while, in the model 3 the diameter of pipe 15 is modified from 150 mm to 175 mm.

In the initialization of the procedure, it is assumed that model 1 is the actual model. Given the nodal demands of Table 2, the simulation model uses the roughness values ε** to generate the nodal pressure reported in Table 2, assumed as experimental pressure data at nodes where sensors are installed. Water distribution network simulation model EPANET [7] was used to calculate nodal pressures, but other software can be employed.

Table 1. Pipe characteristics

Pipe	Diameter (mm)	Length (m)	Roughness ε** (mm)	Initial estimated roughness ε * (mm)
1	250	200	0.60	0.5
2	150	400	0.30	0.3
3	200	300	0.10	0.7
4	300	190	0.70	0.4
5	80	210	0.40	0.5
6	200	300	1.50	0.7
7	150	160	0.30	0.3
8	300	200	1.00	0.7
9	200	180	0.10	0.4
10	250	140	0.10	0.7
11	200	360	0.60	0.3
12	150	200	0.40	0.5
13	150	340	0.90	0.4
14	80	180	0.20	0.4
15	150	180	0.05	0.3
16	200	345	0.60	0.5

Table 2. Nodal demand and pressure values

Node	Demand (l/s)	Pressure (m)
1	20	86.63
2	30	67.36
3	20	67.09
4	10	67.10
5	20	70.27
6	30	67.83
7	10	67.81
8	10	79.01
9	20	72.96
10	10	68.07
11	30	67.26

The pipes are divided in four groups with the same roughness coefficient, and the initial estimates of pipe roughness ε^* reported in Table 1 are arbitrarily chosen. The reduction of the parameter dimension is a common engineering practice for reducing the number of unknown parameters in the WDS calibration. The influence of pipe groupings on the model error and the model prediction error was thoroughly analyzed by Mallick *et al.* [8]. For calibrating the network model the approach proposed by Greco and Del Giudice [6] is adopted.

Preliminary experiments are developed by locating five pressure sensors at nodes 2, 3, 4, 6 and 11 and the corresponding "measured pressure data" of Table 2 are used for parameter estimation. In the example, experimental variances are set to 0.01 for each pressure measurement $y_{r,m}$. The errors of experimental responses are assumed to be independent from each other. In this example, when one model assumes a relative probability smaller than 2.5%, the model is discarded [5]. Since some authors argue strongly against the use of the model probabilities in discrimination procedure [9], in Eq. (3) is assumed Z=0.

Results of the preliminary experiments are reported in Table 3. On the basis of the probability values, model 2 can be eliminated, while between models 1 and 3 the discrimination is not possible and new experiments must be designed in order to discriminate between them. Moreover, Table 3 shows that the model 1 has a larger probability to be the best model, while the determinant of the covariance matrix of parameter estimates for model 3 is smaller than value obtained for model 1.

Table 3. Results after preliminary experiments with five sensors

Model	ε_1 (mm)	ε_2 (mm)	ε_3 (mm)	ε_4 (mm)	P_m	$\det V_{\theta,m}$
1	0.56	0.41	0.77	0.45	0.536	$2.72\cdot10^{-31}$
2	0.60	0.44	0.76	0.43	0.006	$3.43\cdot10^{-31}$
3	0.67	0.32	0.82	0.47	0.458	$2.32\cdot10^{-33}$

In order to select the optimal location for the sixth pressure sensor, the Pareto front for the relative estimation function F_{E1} and F_{E2} and the relative discrimination function F_D is illustrated in Figure 2, by considering all the possible cases. The Pareto front shows that the two objectives are conflicting for both the models and suggests that locating the additional sensor at node 5 leads to the more precise roughness estimation, while the highest discriminant function is obtained if one install the sensor at node 8.

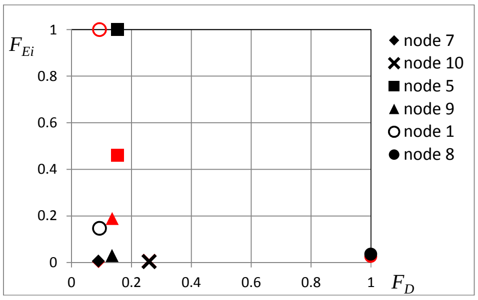

Figure 2. Pareto front for the relative estimation function F_{E1} (red) and F_{E2} (black) and the relative discrimination function F_D for design of the sixth sensor.

Table 4 shows the results obtained after inclusion of the sixth sensor at each of the potential nodes and the new parameter estimation. In any of the cases one of two rival models reaches a relative probability higher than 97.5% and the discrimination is not attained.

Table 4. Model probability after including additional sixth sensor

	node 7	node 10	node 5	node 9	node 1	node 8
P_1	0.40	0.65	0.35	0.10	0.90	0.15
P_2	0.60	0.35	0.65	0.90	0.10	0.85

On the basis of results reported in Figure 2, the sixth sensor is installed at node 8. Then, the procedure is repeated by considering all the remaining nodes for the installation of the seventh sensor. In Figure 3 is illustrated the Pareto front for the two objective functions with seven sensors, at all the possible positions. The analysis of the front shows that the optimal solution is installing the additional sensor at node 5, where for model 2 both objective functions are equal to 1.

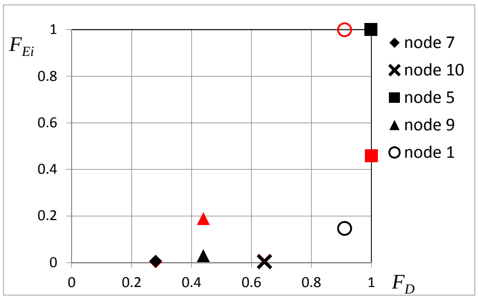

Figure 3. Pareto front for the relative estimation function F_{E1} (red) and F_{E2} (black) and the relative discrimination function F_D for design of the seventh sensor.

The results obtained for seven sensors and after the new parameter estimation are reported in Table 5 for all the nodes considered. When the seventh sensor is added at node 5, the model probabilities indicate that model 2 results not adequate and the model 1 is selected as the best model. Nowhere else the discrimination is possible. Furthermore, after the sequential procedure, the final value of the determinant of the posterior covariance matrix of parameter estimates $\det\tilde{V}_{\theta,1}$ is $1.20 \cdot 10^{-34}$, a value significantly smaller than that obtained initially for the selected model and reported in Table 3.

Table 5. Model probability after including additional seventh sensor

	node 7	node 10	node 5	node 9	node 1
P_1	0.15	0.08	0.98	0.18	0.06
P_2	0.85	0.92	0.02	0.82	0.94

CONCLUSIONS

This work presented an approach for discriminating rival models in WDS analysis and simultaneously improving the precise parameter estimation during model calibration. The proposed procedure was largely used in the field of the chemical engineering.

As shown in the example, the procedure allows the discrimination of the most suitable model among those available during the model building and permits to reduce significantly the uncertainty in the parameter estimates, by selecting the best additional experiments. Moreover, the analysis of the Pareto front shows when the two objectives are conflicting and helps the analyst in solving the sample design problem.

REFERENCES

[1] Alberton A. L., Schwaab M., Biscaia E. C. and Pinto J. C, "Sequential experimental design based on multiobjective optimization procedures", *Chemical Engineering Science*, Vol. 65, (2010), pp 5482-5494.

[2] Schwaab M., Monteiro J. L. and Pinto J. C, "Sequential experimental design for model discrimination taking into account the posterior covariance matrix of differences between model predictions", *Chemical Engineering Science*, Vol. 63, (2008), pp 2408-2419.

[3] Darvini G., "Model discrimination in water distribution systems", CCWI2013, *Procedia Engineering*, Vol. 70, (2014), pp 419-428, doi:10.1016/j.proeng.2014.02.047.

[4] Bard Y., "*Nonlinear Parameter Estimation*", Academic Press, New York, (1974).

[5] Schwaab M., Silva F. M., Queipo C. A., Barreto Jr. A. G., Nele M. and Pinto J. C., "A new approach for sequential experimental design for model discrimination", *Chemical Engineering Science*, Vol. 61, (2006), pp 5791-5806.

[6] Greco M. and Del Giudice G., "New approach to water distribution network calibration", *Journal of Hydraulic Engineering*, Vol. 125, No. 8, (1999), pp 849-854.

[7] Rossman L. A., "EPANET 2 users manual", Water Supply and Water Resources Division, National Risk Management Research Laboratory, (2000).

[8] Mallick K. N., Ahmed I., Tickle K. S. and Lansey K. E., "Determining pipe groupings for water distribution networks", *Journal of Water Resources Planning and Management*, Vol. 128, (2002), pp 130-139.

[9] Buzzi-Ferraris G. and Manenti F., "Kinetic models analysis", *Chemical Engineering Science*, Vol. 64, (2009), pp 1061-1074.