UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING
DEPARTMENT OF MECHANICAL ENGINEERING

Bayesian Optimal Experimental Design for Leakage Detection in Water Distribution Networks

Diploma Thesis of

Georgios Trousas

Thesis supervisor: **Prof. Costas Papadimitriou**

Submitted in partial fulfillment of the requirements for the Mechanical Engineering Diploma from University of Thessaly

Volos, October 2020

Thesis Committee

Prof. Costas Papadimitriou
Professor of Structural Dynamics, Department of Mechanical Engineering, University of Thessaly

Prof. Dimitrios Valougeorgis
Professor of Mesoscale Methods in Flows and Transport Phenomena, Department of Mechanical Engineering, University of Thessaly

Prof. Athanasios Ziliaskopoulos
Professor of Optimization of Production/Transportation Systems, Department of Mechanical Engineering, University of Thessaly

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Costas Papadimitriou for his continuous guidance and support through the development of this thesis. His willingness and availability to help in addressing the issues encountered have been a tremendous help in being able to finish this study.

I would also like to thank the teaching stuff of the Department of Mechanical Engineering University of Thessaly for the knowledge of engineering they have imparted to me and for keeping me motivated to continue studying throughout these years.

Abstract

This thesis investigates the problem of optimizing the location of pressure and/or flow rate sensors in water distribution networks (WDN) in order to maximize the information obtained from the data for identifying leakage in the network. Leakage detection is treated as a Bayesian system identification problem, integrating hydraulic models of steady-state flow in WDN with measurements. The formulation of the hydraulic model is presented and the Bayesian methodology for inferring the location and magnitude of leakage using measurements from pressure and flow rate sensors is developed. Each possible leakage scenario is considered as a distinct model of the system and by comparing the evidence and the posterior PDFs of the models, the most probable one is chosen. The proposed methodology is expanded to also include uncertainties in model parameters to be able to make robust predictions. To maximize the information provided by the measurements for inferring the leakage parameters, optimal sensor configurations are found by solving an optimization problem with the sensor configurations as design variables and the posterior information entropy of the parameters as the objective function. To calculate the multidimensional integrals arising in the objective function an asymptotic and a sampling approximation method is applied. Two computationally efficient heuristic sequential sensor placing algorithms are used to solve the optimization problem. The effectiveness of the resulting optimal sensor configurations is illustrated by using them to simulate measurements which are then passed to the leakage detection methodology and the accuracy of the predictions is compared to that of suboptimal sensor configurations.

Table of Contents

Thesis Committee 2

Acknowledgements 3

Abstract 4

List of Figures 7

List of Tables 11

Nomenclature 12

1. Introduction 15

2. Hydraulic model formulation 17

2.1 Frictional losses in developed pipe flow 17

2.1.1 Losses due to wall shear 17

2.1.2 Minor losses 21

2.1.3 Pumps 21

2.2 System of hydraulic equations 22

2.2.1 Continuity balance 22

2.2.2 Energy balance 23

2.2.3 Formulating a more efficient system of equations 24

2.3 Introducing a leaking pipe into the network 28

2.4 Sensitivities of hydraulic equations 28

2.4.1 Sensitivities of friction factor 29

2.4.2 Sensitivities of the model outputs with respect to the amount of leakage 33

2.4.3 Jacobian matrix for the non-linear system solver 36

3. Bayesian inference framework for leakage detection 38

3.1 Bayesian parameter estimation framework 38

3.1.1 Formulating the posterior PDF 40

3.1.2 Selecting the most probable leakage model 41

4. Bayesian optimal sensor placement formulation 43

4.1 Objective function formulation 43
4.1.1 Asymptotic approximation of information entropy 44
4.1.2 Sampling approximation of the mutual information 46
4.1.3 Monitoring an area of the network 47
4.2 Optimal sensor placement methodology 48
4.2.1 FSSP-BSSP Heuristic algorithms 48
5. Presentation of network 50
6. Optimal sensor placement results 52
6.1 Discussion of OSP results 71
7. Evaluation of optimal sensor configurations 75
7.1 Discussion of leakage detection results 89
8. Conclusions 92
9. References 93

List of Figures

Figure 1 Example of equation (30) applied to junction 3 of the illustrated network 22
Figure 2 Loop consisted of three pipes illustrating the vector values 26
Figure 3 Pseudoloop consisted of four pipes illustrating the vector values 27
Figure 4 Splitting a pipe in two and adding a node 28
Figure 5 Water Distribution Network 50
Figure 6 Monitored area of the network shown in blue 50
Figure 7 Flow rate 51
Figure 8 Pressure head 51
Figure 9 Objective function for pressure sensors 52
Figure 10 Objective function for flow-rate sensors 52
Figure 11 Best pressure sensor locations 53
Figure 12 Best flow-rate sensor locations 53
Figure 13 Worst pressure sensor locations 53
Figure 14 Worst flow-rate sensor locations 53
Figure 15 Objective function for pressure sensors 54
Figure 16 Objective function for flow-rate sensors 54
Figure 17 Best pressure sensor locations 54
Figure 18 Best flow-rate sensor locations 54
Figure 19 Worst pressure sensor locations 54
Figure 20 Worst flow-rate sensor locations 54
Figure 21 Objective function for pressure sensors 55
Figure 22 Objective function for flow-rate sensors 55
Figure 23 Best pressure sensor locations 55
Figure 24 Best flow-rate senor location 55
Figure 25 Worst pressure sensor locations 55
Figure 26 Worst flow-rate sensor locations 55
Figure 27 Objective function for pressure sensors 56
Figure 28 Objective function for flow-rate sensors 56
Figure 29 Best pressure sensor locations 56
Figure 30 Best flow-rate sensor locations 56
Figure 31 Worst pressure sensor locations 57
Figure 32 Worst flow-rate sensor locations 57

Figure 33 Objective function for pressure sensors 57
Figure 34 Objective function for flow-rate sensors 57
Figure 35 Best pressure sensor locations 58
Figure 36 Best flow-rate sensor locations 58
Figure 37 Worst pressure sensor locations 58
Figure 38 Worst flow-rate sensor locations 58
Figure 39 Objective function for pressure sensors 58
Figure 40 Objective function for flow-rate sensors 58
Figure 41 Worst pressure sensor locations 59
Figure 42 Worst flow-rate sensor locations 59
Figure 43 Best flow-rate sensor locations 59
Figure 44 Best pressure sensor locations 59
Figure 45 Objective function for pressure sensors 60
Figure 46 Objective function for flow-rate sensors 60
Figure 47 Best pressure sensor locations 60
Figure 48 Best flow-rate sensor locations 60
Figure 49 Worst pressure sensor locations 60
Figure 50 Worst flow-rate sensor locations 60
Figure 51 Objective function for pressure sensors 61
Figure 52 Objective function for flow-rate sensors 61
Figure 53 Best pressure sensor locations 61
Figure 54 Best flow-rate sensor locations 61
Figure 55 Worst pressure sensor locations 61
Figure 56 Worst flow-rate sensor locations 61
Figure 57 Objective function for pressure sensors 62
Figure 58 Objective function for flow-rate sensors 62
Figure 59 Best pressure sensor locations 62
Figure 60 Best flow-rate sensor locations 62
Figure 61 Worst flow-rate sensor locations 63
Figure 62 Worst pressure sensor locations 63
Figure 63 Objective function for pressure sensors 63
Figure 64 Objective function for flow-rate sensors 63
Figure 65 Best pressure sensor locations 63
Figure 66 Best flow-rate sensor locations 63

Figure 67 Worst pressure sensor locations.....64
Figure 68 Worst flow-rate sensor locations.....64
Figure 69 Objective function for pressure sensors.....64
Figure 70 Objective function for flow-rate sensors.....64
Figure 71 Best pressure sensor locations.....64
Figure 72 Best flow-rate sensor locations.....64
Figure 73 Worst pressure sensor locations.....65
Figure 74 Worst flow-rate sensor locations.....65
Figure 75 Objective function for pressure sensors.....66
Figure 76 Objective function for flow-rate sensors.....66
Figure 77 Best pressure sensor locations.....66
Figure 78 Best flow-rate sensor locations.....66
Figure 79 Worst pressure sensor locations.....66
Figure 80 Worst flow-rate sensor locations.....66
Figure 81 Objective function for pressure sensors.....67
Figure 82 Objective function for flow-rate sensors.....67
Figure 83 Best pressure sensor locations.....67
Figure 84 Best flow-rate sensor locations.....67
Figure 85 Worst pressure sensor locations.....67
Figure 86 Worst flow-rate sensor locations.....67
Figure 87 Objective function for pressure sensors.....68
Figure 88 Objective function for flow-rate sensors.....68
Figure 89 Best pressure sensor locations.....68
Figure 90 Best flow-rate sensor locations.....68
Figure 91 Worst pressure sensor locations.....68
Figure 92 Worst flow-rate sensor locations.....68
Figure 93 Objective function for pressure sensors.....69
Figure 94 Objective function for flow-rate sensors.....69
Figure 95 Best pressure sensor locations.....69
Figure 96 Best flow-rate sensor locations.....69
Figure 97 Worst pressure sensor locations.....70
Figure 98 Worst flow-rate sensor locations.....70
Figure 99 Objective function for pressure sensors.....70
Figure 100 Objective function for flow-rate sensors.....70

Figure 101 Best pressure sensor locations 71
Figure 102 Best flow-rate sensor locations..... 71
Figure 103 Worst pressure sensor locations 71
Figure 104 Worst flow-rate sensor locations..... 71

List of Tables

Table 1 Results of flow-rate sensors for leaking pipe 3 and $\sigma = 0.01$ 76
Table 2 Results of flow-rate sensors for leaking pipe 3 and $\sigma = 0.05$ 77
Table 3 Results of flow-rate sensors for leaking pipe 12 and $\sigma = 0.01$ 77
Table 4 Results of flow-rate sensors for leaking pipe 12 and $\sigma = 0.05$ 77
Table 5 Results of flow-rate sensors for leaking pipe 17 and $\sigma = 0.01$ 78
Table 6 Results of flow-rate sensors for leaking pipe 17 and $\sigma = 0.05$ 78
Table 7 Results of flow-rate sensors for leaking pipe 19 and $\sigma = 0.01$ 79
Table 8 Results of flow-rate sensors for leaking pipe 19 and $\sigma = 0.05$ 79
Table 9 Results of pressure sensors for leaking pipe 3 80
Table 10 Results of pressure sensors for leaking pipe 12 80
Table 11 Results of pressure sensors for leaking pipe 17 80
Table 12 Results of pressure sensors for leaking pipe 19 81
Table 13 Results of flow-rate sensors for leaking pipe 3 and $\sigma = 0.01$ 82
Table 14 Results of flow-rate sensors for leaking pipe 3 and $\sigma = 0.05$ 82
Table 15 Results of flow-rate sensors for leaking pipe 12 and $\sigma = 0.01$ 83
Table 16 Results of flow-rate sensors for leaking pipe 12 and $\sigma = 0.05$ 83
Table 17 Results of flow-rate sensors for leaking pipe 17 and $\sigma = 0.01$ 83
Table 18 Results of flow-rate sensors for leaking pipe 17 and $\sigma = 0.05$ 84
Table 19 Results of flow-rate sensors for leaking pipe 19 and $\sigma = 0.01$ 84
Table 20 Results of flow-rate sensors for leaking pipe 19 and $\sigma = 0.05$ 85
Table 21 Results of flow-rate sensors for leaking pipe 3 and $\sigma = 0.01$ 86
Table 22 Results of flow-rate sensors for leaking pipe 3 and $\sigma = 0.05$ 86
Table 23 Results of flow-rate sensors for leaking pipe 12 and $\sigma = 0.01$ 86
Table 24 Results of flow-rate sensors for leaking pipe 12 and $\sigma = 0.05$ 87
Table 25 Results of flow-rate sensors for leaking pipe 17 and $\sigma = 0.01$ 87
Table 26 Results of flow-rate sensors for leaking pipe 17 and $\sigma = 0.05$ 88
Table 27 Results of flow-rate sensors for leaking pipe 19 and $\sigma = 0.01$ 88
Table 28 Results of flow-rate sensors for leaking pipe 19 and $\sigma = 0.05$ 88

Nomenclature

A : Cross-sectional area of a pipe
a : Variable used in the Cheng friction factor formula
a_1 : Constant used in the friction factor equation
B : Sign matrix for the loop equations
b : Variable used in the Cheng friction factor formula.
$\underline{b}$: Selection vector for the loop equation
b_1 : Constant used in the friction factor equation
C : Hazen-Williams coefficient, selection matrix for the continuity equations
$\underline{c_i}$: Selection vector for the continuity equation in node i
c_1 : Constant which depends on the prior and evidence
c_2 : Constant which depends on the prior and evidence
D : Set of measurements
d : Diameter of a pipe
$\underline{\tilde{d}}$: Vector with all the nodal demands
$\tilde{d}_i$: Demand at node i
DP_f : Variable that contains all non-flow dependent variables in the frictional loss equation due to components
DP_l : Variable that contains all non-flow dependent variables in the frictional loss equation due to shear stress from pipe walls
DP_p : Variable that contains all non-flow dependent variables in the frictional loss equation due to the presence of a pump
E : Selection matrix for the energy balance equations
$\underline{e}$: Prediction error
$\underline{e_j}$: Selection vector for the energy balance in pipe j
F : Number of fixed grade nodes.
FA, FB : Variables used in the cubic interpolation from the Moody diagram.
f : Friction factor
f_1, f_2, f_3 : Variables used in the Cheng friction factor formula
g : Gravitational acceleration
$\underline{g}$: Model predictions that correspond to the measured degrees of freedom
$\underline{\tilde{g}}$: Model predictions
H : Hessian matrix
$\underline{h}$: Vector with all the hydraulic heads
h_1, h_2 : Coefficients in the pump three-point curve equation
h_f : Head loss due to components in a pipe
$\underline{h_f}$: Vector with the head losses due to components in all the pipes
$h_{f,j}$: Head loss due to components in pipe j
h_l : Head loss due to wall shear in a pipe
$\underline{h_l}$: Vector with the head losses due to wall shear in all the pipes
$h_{l,j}$: Head loss due to wall shear in pipe j
h_m : Shutoff head of a pump

h_p : Head loss due to pump in a pipe

$\underline{h_p}$: Vector of head losses due to pump in all the pipes

$h_{p,j}$: Head loss due to a pump in pipe j

$h_{\underline{\theta}}$: Information entropy of parameters $\underline{\theta}$

I : Identity matrix

J : Number of junctions, Jacobian matrix, Deviation between model predictions and measurements

J_1, J_2, J_3 : Partitions of the Jacobian matrix

K : Loss coefficient due to components

K_1 : Loss coefficient of the Hazen-Williams formula

K_2 : Coefficient of the Chezy-Manning formula

$\underline{k}^*$: Selection vector for the known hydraulic heads

L : Number of loops, Minus logarithm of the posterior

$\tilde{L}$: Minus logarithm of the likelihood

l : Length of a pipe

M : Selection matrix that corresponds model predictions to measured Degrees of Freedom

M_i : Model which has leakage at pipe i

m : Exponent in the Hazen-Williams equation

N : Number of nodes

N_s : Number of sensors

N_θ : Number of parameters to be updated

$\widetilde{N}$: Number of samples

n : Manning roughness coefficient

P : Number of pipes

p : Power of a pump

Q : Fisher information matrix

$\underline{Q}$: Vector with the sign of flow rates

Q_π : Matrix related to the prior

q : Flow rate in a pipe

$\underline{q}$: Vector with flow rates of all pipes

q_j : Flow rate in pipe j

R : Resistance coefficient in the frictional loss equation

$\hat{R}$: Fraction of the Reynolds number used the cubic interpolation from Moody diagram

Re : Reynolds number

S : Number of pseudoloops

U : Information entropy based objective function

U' : Information entropy based objective function conditioned also on nuisance parameters

u : Average velocity of fluid in a pipe

$\underline{u}^*$: Selection vector for the unknown hydraulic heads

$\underline{w}$: Vector with the total head loss of all the pipes

w_i : Sparse grid or Monte Carlo point weights

x : Flow exponent in the frictional loss equation

$X1, X2, X3, X4$: Coefficients of the cubic interpolation from Moody diagram.

$\underline{y}$: Vector of measurements

y_i : Measurement i

$Y2, Y3$: Variables used in the cubic interpolation from the Moody diagram.
Z : Selection matrix for all the pseudoloop equations
$\underline{z}$: Selection vector for a pseudoloop equation

γ : Specific gravity of the fluid
ΔH : Difference between the posterior and the prior information entropy
Δh_j : Total head loss in pipe j
Δr : Head difference between endpoints in a pseudoloop equation
$\underline{\Delta r}$: Vector with head differences for all pseudoloop equations
$\underline{\Delta \varphi}$: Perturbation of the model nuisance parameters
$\underline{\delta}$: Coordinates of sensors
$\underline{\delta_i}$: Kronecker delta vector
$\underline{\delta_{opt}}$: Optimal sensor configuration
ε : Roughness of a pipe
η : Efficiency of a pump
$\underline{\theta}$: Model parameters to be updated
$\underline{\hat{\theta}}$: Most probable values of model parameters
θ_l *:* Amount of leakage
κ : Variable used in the friction factor equation
λ_q : Variable used in the friction factor equation
μ : Viscosity of fluid
ρ : Density of fluid
Σ : Prediction error covariance matrix
$\underline{\sigma}$: Parameters of prediction error covariance matrix
$\underline{\hat{\sigma}}$: Most probable values of prediction error covariance matrix parameters
$\underline{\varphi}$: Model nuisance parameters

1. Introduction

Pipe networks such as water distribution networks are an important asset encountered in the industrialized society. The size of the networks can vary from small ones, such as in heating systems to massive ones, such as in large cities, consisted of thousands of pipes. One of the most common problems of these networks is leakage and when the size of the network is taken into consideration locating one can be difficult. Given enough time leaks can accumulate in the network resulting in significant portion of the water supply being wasted. To prevent this, a leakage detection methodology needs to be implemented to identify leaks early and provide crucial information that will guide the decision making of the operation.

Some leakage detection methodologies that have been proposed are acoustic logging [1], ground penetrating radar [2], transient-based methods [3], and model-based methods [4]. The model-based methods use experimental measurements to update the model parameters by minimizing the discrepancy between the model predictions and the measurements, resulting in a model which represents the condition of the actual system. The method that is presented here also relies on a model of the system and uses a Bayesian system identification methodology, such as [5], [6], to infer the location and the magnitude of leakage. The Bayesian framework also enables the opportunity to account for and quantify the unavoidable uncertainties presented in the model parameters and model predictions due to modeling and measurement error respectively. A set of measurements gathered from sensors placed within the network are used to update the parameters from a set of hydraulic models which represent leakage on different pipes. According to the Bayesian model selection methodology, the most probable leakage scenario corresponds to the model which maximizes the evidence of the data.

Herein, the hydraulic model of the system is also developed. It uses a system of equations comprised of the continuity and energy balance equations, [7], [8], to be solved to calculate the pressures and flow rates in the network. To make the model more efficient, by reducing the number of non-linear equations to be solved, the energy balance equations are combined accordingly to formulate new independent equations, resulting in a system of linear and a system of non-linear equations that can be solved independently. To verify the accuracy of the developed model, the commercial software EPANET [9] was used to compare the model outputs.

The quantity and quality of information provided by the measurement data depends on the number and type of sensors, as well as the location that they will be placed. It is of interest then, to develop a methodology that can find the optimal sensor configurations that maximize the information the

measurements provide about the parameters of interest. In this work Optimal Sensor Placement (OSP) methodologies are used that are based on optimizing some index of the information provided by the data. A norm of the Fisher information matrix (FIM) [10] is one such index that has been used in the past. It expresses the sensitivity of the measurements with respect to the parameters to be inferred. Maximization of the FIM can yield the optimal sensor configuration. Equivalently minimization of the Information Entropy (IE) [11] of the model parameters, indicative of leakage in the network, can yield the optimal sensor configuration. The IE quantifies the posterior uncertainty of the parameters of interest and is related to the determinant of the FIM. In order to account for the uncertainty in the measurements and the model parameters, the Bayesian framework is applied to the IE to formulate an expression that results in sensor configurations which are robust to uncertainties [12].

In this thesis the robust IE is used as the measure of uncertainty [12]. This chosen objective function contains a multidimensional integral over the model parameter space. To evaluate this, two approximations are provided. The first one is an asymptotic approximation which gives more accurate estimates as the amount of data is increased [11] and the second one is a Monte Carlo sampling approximation. The accuracy of the sampling approximation depends on the number of samples that will be drawn from the prior of the model parameters and the prediction error parameters, at the cost of more computational time. Since an exhaustive evaluation over all possible sensor configurations is computationally prohibitive, two heuristic sequential sensor placing algorithms are used [11] to estimate the optimal sensor configuration. To verify the results of the OSP methodology the resulting optimal sensor configurations are applied to detect simulated leaks. The effectiveness of the optimal configuration is measured by comparing the accuracy of locating correctly the leaking pipe and the posterior uncertainty of the leakage parameters to the accuracy achieved by suboptimal sensor configurations.

The organization of the thesis is as follows. The formulation of the hydraulic model is presented in Chapter 2. The Bayesian inference and model selection methodology used to infer the amount of leakage and locate the leaking pipe is presented in Chapter 3. In Chapter 4 the information entropy based objective function and the methodology for optimal sensor placement will be introduced. The optimal sensor placement methodology is then applied in an artificial water distribution network, which is outlined in Chapter 5, and the results are presented and discussed in Chapter 6. The effectiveness of the resulting optimal sensor configurations is evaluated and discussed in Chapter 7. Finally, conclusions are drawn in Chapter 8.

2. Hydraulic model formulation

In this chapter the steady-state equations related to the hydraulic system are formulated. These equations constitute the model of a water distribution network that will be used to calculate the flow rate in pipes and pressure in nodes, which are the model outputs. The Bayesian inference and the optimal sensor placement methodology developed in the following chapters require also the derivatives of the model outputs (flow rates and pressures) with respect to the model parameters. Thus, all the necessary equations and systems of equations for the derivates of the flow rates and pressures with respect to the model parameters are analytically formulated as well. Since the hydraulic model will be used to simulate the presence of a leaking pipe, some changes are required to be done to the original model of the network in order to model the leakage. The method that is used to incorporate the leakage is to split the pipe into two pipes and add an extra node between with nodal demand to represent the amount of leakage. This will give the model the capability to simulate the hydraulic changes that will happen in the network and by monitoring these changes, the leaking pipe will be able to be located. The hydraulic equations are based on the following fluid mechanics book [7], which are extended in this chapter to provide the equations for the derivatives of the flow rate and pressure with respect to the model parameters.

2.1 Frictional losses in developed pipe flow

When a liquid flows in pipes it is subjected to frictional losses. These losses gradually reduce the energy of the liquid or equivalently the hydraulic head of the liquid. The hydraulic head is what makes the liquid able to flow spontaneously from regions of the network with higher hydraulic head to regions with lower hydraulic head. If the altitude difference between the sources of water and the places where the water is consumed is not adequate to overcome the frictional loses and still have enough pressure where the water is consumed, then pumps are required to increase the hydraulic head of the liquid. In the following sections the types of frictional losses are discussed with greater detail.

2.1.1 Losses due to wall shear

Losses due to wall shear is a type of loss that acts along the length of pipes due to friction that arises between the internal pipe walls and the water that flows within. The head loss in a pipe that results from the wall shear in a developed flow is given by an equation of the following form

$$h_l = Rq^x \tag{1}$$

where h_l is the head loss due to wall shear, q is the flow rate and R, x are the resistance coefficient and the flow exponent respectively that depend on the exact friction formula that will be used. Some

of the formulas that can be used in equation (1) and are also used in the EPANET software [9] are the following:

- Hazen-Williams formula
- Darcy-Weisbach formula
- Chezy-Manning formula

i. Hazen-Williams

The only liquid that Hazen-Williams formula can be applied to is water.

x=1.85

$$R = \frac{K_1 l}{C^x d^m} \qquad K_1 = \begin{cases} 10.59 & \text{SI units} \\ 4.72 & \text{English units} \end{cases} \tag{2}$$

where C is the Hazen-Williams coefficient, d is the pipe diameter, l is the length of the pipe and m=4.87

ii. Darcy-Weisbach

The Darcy-Weisbach formula is the formula that is used throughout this thesis and is the most accurate of the three. It can be applied to all kinds of liquids and flow regimes.

x=2

$$R = \frac{fl}{2gdA^2} \tag{3}$$

where f is the friction factor, g is the gravitational acceleration and A is the cross sectional area of the pipe

In the cases where the pipe is circular (4) can be applied to (3) resulting to the equation (5)

$$A = \frac{\pi d^2}{4} \tag{4}$$

$$R = \frac{8fl}{g\pi^2 d^5} \tag{5}$$

The friction factor f is a dimensionless shear and depends on various flow related quantities. A dimensional analysis shows that f is a function of the Reynolds number and the pipe relative roughness.

$$f = f\left(Re, \frac{\varepsilon}{d}\right) \tag{6}$$

where Re is the Reynolds number defined in (7), ε is the roughness of the pipe measured in units of length and ε/d is the relative pipe roughness

$$Re = \frac{|u|\rho d}{\mu} \tag{7}$$

where u is the average velocity of the liquid, ρ is the density of the liquid and μ is the viscosity of the liquid. The absolute value is used for the average velocity because in the software the sign of the flow rate ($q \propto u$) corresponds to the relative direction in which the liquid is flowing in a pipe. The Reynolds number can be written as a function of the flow rate in a circular pipe using (4) and (8) to get equation (9).

$$q = uA \tag{8}$$

$$Re = \frac{4\rho|q|}{\pi d\mu} \tag{9}$$

Values of f as a function to the Reynolds number and relative pipe roughness can be found on the Moody diagram. However, in this case the friction factor will be calculated by the software, so an analytical expression needs to be used. A significant number of empirical equations have been developed overtime to represent the different regimes of the Moody diagram with varying degrees of accuracy. The different regimes of the Moody diagram are comprised of laminar flow, transitional flow and turbulent flow, which can be either smooth or rough turbulent flow. EPANET [9] in order to have accuracy over all these regimes uses three different functions of f, one for each flow regime. The equations are the following.

- Hagen-Poiseuille formula for $Re < 2000$

$$f = \frac{64}{Re} \tag{10}$$

- Swamee and Jain approximation to the Colebrook-White equation for $Re > 4000$

$$f = \frac{1.325}{\left[\ln\left(\frac{\varepsilon}{3.7d} + \frac{5.74}{Re^{0.9}}\right)\right]^2} \tag{11}$$

Note: Equation (11) in the EPANET manual has 0.25 in the numerator which is the value that will arise if the natural logarithm in the denominator is changed to base 10 logarithm.

- Cubic interpolation from the Moody diagram for $2000 < Re < 4000$

$$f = \left(X1 + \hat{R}\left(X2 + \hat{R}(X3 + X4)\right)\right) \tag{12}$$

$$\hat{R} = \frac{Re}{2000} \tag{13}$$

$$X1 = 7FA - FB \tag{14}$$

$$X2 = 0.128 - 17FA + 2.5FB \tag{15}$$

$$X3 = -0.128 + 13FA - 2FB \tag{16}$$

$$X4 = \hat{R}(0.032 - 3FA + 0.5FB) \tag{17}$$

$$FA = (Y3)^{-2} \tag{18}$$

$$FB = FA\left(2 - \frac{0.00514215}{(Y2)(Y3)}\right) \tag{19}$$

$$Y2 = \frac{\varepsilon}{3.7d} + \frac{5.74}{Re^{0.9}} \tag{20}$$

$$Y3 = -0.86859 \ln\left(\frac{\varepsilon}{3.7d} + \frac{5.74}{4000^{0.9}}\right) \tag{21}$$

The problem that was encountered with this set of equations is that the second derivative with respect to flow rate was not continuous at the intersection points and caused problems when it was used by the numerical scheme to solve the system of non-linear hydraulic equations.

Instead, a single more complex equation was used that combined expressions for all flow regimes and pipe roughness.

- Cheng formula [13]

$$f = \left(\frac{64}{Re}\right)^{a}\left(1.8\log\frac{Re}{6.8}\right)^{2(a-1)b}\left(2\log\frac{3.7d}{\varepsilon}\right)^{2(a-1)(1-b)} \tag{22}$$

$$a = \frac{1}{1 + \left(\frac{Re}{2720}\right)^{9}} \tag{23}$$

$$b = \frac{1}{1 + \left(\frac{Re}{160\frac{d}{\varepsilon}}\right)^{2}} \tag{24}$$

iii. <u>Chezy-Manning</u>

The Chezy-Manning formula is mostly used for open channel flows.

x=2

$$R = \frac{10.29n^2 l}{K_2 d^{5.33}} \quad K_2 = \begin{cases} 1 & \text{SI units} \\ 2.22 & \text{English units} \end{cases} \tag{25}$$

where n is the Manning roughness coefficient

2.1.2 Minor losses

Minor losses are a type of loss that occurs due to the presence of different components in the pipes. These components cause discrete reductions in the hydraulic head of the liquid and are caused by secondary flows or flow separation. Some of the components that pipes have and can cause minor losses are bends, elbows, valves, changes in pipe diameter and other fittings. The minor losses are given by the equation

$$h_f = \frac{8K}{g\pi^2 d^4} q^2 \tag{26}$$

where h_f is the head loss due to the presence of a component and K is the loss coefficient of the component, which is calculated experimentally and can be found in tables [14]

2.1.3 Pumps

Pumps are devices which are added to water distribution networks to increase the hydraulic head of the liquid. This is done to overcome energy losses due to friction, increase the pressure of water at the consumption points or make the water go to higher altitudes than the altitude of the source. For large networks where the total length of the pipes that the water must travel is significant, multiple pumps are usually required to overcome the losses. Pumps, for the purpose of this thesis, will be considered as negative losses and added to the total energy loss of the pipe. If the exact characteristics of the pump are known, then the pump curve is also available. When this is the case the head loss of the pump (negative term) can be calculated using a three-point curve with 3 parameters that can be calculating by fitting 3 points from the pump curve.

$$h_p = -\left(h_m - h_1 q^{h_2} \right) \tag{27}$$

where h_p is the head loss due to the pump, h_m is the shutoff head of the pump and h_1, h_2 are parameters that are determined by fitting the equation to the pump curve.

If the pump curve is not available, which is usually the case in the preliminary design of the network, then the following equation can be used instead which assumes a constant power pump.

$$h_p = -\frac{p\eta}{q\gamma} \tag{28}$$

where p is the power of the pump, η is the efficiency of the pump and γ is the specific gravity of the fluid

$$\gamma = \rho g \tag{29}$$

2.2 System of hydraulic equations

The unknown variables of the hydraulic model that must be calculated are the flow rate in each pipe and the pressure in each interior node or junction. Junction is every node that is not a reservoir. Reservoirs are fixed grade nodes and their hydraulic head is known. In order to calculate these quantities, a system of equations has to be formulated. The equations that will be used are the continuity equation in junctions and the energy balance in pipes.

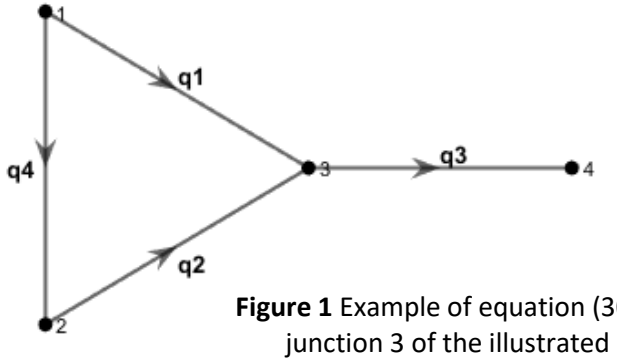
Let P be the number of pipes, J be the number of interior nodes or junctions and N the number of all the nodes. The number of unknown variables is equal to $P + J$ for the flow rate in P pipes and the pressure in J junctions. The continuity and energy balance equations are formulated below.

2.2.1 Continuity balance

The continuity balance specifies that the total inflow of water in a junction has to be equal to the sum of total outflow plus the consumption at the junction. The mathematical notation of this description is the following

$$\underline{c_i}^T \underline{q} = \tilde{d}_i \tag{30}$$

where $\underline{q} \in \mathbb{R}^P$ is a vector with the flow rates of every pipe, $\tilde{d}_i$ is the demand of the junction i and $\underline{c_i} \in \mathbb{R}^P$, $i = 1, \dots, J$ is a selection vector that corresponds the junction i to all the pipes that are directly connected to it and has values -1, 0, 1. The value -1 is used when the pipe that corresponds to the specific element of the vector has junction i as start node and thus removes water from this junction, the definition of pipes with start and end nodes is addressed in the next paragraph. The value 1 is used when the respective pipe has junction i as end node and thus adds water to this node. If a pipe is not connected directly to junction i then the respective element of the vector has the value 0. An example of this is illustrated in Figure 1, the arrows showcase the positive flow direction in the pipes.



$$[1, 1, -1, 0] \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \tilde{d}_3 \Rightarrow$$

$$q1 + q2 - q3 = \tilde{d}_3$$

Figure 1 Example of equation (30) applied to junction 3 of the illustrated network

Each pipe is defined with a start node and an end node. When the water flows from the start to the end node the flow is considered positive in that pipe. When the water flows in the opposite direction the flow is considered negative. If a pipe contains a pump, then the start and end nodes are defined accordingly so that the outflow of the pump is in the positive direction.

Applying equation (30) to every junction results to the following form

$$C\underline{q} = \underline{\tilde{d}} \tag{31}$$

where C is a matrix defined as $C = \left[\underline{c}_1, \dots, \underline{c}_J\right]^T$ and $\underline{\tilde{d}}$ is a vector containing all the demands $\underline{\tilde{d}} = \left[\tilde{d}_1, \dots, \tilde{d}_J\right]^T$

2.2.2 Energy balance

The total energy loss in a pipe, or equivalently the head loss, is equal to the sum of the three loss terms formulated in the previous section.

$$\Delta h_j = sign(q_j)\left[h_{l,j} + h_{f,j} + h_{p,j}\right] \tag{32}$$

where Δh_j is the head loss in pipe j and is defined as:

$$\Delta h_j = \underline{e}_j^T \underline{h} \tag{33}$$

where $\underline{h} \in \mathbb{R}^N$ is a vector that contains the hydraulic head of all the nodes, reservoirs included and $\underline{e}_j \in \mathbb{R}^N$ $j = 1, \dots P$ is a selection vector that corresponds pipe j to its start and end nodes and has values -1,0,1. The value -1 is used when the node that corresponds to the specific element of the vector is the end node of pipe j. The value 1 is used when the respective node is the start node of pipe j and 0 is used if the node is not connected directly to pipe j.

The head loss in a pipe is defined as the head in the start node minus the head in the end node. In a pipe with the water flowing in the positive direction the hydraulic head in the start node is higher than the head in the end node and the head loss has to be a positive quantity for a positive water flow. If the flow turns out to be negative, water goes from the end node towards the start node and the head loss is negative. The loss terms as are formulated are always positive quantities, so the $sign(\dots)$ function is used to correct this. If a pump does not exist in the pipe j the term $h_{p,j}$ is 0. The $sign(q)$ function is defined as

$$sign(q) = \begin{cases} -1, & q < 0 \\ 1, & q > 0 \\ 0, & q = 0 \end{cases} \tag{34}$$

Applying equation (32) to every pipe results to the following form

$$E\underline{h} = \underline{Q}.*\left(\underline{h}_l + \underline{h}_f + \underline{h}_p\right) \tag{35}$$

where E is a matrix defined as $E = \left[\underline{e}_1, \dots, \underline{e}_P\right]^T$, $\underline{h}_l$, $\underline{h}_f$, $\underline{h}_p$ are vectors which contain the loss term for every pipe and $\underline{Q} = sign\left(\underline{q}\right)$

The period character (.) in the equation (35), which is also used throughout this thesis, is equivalent to the period (.) used in MATLAB to distinguish between array operations and matrix operations. When the period is used before an operator, it means that the operator that follows will execute an element by element operation on the corresponding elements of vectors or matrices.

The way the equation (35) is formulated however, is not the usual $Ax = b$ linear system form. The problem is that some of the values of the vector $\underline{h}$ are already known, since it contains the head of the fixed grade nodes as well, which have known hydraulic head. In order to fix this the following equation is used instead to calculate the hydraulic head.

$$E(:, \underline{u}^*)\underline{h}(\underline{u}^*) = \underline{Q}.*\left(\underline{h}_l + \underline{h}_f + \underline{h}_p\right) - E(:, \underline{k}^*)\underline{h}(\underline{k}^*) \tag{36}$$

where the $E(\dots, \dots)\underline{h}(\dots)$ notation is the notation used in MATLAB where specific elements of the matrix can be selected using a selection vector in each coordinate. $\underline{u}^*$ is a selection vector that selects all the unknown hydraulic heads, $\underline{k}^*$ is a vector that selects all the known hydraulic heads and the (:) notation denotes all rows of the matrix.

The equations (31) and (36), for all junctions and all pipes respectively constitute a system of $P + J$ non-linear equations that can be solved to calculate the flow rates and the hydraulic heads in the network. In order to calculate the static pressure of each junction the altitude of each junction has to be subtracted from its hydraulic head. The hydraulic head is a sum of the static pressure of the water, in units of length, and the height from a plane of reference (sea level) which corresponds to the potential energy of the liquid.

2.2.3 Formulating a more efficient system of equations

By solving the system of equations formulated above it is possible to get the hydraulic quantities of interest. However, this is inefficient since the system contains only non-linear equations. It is possible to alter the equations in such a way that the number of non-linear equations that must be solved will be reduced. The number of non-linear equations that will be removed will be replaced by the same number of linear equations that can be solved independently of the non-linear ones. If attention is

paid to the equation (36) it can be noticed that it is a linear system of equations with respect to the hydraulic head and if the flow rates were known the hydraulic head could be calculated by solving this linear system. In order to calculate the flow rates, the network will be split into loops (closed paths) and pseudoloops (paths between fixed grade nodes) and use the fact that loops have zero head gain at the end points (an arbitrarily chosen node of the loop) and pseudoloops have known head at the end points. In this way the unknown head terms can be canceled out from equation (36) and make the system of equations for calculating flow rates independent of calculating the pressure head.

i. Energy balance in loops

Loops are closed paths in the network. In other words, by following a loop the starting node (arbitrarily chosen) can be reached without crossing any pipe more than once. In order to find the loops in the network the DFS algorithm is utilized [15], [16].

Let L be the number of independent loops in the network (resulting in linearly independent equations) which is given by

$$L = P - N + 1 \tag{37}$$

and $\underline{w} \in \mathbb{R}^P$ be a vector containing the total head loss of every pipe defined as follows

$$\underline{w} = \underline{Q}.*\left(\underline{h}_l + \underline{h}_f + \underline{h}_p\right) \tag{38}$$

The energy balance around a single loop can be written as

$$\underline{b}^T \underline{w} = \underline{0} \tag{39}$$

where $\underline{b} \in \mathbb{R}^P$ is a selection vector with values -1,0,1. The value -1 is used when the pipe that corresponds to the specific element of the vector has alignment opposite to the alignment of the loop. The value 1 is used when the respective pipe has the same alignment as the loop and 0 is used when the respective pipe is not part of the loop. An example is illustrated in Figure 2. The alignment of the loop is arbitrarily defined. The same equation written for all the loops is the following.

$$B\underline{w} = \underline{0} \tag{40}$$

where B is a matrix defined as $B = \left[\underline{\mathrm{b}}_1, \ldots, \underline{\mathrm{b}}_L\right]^T$

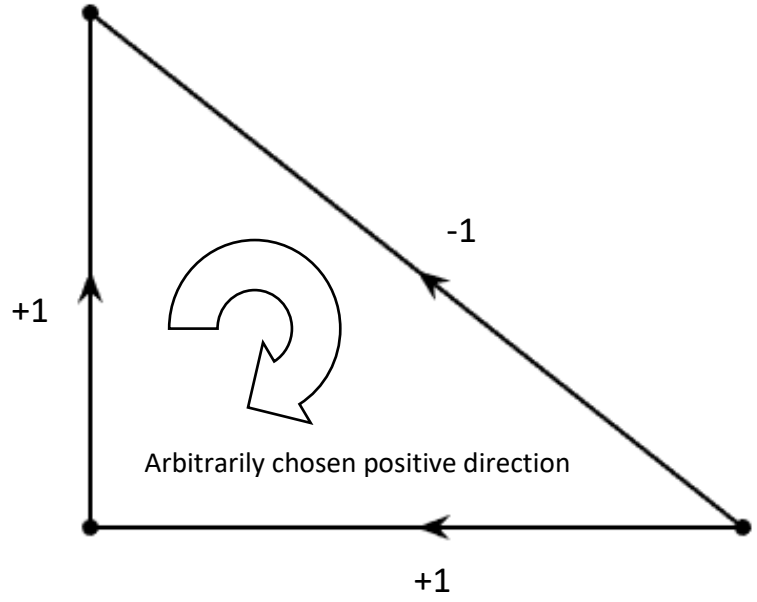


Figure 2 Loop consisted of three pipes illustrating the vector values

ii. Energy balance in pseudoloops

Pseudoloops are paths between fixed grade nodes (reservoirs). The hydraulic head of the fixed grade nodes is known so the head difference between the start fixed grade node of the pseudoloop and the end fixed grade node is known as well. This is needed in order to avoid the yet unknown head terms. The pseudoloops that are used in the software have a specific arbitrarily chosen fixed grade node as the start node and the rest of the fixed grade nodes become the end node of the pseudoloop one at a time, resulting to number of pseudoloops equal to the number of fixed grade nodes minus one. This selection of pairs of fixed grade nodes for the start and end nodes is not unique, other pairs can be chosen as long as the resulting pseudoloop equations are linearly independent [15]. The path between the start node and the end node of the pseudoloops is found using the MATLAB *shortestpath* function. Let S be the number of pseudoloops in the network and F the number of fixed grade nodes. Note that

$$N = F + J \tag{41}$$

The number of independent pseudoloops in a network (resulting in linearly independent equations) is

$$S = F - 1 \tag{42}$$

The energy balance in a single pseudoloop can be written as

$$\underline{z}^T \underline{w} + \Delta r = 0 \tag{43}$$

where $\underline{z} \in \mathbb{R}^P$ is a selection vector with values -1,0,1. The value -1 is used when the pipe that corresponds to the specific element of the vector has alignment opposite to the alignment of the pseudoloop. The value 1 is used when the respective pipe has the same alignment as the pseudoloop and 0 is used when the respective pipe is not part of the pseudoloop. An example is illustrated in Figure 3. Δr is the difference between the head of the start fixed grade node of the pseudoloop and the head of the end fixed grade node.

The same equation written for all the pseudoloops is the following.

$$Z\underline{w} + \underline{\Delta r} = \underline{0} \tag{44}$$

where Z is a matrix defined as $Z = \left[z_1, \dots, \underline{z}_S\right]^T$ and $\underline{\Delta r}$ is a vector containing the head differences of the endpoints of the pseudoloops being used.

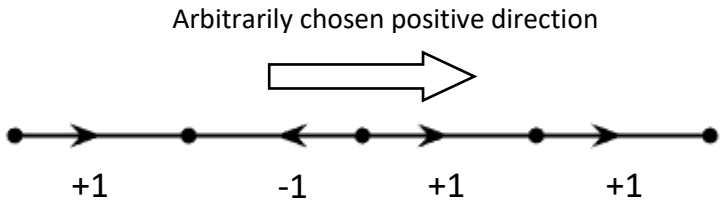


Figure 3 Pseudoloop consisted of four pipes illustrating the vector values

To sum up, in order to calculate the P unknown flow rates in the pipes the system of non-linear equations (31), (40) and (44) with J, L and S equations respectively has to be solved simultaneously using a numerical scheme. It can be shown using equations (37), (41) and (42) that $P = J + L + S$. In the developed software the numerical method that is used is the trust-region-dogleg algorithm which can be found in the *fsolve* function in MATLAB. Now that the flow rates are known, the linear equation (36) can be used to calculate the hydraulic head of each junction. In order to calculate the pressure of the junctions as well, as mentioned before the altitude must be subtracted from the total hydraulic head resulting in the junction pressure given in the same units as the hydraulic head.

2.3 Introducing a leaking pipe into the network

In order to perform leakage detection, a leaking pipe has to be introduced to the model. The method that is used to introduce the leakage is the following. First, a pipe that the leakage will be introduced is selected. The pipe then is split into two similar pipes with each one of them having the same roughness and diameter and half the length of the original pipe. If the chosen pipe has a pump or components that cause minor losses, then these are only added to one of the two resulting pipes. Then, at the intersection of the two new pipes a new node is introduced. This new node has a demand assigned into it that is equal to the amount of leakage that needs to be introduced. An example of this is illustrated in Figure 4. This results into a model that has one more pipe and one more node than the original network. The numbering of the new node is one more than the previous highest node number and the numbering of the new pipe is one more than the previous highest pipe number. When a leakage is added using this method, the vectors and matrices used to model the network have to be altered accordingly. These new augmented vectors and matrices, if used in solving the hydraulic equations formulated above, will result in model outputs that will consider the presence of a leaking pipe. The matrices used in the following chapters are assumed to be augmented with the extra pipe and node resulting from adding the leakage. For simplicity the same symbols will be used to describe them.

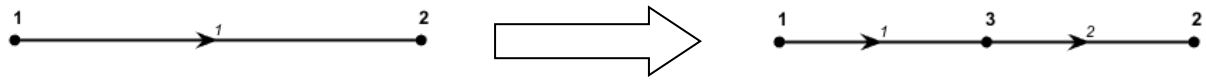


Figure 4 Splitting a pipe in two and adding a node

2.4 Sensitivities of hydraulic equations

In the following chapters, the Bayesian parameter inference framework is applied for leakage detection and the information theory is used to find the locations where if sensors are placed, will result in very informative measurements. The application of these methodologies requires the use of derivatives of system outputs (flow rates and pressures) with respect to the model parameters. Also, derivatives are required in the trust-region-dogleg algorithm used for solving the system of hydraulic equations. This part of the chapter will therefore present the necessary equations that will be used to calculate these derivatives. It is noted that the following derivatives can also be calculated numerically, however this will increase the required running time of the software significantly, since the hydraulic system will be solved a considerable amount of times and is therefore discouraged.

2.4.1 Sensitivities of friction factor

Firstly, the derivatives of the Cheng friction factor with respect to flow rate will be formulated.

2.4.1.1 Cheng friction factor

i. 1st derivative of the Cheng friction factor with respect to flow rate q

The Cheng friction factor was introduced in equation (22)

For simplicity the equation is split into three terms.

$$f = f_1 f_2 f_3 \tag{45}$$

where

$$f_1 = \left(\frac{64}{Re}\right)^a \tag{46}$$

$$f_2 = \left(1.8 \log \frac{Re}{6.8}\right)^{2(a-1)b} \tag{47}$$

$$f_3 = \left(2 \log \frac{3.7d}{\varepsilon}\right)^{2(a-1)(1-b)} \tag{48}$$

Formulating the 1st derivatives

$$\frac{\partial f}{\partial q} = \frac{\partial f_1}{\partial q} f_2 f_3 + f_1 \frac{\partial f_2}{\partial q} f_3 + f_1 f_2 \frac{\partial f_3}{\partial q} \tag{49}$$

where

$$\frac{\partial f_1}{\partial q} = f_1 \left(\frac{\partial a}{\partial q} \ln \frac{64}{Re} - a Re^{-1} \frac{\partial Re}{\partial q}\right) \tag{50}$$

$$\frac{\partial f_2}{\partial q} = 2 f_2 \left[\left(\frac{\partial a}{\partial q} b + (a-1)\frac{\partial b}{\partial q}\right) \ln\left(1.8 \log \frac{Re}{6.8}\right) + \frac{(a-1)b}{Re \ln 10 \log \frac{Re}{6.8}} \frac{\partial Re}{\partial q}\right] \tag{51}$$

$$\frac{\partial f_3}{\partial q} = 2 f_3 \left(\frac{\partial a}{\partial q}(1-b) - (a-1)\frac{\partial b}{\partial q}\right) \ln\left(2 \log \frac{3.7d}{\varepsilon}\right) \tag{52}$$

$$\frac{\partial a}{\partial q} = -\frac{9}{2720^9} a^2 Re^8 \frac{\partial Re}{\partial q} \tag{53}$$

$$\frac{\partial b}{\partial q} = -\frac{2}{\left(160 \frac{d}{\varepsilon}\right)^2} b^2 Re \frac{\partial Re}{\partial q} \tag{54}$$

$$\frac{\partial Re}{\partial q} = \frac{4\rho \; sign(q)}{\pi d \mu} \tag{55}$$

A problem that was encountered with the first derivative of the Cheng friction factor is that in equation (51) if $Re < 6.8$ the quantity $\log(Re/6.8)$ is negative which is then used as input

in a logarithm resulting in complex values. This problem is resolved if the term on the left of the logarithm is 0, since the product will be 0 eliminating the complex values. The way the Cheng friction factor is formulated uses the a and b parameters as weights to choose which terms should be used depending on the flow regime and smooth or rough turbulent flow respectively. Parameter a approaches 1 for laminar flow and for $Re < 6.8$ can be considered as exactly 1 thus making the $(a-1)$ term to be 0. In this case the term $\partial a/\partial q$ should also be 0 since the rate of change is insignificant. However, this is not the case. If $Re = 6.8$ is used to calculate the $\partial a/\partial q$ derivative, the result is of the order 10^{-18} thus not eliminating the complex term. The solution applied to this is that if $Re < 6.8$ the $\partial a/\partial q$ term is forced to be 0.

ii. 2nd derivative of the Cheng friction factor with respect to flow rate q

Formulating the 2nd derivatives

$$\frac{\partial^2 f}{\partial q^2} = \frac{\partial^2 f_1}{\partial q^2} f_2 f_3 + f_1 \frac{\partial^2 f_2}{\partial q^2} f_3 + f_1 f_2 \frac{\partial^2 f_3}{\partial q^2} + 2\frac{\partial f_1}{\partial q}\frac{\partial f_2}{\partial q} f_3 + 2\frac{\partial f_1}{\partial q} f_2 \frac{\partial f_3}{\partial q} + 2 f_1 \frac{\partial f_2}{\partial q}\frac{\partial f_3}{\partial q} \tag{56}$$

$$\frac{\partial^2 f_1}{\partial q^2} = \frac{\partial f_1}{\partial q}\left(\frac{\partial a}{\partial q}\ln\frac{64}{Re} - \frac{a}{Re}\frac{\partial Re}{\partial q}\right) + f_1\left(\frac{\partial^2 a}{\partial q^2}\ln\frac{64}{Re} - 2\frac{\partial a}{\partial q} Re^{-1}\frac{\partial Re}{\partial q} + a\, Re^{-2}\left(\frac{\partial Re}{\partial q}\right)^2\right) \tag{57}$$

$$\begin{aligned}\frac{\partial^2 f_2}{\partial q^2} = {} & 2\frac{\partial f_2}{\partial q}\left[\left(\frac{\partial a}{\partial q} b + (a-1)\frac{\partial b}{\partial q}\right)\ln\left(1.8\log\frac{Re}{6.8}\right) + \frac{(a-1)b}{Re\ln 10\log\frac{Re}{6.8}}\frac{\partial Re}{\partial q}\right] \\ & + 2 f_2\left[\ln\left(1.8\log\frac{Re}{6.8}\right)\left(\frac{\partial^2 a}{\partial q^2} b + 2\frac{\partial a}{\partial q}\frac{\partial b}{\partial q} + (a-1)\frac{\partial^2 b}{\partial q^2}\right)\right. \\ & + \left(\frac{\partial a}{\partial q} b + (a-1)\frac{\partial b}{\partial q}\right)\left(1.8\log\frac{Re}{6.8}\right)^{-1}\frac{1.8}{Re\ln 10}\frac{\partial Re}{\partial q} \\ & + \frac{\partial Re}{\partial q}\left(\frac{\partial a}{\partial q}\frac{b}{Re\ln 10\log\frac{Re}{6.8}} + \frac{\partial b}{\partial q}\frac{a-1}{Re\ln 10\log\frac{Re}{6.8}} - Re^{-2}\frac{(a-1)b}{\ln 10\log\frac{Re}{6.8}}\frac{\partial Re}{\partial q}\right. \\ & \left.\left. - \log^{-2}\frac{Re}{6.8}\ \frac{(a-1)b}{(Re\ln 10)^2}\frac{\partial Re}{\partial q}\right)\right]\end{aligned} \tag{58}$$

$$\frac{\partial^2 f_3}{\partial q^2} = 2\frac{\partial f_3}{\partial q}\left(\frac{\partial a}{\partial q}(1-b) - (a-1)\frac{\partial b}{\partial q}\right)\ln\left(2\log\frac{3.7d}{\varepsilon}\right) + 2f_3\left(\frac{\partial^2 a}{\partial q^2}(1-b) - 2\frac{\partial a}{\partial q}\frac{\partial b}{\partial q} - (a-1)\frac{\partial^2 b}{\partial q^2}\right)\ln\left(2\log\frac{3.7d}{\varepsilon}\right) \quad (59)$$

$$\frac{\partial^2 a}{\partial q^2} = \frac{-9}{2720^9}\left(2aRe^8\frac{\partial a}{\partial q}\frac{\partial Re}{\partial q} + 8a^2Re^7\left(\frac{\partial Re}{\partial q}\right)^2\right) \quad (60)$$

$$\frac{\partial^2 b}{\partial q^2} = -\frac{2}{\left(160\frac{d}{\varepsilon}\right)^2}\left(2bRe\frac{\partial b}{\partial q}\frac{\partial Re}{\partial q} + \left(b\frac{\partial Re}{\partial q}\right)^2\right) \quad (61)$$

A problem with complex numbers was encountered in the second derivative as well. The quantity $\partial^2 a/\partial q^2$ was also forced to be 0 to resolve the complex numbers from $\partial^2 f_2/\partial q^2$. The maximum value that $\partial^2 a/\partial q^2$ can get is for $Re = 6.8$ and is of the order of magnitude 10^{-12}.

2.4.1.2 EPANET friction factors

i. 1st derivative of the EPANET friction factors with respect to flow rate q

For simplicity, let κ and λ_q be variables defined as follows

$$\kappa = \frac{\pi d\mu}{4\rho} \quad (62)$$

$$\lambda_q = \frac{\varepsilon}{3.7d} + 5.74\left(\frac{\kappa}{|q|}\right)^{0.9} \quad (63)$$

The first derivatives of the three equations used for calculating the friction factor by EPANET for each flow regime are the following

$$\frac{\partial f}{\partial q} = -\frac{64\kappa}{sign(q)q^2}, \quad Re < 2000 \quad (64)$$

$$\frac{\partial f}{\partial q} = 13.6899\left(\ln\lambda_q\right)^{-3}\lambda_q^{-1}\kappa^{0.9}|q|^{-1.9}sign(q), \quad Re > 4000 \quad (65)$$

For the transitional regime the following constants are introduced

$$a_1 = 0.00514215, \qquad b_1 = 0.86859$$

$$\frac{\partial f}{\partial q} = \frac{\partial X1}{\partial q} + \frac{\partial \hat{R}}{\partial q}\left(X2 + \hat{R}(X3+X4)\right) + \hat{R}\left[\frac{\partial X2}{\partial q} + \frac{\partial \hat{R}}{\partial q}(X3+X4) + \hat{R}\left(\frac{\partial X3}{\partial q} + \frac{\partial X4}{\partial q}\right)\right], \qquad 2000 < Re < 4000 \quad (66)$$

where

$$\frac{\partial X1}{\partial q} = -\frac{\partial FB}{\partial q} \tag{67}$$

$$\frac{\partial X2}{\partial q} = 2.5\frac{\partial FB}{\partial q} \tag{68}$$

$$\frac{\partial X3}{\partial q} = -2\frac{\partial FB}{\partial q} \tag{69}$$

$$\frac{\partial X4}{\partial q} = \frac{\partial \hat{R}}{\partial q}(0.032 - 3FA + 0.5FB) + 0.5\hat{R}\frac{\partial FB}{\partial q} \tag{70}$$

$$\frac{\partial \hat{R}}{\partial q} = \frac{sign(q)}{2000\kappa} \tag{71}$$

$$\frac{\partial Re}{\partial q} = \frac{sign(q)}{\kappa} \tag{72}$$

$$\frac{\partial FB}{\partial q} = FA\left(\frac{a_1}{(Y_2)^2 Y_3}\frac{\partial Y_2}{\partial q}\right) \tag{73}$$

$$\frac{\partial Y_2}{\partial q} = -5.166Re^{-1.9}\frac{\partial Re}{\partial q} \tag{74}$$

ii. 2nd derivative of the EPANET friction factors with respect to flow rate q

The second derivatives of the three equations used for calculating the friction factor by EPANET for each flow regime are the following

$$\frac{\partial^2 f}{\partial q^2} = \frac{128\kappa}{|q|^3}, \quad Re < 2000 \tag{75}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial q^2} = {} & 13.6899\kappa^{0.9}\left(\ln\left(\lambda_q\right)\right)^{-3}\Big[15.498\left(\ln\left(\lambda_q\right)\right)^{-1}\left(\lambda_q\right)^{-2}|q|^{-3,8}\kappa^{0.9} \\ & + 5.166\left(\lambda_q\right)^{-2}|q|^{-3.8}\kappa^{0.9} - 1.9\left(\lambda_q\right)^{-1}|q|^{-2.9}\Big], \quad Re > 4000\end{aligned} \tag{76}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial q^2} = {} & \frac{\partial^2 X1}{\partial q^2} + 2\frac{\partial \hat{R}}{\partial q}\left[\frac{\partial X2}{\partial q} + \frac{\partial \hat{R}}{\partial q}(X3 + X4) + \hat{R}\left(\frac{\partial X3}{\partial q} + \frac{\partial X4}{\partial q}\right)\right] \\ & + \hat{R}\left[\frac{\partial^2 X2}{\partial q^2} + 2\frac{\partial \hat{R}}{\partial q}\left(\frac{\partial X3}{\partial q} + \frac{\partial X4}{\partial q}\right) + \hat{R}\left(\frac{\partial^2 X3}{\partial q^2} + \frac{\partial^2 X4}{\partial q^2}\right)\right], \\ & 2000 < Re < 4000\end{aligned} \tag{77}$$

where

$$\frac{\partial^2 X1}{\partial q^2} = -\frac{\partial^2 FB}{\partial q^2} \tag{78}$$

$$\frac{\partial^2 X2}{\partial q^2} = 2.5\frac{\partial^2 FB}{\partial q^2} \tag{79}$$

$$\frac{\partial^2 X3}{\partial q^2} = -2\frac{\partial^2 FB}{\partial q^2} \tag{80}$$

$$\frac{\partial^2 X4}{\partial q^2} = \frac{\partial \hat{R}}{\partial q}\frac{\partial FB}{\partial q} + 0.5\hat{R}\frac{\partial^2 FB}{\partial q^2} \tag{81}$$

$$\frac{\partial^2 FB}{\partial q^2} = FA\frac{a_1}{Y3}\left(\frac{1}{Y2^2}\frac{\partial^2 Y2}{\partial q^2} - \frac{2}{Y2^3}\left(\frac{\partial Y2}{\partial q}\right)^2\right) \tag{82}$$

$$\frac{\partial^2 Y2}{\partial q^2} = 9.8154\mathrm{Re}^{-2.9}\frac{\partial Re}{\partial q} \tag{83}$$

2.4.2 Sensitivities of the model outputs with respect to the amount of leakage

In this part, the system of equations required to calculate the derivatives of the model outputs (flow rate and pressure) with respect to the amount of leakage is formulated. These derivatives are required in order to quantify the uncertainty of the parameter that is inferred through the Bayesian framework for uncertainty quantification. The parameter that is inferred in this case is the amount of leakage that escapes through a damaged pipe in the network. The calculated uncertainty, along with other quantities that will be introduced later, will be used to find which pipe is the most probable to be damaged. Moreover, these derivatives are required in calculating the Fisher information matrix. This is a quantity that is used in Chapter 4 to evaluate the effectiveness of each sensor configuration, thus making it possible to find which sensor configurations are optimal.

2.4.2.1 First derivatives of the model outputs with respect to the amount of leakage

i. Continuity equation

The first derivative of the continuity equation (31) with respect to the amount of leakage is

$$C\frac{\partial \underline{q}}{\partial \theta_l} = \underline{\delta_i} \tag{84}$$

where θ_l is the amount of leakage introduced at the new node i and δ_i is a Kronecker delta vector with all elements being 0 except the i^{th} element which has value 1 and is illustrated in (85)

$$\underline{\delta_i} = \begin{Bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{Bmatrix} \leftarrow i^{th} row \tag{85}$$

The i^{th} row is the row which corresponds to the new node that was added in order to introduce the leakage.

ii. Energy balance in loops

For simplicity the following three terms are introduced.

$$DP_l = \frac{8l}{g\pi^2 d^5} \tag{86}$$

$$DP_f = \frac{8K}{g\pi^2 d^4} \tag{87}$$

$$DP_p = \frac{p\eta}{\gamma} \tag{88}$$

These variables contain all non-flow-rate related variables from the three loss terms formulated in chapter 2.1 and are known before the system is solved. When they are used as vectors, they denote the term on the right hand side for all pipes.

The first derivative of the energy balance in loops, equation (40), with respect to the amount of leakage is

$$\left[B.*\left(\partial \underline{w}./\partial \underline{q}\right)^T\right]\frac{\partial \underline{q}}{\partial \theta_l} = \underline{0} \tag{89}$$

where

$$\partial \underline{w}./\partial \underline{q} = Q.*\left(\underline{DP_l}.*\left(\underline{q}.^2 .*\left(\partial \underline{f}./\partial \underline{q}\right) + 2\underline{f}.*\underline{q}\right) + 2\underline{DP_f}.*\underline{q} + \underline{DP_p}.*\underline{q}.^{-2}\right) \tag{90}$$

All operators preceded by the period character (.) are element by element operators. The operator (.*) in equation (89) denotes that matrix B has each row multiplied element by element by the row vector $\left(\partial \underline{w}./\partial \underline{q}\right)^T$. It is noted that the matrix in the square brackets in (89) is known since the flow rate has already been solved for.

iii. Energy balance in pseudoloops

The first derivative of the energy balance in pseudoloops, equation (44), with respect to the amount of leakage is

$$\left[Z.*\left(\partial\underline{w}./\partial\underline{q}\right)^{T}\right]\frac{\partial\underline{q}}{\partial\theta_l}=\underline{0} \tag{91}$$

where $\partial\underline{w}./\partial\underline{q}$ is defined in equation (90)

Now that the necessary equations are formulated the derivatives of the flow rate with respect to the amount of leakage can be calculated. The system of equations that has to be solved simultaneously is comprised of the equations (84), (89) and (91). It is noted that this system of equations is linear since all the quantities used to calculate the matrix on the left side are known, with the only unknown quantity being the derivatives which are linearly related.

In order to calculate the first derivative of the pressure (same result as the hydraulic head derivative) at junctions with respect to the amount of leakage, the equation (36) will be used to get the following

$$E\left(:,\underline{u}^{*}\right)\frac{\partial\underline{h}\left(\underline{u}^{*}\right)}{\partial\theta_i}=\frac{\partial\underline{w}}{\partial\theta_l} \tag{92}$$

using the chain rule

$$\frac{\partial\underline{w}}{\partial\theta_l}=\left(\partial\underline{w}./\partial\underline{q}\right).*\frac{\partial\underline{q}}{\partial\theta_l} \tag{93}$$

All the terms on the right-hand side of equation (93) are known. The set of equations (92) constitute a linear system of algebraic equations for the unknown derivatives.

2.4.2.2 Second derivatives of the model outputs with respect to the amount of leakage

i. Continuity equation

The second derivative of the continuity equation (31) with respect to the amount of leakage is

$$C\frac{\partial^2\underline{q}}{\partial\theta_l^2}=\underline{0} \tag{94}$$

ii. Energy balance in loops

The second derivative of the energy balance in loops, equation (40), with respect to the amount of leakage is

$$\left[B.*\left(\partial\underline{w}./\partial\underline{q}\right)^{T}\right]\frac{\partial^{2}\underline{q}}{\partial\theta_{l}^{2}}=-\left[B.*\left(\partial^{2}\underline{w}./\partial\underline{q}^{2}\right)^{T}\right]\left(\frac{\partial\underline{q}}{\partial\theta_{l}}\right).^{2} \tag{95}$$

where

$$\begin{aligned}\partial^{2}\underline{w}./\partial\underline{q}^{2}=Q.*\Big(&\underline{DP_{l}}.*\Big(2\left(\partial\underline{f}./\partial\underline{q}\right).*\underline{q}\\&+\underline{q}.^{2}.*\left(\partial^{2}\underline{f}./\partial\underline{q}^{2}\right)+2\underline{q}.*\left(\partial\underline{f}./\partial\underline{q}\right)+2\underline{f}\Big)+2\underline{DP_{f}}+\underline{DP_{p}}.\\&*\left(-2\underline{q}.^{-3}\right)\Big)\end{aligned} \tag{96}$$

iii. Energy balance in pseudoloops

The second derivative of the energy balance in pseudoloops, equation (44), with respect to the amount of leakage is

$$\left[Z.*\left(\partial\underline{w}./\partial\underline{q}\right)^{T}\right]\frac{\partial^{2}\underline{q}}{\partial\theta_{l}^{2}}=-\left[Z.*\left(\partial^{2}\underline{w}./\partial\underline{q}^{2}\right)^{T}\right]\left(\frac{\partial\underline{q}}{\partial\theta_{l}}\right).^{2} \tag{97}$$

The equations (94), (95) and (97) constitute the system of linear equations that result to the second derivatives of the flow rate with respect to the amount of leakage.

In order to calculate the second derivative of the pressure at junctions with respect to the amount of leakage, the equation (36) will be used to get the following

$$E\left(:,\underline{u}^{*}\right)\frac{\partial\underline{h}^{2}\left(\underline{u}^{*}\right)}{\partial\theta_{l}^{2}}=\frac{\partial^{2}\underline{w}}{\partial\theta_{l}^{2}} \tag{98}$$

using the chain rule

$$\frac{\partial^{2}\underline{w}}{\partial\theta_{l}^{2}}=\left(\partial^{2}\underline{w}./\partial\underline{q}^{2}\right).*\left(\frac{\partial\underline{q}}{\partial\theta_{l}}\right).^{2}+\left(\partial\underline{w}./\partial\underline{q}\right).*\frac{\partial^{2}\underline{q}}{\partial\theta_{l}^{2}} \tag{99}$$

All terms on the right-hand side of equation (99) are known.

2.4.3 Jacobian matrix for the non-linear system solver

The computational time needed to calculate the flow rates of the network through the non-linear system of equations (31), (40) and (44) can be reduced if the Jacobian matrix is provided to the solver. This matrix contains the derivatives of every equation in the system with every variable that is to be

determined. The three parts of the matrix, corresponding to the three sets of equations that the system is comprised of, are formulated.

$$J = \begin{bmatrix} J_1 \\ J_2 \\ J_3 \end{bmatrix} \tag{100}$$

where J_1, J_2, J_3 are partitions of the Jacobian matrix and represent the continuity, the loop and the pseudoloop equations respectively

i. Jacobian of the continuity equations

The Jacobian of the continuity equations (31) results to be the matrix C. The reason for this is that the demand θ_l is a constant, so it is canceled by the derivatives and the matrix C is independent of the variables, so the derivatives affect only the vector $\underline{q}$. When the Jacobian of the vector $\underline{q}$ is calculated, the result is the identity matrix I which when multiplied with any other matrix will result to the other matrix.

$$J_1 = \left[C \frac{\partial \underline{q}}{\partial q_1}, \dots, C \frac{\partial \underline{q}}{\partial q_P} \right] = \left[C \begin{Bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{Bmatrix}, \dots, C \begin{Bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{Bmatrix} \right] = CI = C \tag{101}$$

ii. Jacobian of the loop equations

The Jacobian of the loop equations (40) is

$$J_2 = \left[B \frac{\partial \underline{w}}{\partial q_1}, \dots, B \frac{\partial \underline{w}}{\partial q_P} \right] = B \left[\frac{\partial \underline{w}}{\partial q_1}, \dots, \frac{\partial \underline{w}}{\partial q_P} \right] \tag{102}$$

iii. Jacobian of the pseudoloop equations

The Jacobian of the pseudoloop equations (44) is similar to the loop equations since the head difference of the fixed grade nodes is independent of the flow rates and is thus canceled.

$$J_3 = \left[Z \frac{\partial \underline{w}}{\partial q_1}, \dots, Z \frac{\partial \underline{w}}{\partial q_P} \right] = Z \left[\frac{\partial \underline{w}}{\partial q_1}, \dots, \frac{\partial \underline{w}}{\partial q_P} \right] \tag{103}$$

3. Bayesian inference framework for leakage detection

In this chapter, the leakage detection methodology and the related formulation will be developed. The leakage detection will be accomplished using a Bayesian inference framework for parameter estimation and model selection. The objective is to identify the location and magnitude of leakage in a network using data acquired through hydraulic sensors. For this, a family of parameterized hydraulic models are introduced which share the same nominal hydraulic model and differ from each other from the set of parameters specified for each model. Each model is introduced to monitor leakage at a particular location of the network so that it corresponds to a leakage scenario. For example, a model can have a single parameter that accounts for the magnitude of leakage at a single pipe. The identification of the magnitude of leakage is accomplished by using Bayesian model selection methodology to estimate the most probable model given the data, as well as rank the rest of the models. The most probable model gives the leakage location(s) due to its association with a leakage scenario. Bayesian inference which is used to estimate the values of the parameters of each model, indicates the amount of leakage when applied to the most probable model. The posterior distribution of the model parameters developed for each model is used to formulate the optimal sensor placement methodology in Chapter 4. Since there are no measurements from an actual water distribution network (WDN) to demonstrate the effectiveness of the leakage detection technique, the measurements are created artificially from a perturbed model of the system. Results for leakage detection are presented in Chapter 7.

3.1 Bayesian parameter estimation framework

Let $\underline{\tilde{g}}\left(\underline{\theta},\underline{\varphi}\right) \in \mathbb{R}^{P+J}$ be the predictions of flow rates at all pipes and pressures at all junctions obtained from a model of steady-state flow in a WDN, with $\underline{\theta} \in \mathbb{R}^{N_\theta}$ being a vector of the model parameters that we are interested to update and $\underline{\varphi}$ being the rest of the model parameters that are not updated using the data. A set of measurements $D \equiv \underline{y}(\underline{\delta}) \in \mathbb{R}^{N_s}$ obtained from N_s number of sensors (flow rate and/or pressure sensors) placed at locations $\underline{\delta}$ is going to be used to update the parameters. To account for measurement and model error in the formulation, the following prediction error expression is used to relate the measurements to the model predictions.

$$\underline{y}(\underline{\delta}) = \underline{g}\left(\underline{\theta},\underline{\varphi}\right) + \underline{e} \tag{104}$$

where $\underline{e}$ is the prediction error term which can be assumed to follow a Gaussian distribution with zero mean and covariance matrix $\Sigma(\underline{\sigma}) \in \mathbb{R}^{N_s \times N_s}$ parameterized by the vector $\underline{\sigma}$

In order to correspond the model predictions with the degrees of freedom being measured the vector $\underline{g}\left(\underline{\theta}, \underline{\varphi}\right)$ is defined as follows

$$\underline{g}\left(\underline{\theta}, \underline{\varphi}\right) = M\underline{\tilde{g}}\left(\underline{\theta}, \underline{\varphi}\right) \tag{105}$$

where M is a selection matrix that keeps only the model predictions that correspond to the degrees of freedom that are being measured

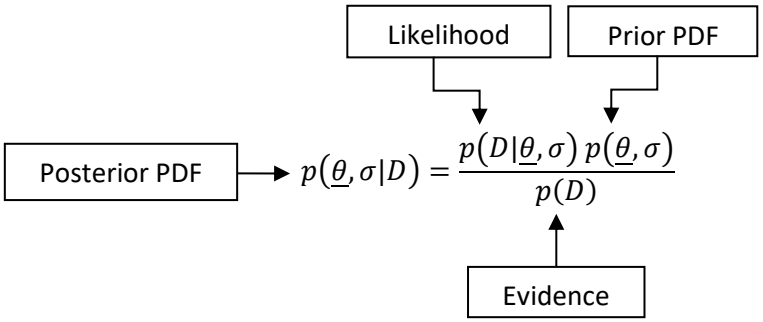
The covariance matrix that is going to be used for the prediction error is the following

$$\Sigma(\underline{\sigma}) = \sigma^2\, diag\{y_i^2\} \tag{106}$$

where y_i is the measurement of the i^{th} degree of freedom

This covariance matrix is diagonal and thus the error between different degrees of freedom is not correlated. For more information on correlated prediction error the following papers can be consulted [17], [18]. Moreover, the standard deviation of each degree of freedom is assumed to be proportional to its measurement. Since the parameter set $\underline{\sigma}$ is consisted of only one parameter, then $\Sigma(\underline{\sigma}) \equiv \Sigma(\sigma)$. The parameter σ is not known initially and is left to be inferred though the Bayesian methodology. If the value of σ is not of interest, then it can be marginalized out of the equations formulated below which can help in avoiding unidentifiability problems due to lack of sufficient information.

Introducing Bayes' rule



(107)

The objective is to calculate the posterior PDF of the parameters of interest $\underline{\theta}, \sigma$, which quantifies the uncertainty in their values. This is done using data, which are included in the likelihood term in Bayes' rule, and by using prior knowledge about the parameter values quantified by the prior PDF. The evidence term, in the context of parameter inference, is only a normalizing constant, which is used so the posterior can integrate to 1. The evidence, however, will be crucial in comparing the distinct hydraulic models to find the most probable one.

3.1.1 Formulating the posterior PDF

Based on equation (104) the measurements $\underline{y}(\underline{\delta})$ follow a Gaussian distribution $D \sim N\left(\underline{g}\left(\underline{\theta},\underline{\varphi}\right), \Sigma(\sigma)\right)$, since the model predictions are deterministic and the prediction error follows a Gaussian distribution. Thus, the analytical expression of the likelihood is

$$p(D|\underline{\theta},\sigma) = \frac{1}{(2\pi)^{N_s}\sqrt{\det(\Sigma(\sigma))}} \exp\left(-\frac{1}{2}J(\underline{\theta},\sigma)\right) \tag{108}$$

where $J(\underline{\theta},\sigma)$ is a measure of deviation between the model predictions and the measurements and is defined as

$$J(\underline{\theta},\sigma) = \left[\underline{y} - \underline{g}\left(\underline{\theta},\underline{\varphi}\right)\right]^T \Sigma^{-1}(\sigma)\left[\underline{y} - \underline{g}\left(\underline{\theta},\underline{\varphi}\right)\right] \tag{109}$$

Assuming the prior to be uniform, the posterior PDF can be written as

$$p(\underline{\theta},\sigma|D) \propto p(D|\underline{\theta},\sigma) \tag{110}$$

It can be shown that for large number of data the posterior asymptotically follows a Gaussian PDF. In this case, the posterior PDF is completely defined by the most probable values of the model parameter set and the covariance matrix.

The most probable values of $\underline{\theta},\sigma$ are the values that maximize the posterior PDF or equivalently minimize the function

$$L(\underline{\theta},\sigma) = -\ln\left(p(\underline{\theta},\sigma|D)\right) \tag{111}$$

This function is chosen because it varies much more slowly than the original posterior, making the quadratic approximation used more accurate. After some manipulation of equation (111) the result is

$$L(\underline{\theta},\sigma) = \frac{N_s}{2}\ln(2\pi) + N_s \ln\sigma + \frac{1}{2}\sum_{i=1}^{N_s}\ln y_i^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{N_s}\frac{\left(y_i - g_i\left(\underline{\theta},\underline{\varphi}\right)\right)^2}{y_i^2} + c_1 \tag{112}$$

where c_1 is a constant which depends on the prior, which is assumed constant, and evidence, which is constant, and doesn't affect the optimal values of the parameters

To determine the most probable values $\hat{\theta}_l, \hat{\sigma}$ a minimization scheme is applied on equation (112). The optimization algorithm that is used is the interior-point algorithm of the *fmincon* function in MATLAB. In order to speed up the process, the derivatives of the equation (112) with respect to the parameters of interest are provided analytically to the algorithm and are the following

$$\frac{\partial L(\underline{\theta},\sigma)}{\partial \theta_l} = -\frac{1}{\sigma^2}\sum_{i=1}^{N_s}\frac{y_i - g_i\left(\underline{\theta},\underline{\varphi}\right)}{y_i^2}\frac{\partial g_i\left(\underline{\theta},\underline{\varphi}\right)}{\partial \theta_l} \tag{113}$$

$$\frac{\partial L(\underline{\theta},\sigma)}{\partial \sigma} = \frac{N_s}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{N_s}\frac{\left(y_i - g_i\left(\underline{\theta},\underline{\varphi}\right)\right)^2}{y_i^2} \tag{114}$$

where the equations for computing the derivatives of the flow rates and pressures with respect to the model parameters have already been formulated analytically in Chapter 2.

If the assumption of large number of data cannot be used to approximate the posterior PDF by a Gaussian PDF, then the Transitional-Markov-Chain-Monte-Carlo (TMCMC) methodology [19] can be applied to sample from the posterior PDF as well as estimate the evidence of a model. Details of the use of the TMCMC are given in references [19], [20].

3.1.2 Selecting the most probable leakage model

Let M_i be a model that monitors leakage at pipe i and let $\underline{\theta}_i, \sigma_i$ be the corresponding model parameters. Bayes' rule is applied again in order to select the most probable leakage model within a family of leakage models, as follows

$$p(M_i|D) = \frac{p(D|M_i)p(M_i)}{p(D)} \tag{115}$$

Again, assuming constant prior the posterior PDF can be written as

$$p(M_i|D) \propto p(D|M_i) \tag{116}$$

The quantity on the right hand side of equation (116) is the evidence term of the Bayes' rule in equation (107). After some manipulation of (116) the result is

$$p(M_i|D) \propto \int p\left(\underline{\theta}_i,\sigma_i|M_i\right)\,\exp\left(-\tilde{L}\left(\underline{\theta}_i,\sigma_i\right)\right)d\theta_i d\sigma_i \tag{117}$$

where $\tilde{L}\left(\underline{\theta},\sigma\right) = L\left(\underline{\theta},\sigma\right) - c_1$ in order to ignore the constant term

The integral in (117) is a Laplace type integral and when the Laplace asymptotic approximation is applied the result is

$$p(M_i|D) \propto p\left(\hat{\underline{\theta}}_i,\hat{\sigma}_i|M_i\right)\exp\left(-\tilde{L}\left(\hat{\underline{\theta}}_i,\hat{\sigma}_i\right)\right)\frac{\left(\sqrt{2\pi}\right)^{N_\theta}}{\sqrt{\det\left(H\left(\hat{\underline{\theta}}_i,\hat{\sigma}_i\right)\right)}} \tag{118}$$

where $\hat{\theta},\hat{\sigma}$ are the most probable values found by minimizing equation (112) and $H\left(\underline{\theta},\sigma\right)$ is the Hessian matrix defined as follows

$$H(\underline{\theta},\sigma)=\begin{bmatrix}\frac{\partial^2 L(\underline{\theta},\sigma)}{\partial\theta_l^2} & \frac{\partial^2 L(\underline{\theta},\sigma)}{\partial\theta_l\partial\sigma}\\ \frac{\partial^2 L(\underline{\theta},\sigma)}{\partial\theta_l\partial\sigma} & \frac{\partial^2 L(\underline{\theta},\sigma)}{\partial\sigma^2}\end{bmatrix} \tag{119}$$

where

$$\frac{\partial^2 L(\underline{\theta},\sigma)}{\partial\theta_l^2}=-\frac{1}{\sigma^2}\sum_{i=1}^{N_s}\left[\frac{\hat{Y}_i-g_i\left(\underline{\theta},\underline{\varphi}\right)}{y_i^2}\frac{\partial^2 g_i\left(\underline{\theta},\underline{\varphi}\right)}{\partial\theta_l^2}-\frac{1}{y_i^2}\left(\frac{\partial g_i\left(\underline{\theta},\underline{\varphi}\right)}{\partial\theta_l}\right)^2\right] \tag{120}$$

$$\frac{\partial^2 L(\underline{\theta},\sigma)}{\partial\theta_l\partial\sigma}=\frac{2}{\sigma^3}\sum_{i=1}^{N_s}\frac{\hat{Y}_i-g_i\left(\underline{\theta},\underline{\varphi}\right)}{y_i^2}\frac{\partial g_i\left(\underline{\theta},\underline{\varphi}\right)}{\partial\theta_l} \tag{121}$$

$$\frac{\partial^2 L(\underline{\theta},\sigma)}{\partial\sigma^2}=-\frac{N_s}{\sigma^2}+\frac{3}{\sigma^4}\sum_{i=1}^{N_s}\frac{\left(\hat{Y}_i-g_i(\underline{\theta})\right)^2}{y_i^2} \tag{122}$$

It is noted that the covariance matrix of the posterior PDF is the inverse of the Hessian matrix evaluated at the most probable values $\Sigma(\sigma)=H^{-1}(\hat{\underline{\theta}},\hat{\sigma})$. The formulation for computing the first and second derivatives of the model prediction quantities with respect to the parameters set $\underline{\theta}$ has been presented in Chapter 2.

In order to simplify the equation (118) a logarithm is used which results in

$$\ln p(M_i|D)=-\tilde{L}(\hat{\underline{\theta}}_i,\hat{\sigma}_i)+\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln\det\left(H(\hat{\underline{\theta}}_i,\hat{\sigma}_i)\right)+\ln p(\hat{\underline{\theta}}_i,\hat{\sigma}_i|M_i)+c_2 \tag{123}$$

where c_2 is a constant resulting from the prior and evidence terms of equation (115) and can be ignored since it does not affect the optimization

Equation (123) is evaluated for each distinct model and the model which gives the highest value is the most probable model to represent the actual system. The parameterization of the most probable model M_i indicates the location of leakage while the value of the parameter θ_l indicates the amount of leakage.

4. Bayesian optimal sensor placement formulation

In this chapter, the formulation of the equations for the optimal sensor placement (OSP) are derived. The purpose of OSP is to find sensor configurations (type, number and location of sensors) that will provide measurements which are the most informative for estimating the parameters of interest presented in chapter 3. Since this methodology is implemented in the design phase, there are no available data to use yet and as a result, both the model parameters and measurements are uncertain. In order to overcome this problem, the prediction error equation (104) will be used again to simulate the measurements. The uncertainty will be dealt with by sampling the prior PDF of the parameters of interest, along with the prediction error term and calculate the expected value of the objective function with respect to these quantities. The objective function that will be used to quantify the amount of information a specific sensor configuration provides, is the information entropy (IE) [21]. IE provides a scalar measure of uncertainty and will measure the posterior uncertainty of the model parameters. The objective function is consisted of a multi-dimensional integral which can be calculated either with sampling or asymptotic approximations [22], both of which are formulated in this thesis. Since the objective is to maximize the amount of information a sensor configuration provides, the objective function will be conditioned on the sensor configurations $\underline{\delta}$ and optimized with respect to $\underline{\delta}$. Furthermore, a robust objective function will be derived by including in the formulation uncertainties in parameters that are not updated though the Bayesian methodology [18].

4.1 Objective function formulation

The objective in OSP is to find sensor configurations whose measurements provide maximum information about the model parameters to be inferred. This translates into minimizing the uncertainty of the parameters of interest after the measurements have been taken into consideration. One quantity that can be used to quantify the information provided is the information entropy. The information entropy provides a scalar measure of the uncertainty in a set of model parameters. Since the design variable is the sensor configuration $\underline{\delta}$, the IE is conditioned on $\underline{\delta}$ and then minimized with respect to $\underline{\delta}$. The IE applied to the posterior PDF of the model parameters is defined as follows and depends only on the provided data D, which implicitly depend on the sensors $\underline{\delta}$ (type, number, location) and the prediction error parameter σ.

$$h_{\underline{\theta}}\left(p\left(\underline{\theta}|\sigma,D,\underline{\delta}\right)\right)=\mathbb{E}_{\underline{\theta}}\left[-\ln p\left(\underline{\theta}|\sigma,D,\underline{\delta}\right)\right]=-\int_{\Theta} p\left(\underline{\theta}|\sigma,D,\underline{\delta}\right)\ln p\left(\underline{\theta}|\sigma,D,\underline{\delta}\right)d\underline{\theta} \quad (124)$$

The prediction error parameter σ in the OSP context is not among the parameters of interest and is not updated.

4.1.1 Asymptotic approximation of information entropy

The integral in equation (124) is a Laplace type integral over the model parameters which can be asymptotically approximated for large number of data and small prediction error by the expression [18], [22]

$$h_{\underline{\theta}}\left(p\left(\underline{\theta}|\sigma, D, \underline{\delta}\right)\right) \sim \frac{N_\theta}{2}(\ln(2\pi)+1) - \frac{1}{2}\ln\left(\det\left[Q\left(\underline{\delta};\underline{\hat{\theta}},\hat{\sigma}\right)+Q_\pi\left(\underline{\hat{\theta}}\right)\right]\right) \quad (125)$$

where $\underline{\hat{\theta}}, \underline{\hat{\sigma}}$ are the most probable values of the parameters to be updated, N_θ is the number of parameters to be updated, $Q\left(\underline{\delta};\underline{\theta},\sigma\right)$ is the Fisher information matrix (FIM) and $Q_\pi\left(\underline{\theta}\right)$ is matrix related to the prior PDF of the model parameter set $\underline{\theta}$ and is defined as

$$Q_\pi\left(\underline{\theta}\right) = -\nabla_{\underline{\theta}}^T \nabla_{\underline{\theta}} \ln p\left(\underline{\theta}\right) \quad (126)$$

The FIM can be asymptotically approximated by [11]

$$Q\left(\underline{\delta};\underline{\theta},\sigma\right) = \nabla_{\underline{\theta}} \underline{g}\left(\underline{\theta};\delta\right)^T \Sigma^{-1}(\sigma) \nabla_{\underline{\theta}} \underline{g}\left(\underline{\theta};\delta\right) \quad (127)$$

where $\nabla_{\underline{\theta}} = \left[\partial/\partial\theta_1, \ldots, \partial/\partial\theta_{N_\theta}\right]$

The covariance matrix $\Sigma(\sigma)$ which is used assumes uncorrelated prediction errors. The standard deviations of the prediction errors are assumed to be proportional to the model predictions and thus the covariance matrix is given by

$$\Sigma\left(\underline{\sigma}\right) = \sigma^2\, diag\left\{g_i^2\right\} \quad (128)$$

where g_i is the model prediction of the i^{th} measured degree of freedom

It can be seen from equation (125) that the value of IE depends on the FIM and the prior. If the prior is chosen to be uniform, which provides no information and for that case can be shown that $Q_\pi\left(\underline{\hat{\theta}}\right) = 0$, then IE depends only on the FIM. The FIM depends on the derivatives of the model predictions with respect to the parameters to be inferred. Higher derivative values indicate model predictions that are sensitive to changes to the parameters, resulting to higher values of the determinant of FIM. Since this is a minimization problem it translates to a maximization problem with respect to FIM, rewarding sensor locations measuring responses which are sensitive to parameter changes. A special case that

can arise is that for relatively few sensors, regardless of the locations placed, the information provided can be inadequate to identify the parameters. This presents itself as a close to zero determinant of the Fisher information matrix. A possible solution to this problem is to use an informative prior, such as a Gaussian prior or any other distribution that results to different than zero determinant due to the existence of the non-singular matrix $Q_\pi(\underline{\theta})$ in equation (125) [18].

The dependence on data on equation (125) comes through the optimal values $\underline{\hat{\theta}}, \underline{\hat{\sigma}}$. Determining the optimal values of the parameters requires information through data, which usually are not available, since the OSP methodology is applied in the design phase of an experiment in order to maximize the information gain through the experiment. A possible solution to this problem is to use nominal parameter values chosen by the designer. Alternatively, the robust IE formulation developed below can be used to account for the uncertainty in the model parameters.

4.1.1.1 Robust formulation

The IE expression formulated above relies on the designer choosing nominal values for the parameters $\underline{\hat{\theta}}, \underline{\hat{\sigma}}$. In order to avoid this, the Bayesian framework will be applied to formulate the robust IE, which relies on the uncertainty of the optimal parameters $\underline{\hat{\theta}}, \underline{\hat{\sigma}}$ quantified by their prior PDF. An equivalent form of expressing the amount of information a sensor configuration provides, is to calculate the difference between the IE of the prior and the IE of the posterior. The IE of the prior is independent of the measurements and, as a result, is a constant which doesn't affect the optimization solution. This measure is equivalent to the Mutual Information by information theory or the expected KL-Divergence. This reduction in the posterior IE will quantify the quality and the amount of information provided by the data. In order to keep it as a minimization problem the negative of the Mutual Information is used, as follows

$$\begin{aligned} U(\underline{\delta}) \equiv \Delta H(\underline{\delta}) &= E_{\underline{\hat{\theta}},\underline{\hat{\sigma}}}\left[E_{\underline{\theta}}\left[-\ln p(\underline{\theta},\underline{\hat{\theta}},\underline{\hat{\sigma}}|\underline{\delta})\right]\right] - E_{\underline{\hat{\theta}},\hat{\sigma}}\left[-\ln p(\underline{\hat{\theta}},\hat{\sigma})\right] \\ &\approx -E_{\underline{\hat{\theta}},\hat{\sigma}}\left[-\ln p(\underline{\hat{\theta}},\hat{\sigma})\right] + \frac{1}{2}N_\theta(\ln(2\pi)+1) \\ &\quad -\frac{1}{2}\int \ln\det\left[Q(\underline{\delta};\underline{\hat{\theta}},\underline{\hat{\sigma}}) + Q_\pi(\underline{\hat{\theta}},\underline{\hat{\sigma}})\right] p(\underline{\hat{\theta}},\underline{\hat{\sigma}}) d\underline{\hat{\theta}} d\underline{\hat{\sigma}} \end{aligned} \tag{129}$$

It is noted that the IE formulation in equation (129) enables the opportunity to also introduce uncertainty on model parameters that are not updated. This can be done by augmenting the set of parameters $\{\underline{\hat{\theta}}, \underline{\hat{\sigma}}\}$ to also include the model parameters $\underline{\varphi}$.

The multidimensional integral in (129) can be calculated numerically using Monte Carlo or sparse grid sampling techniques as follows

$$\int \ln\det\left[Q\left(\underline{\delta};\underline{\hat{\theta}},\underline{\hat{\sigma}}\right)+Q_{\pi}\left(\underline{\hat{\theta}},\underline{\hat{\sigma}}\right)\right]p\left(\underline{\hat{\theta}},\underline{\hat{\sigma}}\right)d\underline{\hat{\theta}}d\underline{\hat{\sigma}}$$
$$=\sum_{j=1}^{\widetilde{N}}w_j\ln\det\left[Q\left(\underline{\delta};\underline{\hat{\theta}}^{(j)},\underline{\hat{\sigma}}^{(j)}\right)+Q_{\pi}\left(\underline{\hat{\theta}}^{(j)},\underline{\hat{\sigma}}^{(j)}\right)\right] \quad (130)$$

where $\widetilde{N}$ is the number of samples or sparse grid points, $\underline{\hat{\theta}}^{(j)},\underline{\hat{\sigma}}^{(j)}$ is the j^{th} sample or sparse grid point in the parameter space and w_j is the weighting factor whose value depends on the numerical technique that is used.

For the robust optimization to calculate the covariance matrix in the FIM two approaches are used. The first one is to use a covariance matrix which depends on the model predictions of each sample, resulting in a different covariance matrix for each sample. This approach better simulates the modeling error which is proportional to the model predictions. The mathematical expression for this is

$$\Sigma_j\left(\underline{\sigma}\right)=\sigma^2\,diag\left\{g_i^2\left(\underline{\theta}_j\right)\right\} \quad (131)$$

The second approach is to use a constant covariance matrix. This approach better simulates the measurement error which depends on the accuracy of the measuring equipment. The mathematical expression that is used is the following

$$\Sigma\left(\underline{\sigma}\right)=\sigma^2\,diag\left\{\left(\frac{1}{\widetilde{N}}\sum_{j=1}^{\widetilde{N}}g_i\left(\underline{\theta}_j\right)\right)^2\right\} \quad (132)$$

For a more accurate representation of reality a combination of the two approaches for the covariance matrix would be preferable, since both kinds of error are present at the same time. This case, however, won't be explored here.

4.1.2 Sampling approximation of the mutual information

Another approach that can be used to formulate the objective function and avoid the asymptotic approximation, which assumes Gaussian posterior and small prediction error, is to use sampling methods straight away [22]. The objective function will be again the negative of the Mutual Information used in the robust formulation in the previous section. An alternative way to express the negative of the Mutual Information conditioned on $\underline{\delta}$ is

$$U\left(\underline{\delta}\right)=-\int_{\Theta}\int_{D}p\left(\underline{\theta},\underline{y}|\underline{\delta}\right)\ln\frac{p\left(\underline{\theta},\underline{y}|\underline{\delta}\right)}{p\left(\underline{\theta}|\underline{\delta}\right)p\left(\underline{y}|\underline{\delta}\right)}d\underline{\theta}d\underline{y} \quad (133)$$

In order to also include the uncertainty of the nuisance parameters $\underline{\sigma}, \underline{\varphi}$ in the formulation, parameters whose uncertainty is not updated, the expectation of (133) with respect to them can be used.

$$U(\underline{\delta}) = E_{\underline{\sigma},\underline{\varphi}}\left[U'\left(\underline{\delta}, \underline{\sigma}, \underline{\varphi}\right)\right] \tag{134}$$

where U' is U conditioned also on the nuisance parameters
In the following equations the nuisance parameters are omitted for simplicity.

By doing some manipulation on equation (133) the following form can be derived [22]

$$U(\underline{\delta}) = -\int_{D}\int_{\Theta} p\left(\underline{\theta}, \underline{y}|\underline{\delta}\right)\left[\ln p\left(\underline{y}|\underline{\theta}, \underline{\delta}\right) - \ln p\left(\underline{y}|\underline{\delta}\right)\right] d\underline{\theta}\, d\underline{y} \tag{135}$$

This form can readily be evaluated with the Monte Carlo sampling method to get the following estimator [23]

$$U(\underline{\delta}) = -\frac{1}{\widetilde{N}}\sum_{i=1}^{\widetilde{N}}\left\{\ln p\left(\underline{y}_i|\underline{\theta}_i, \underline{\delta}\right) - \ln\left\{\frac{1}{\widetilde{N}}\sum_{j=1}^{\widetilde{N}} p\left(\underline{y}_i|\underline{\theta}_j, \underline{\delta}\right)\right\}\right\} \tag{136}$$

where $\widetilde{N}$ is the number of samples, $\underline{\theta}_i$ is the i^{th} sample from the prior and $\underline{y}_i$ is the measurement created from the prediction error equation (104) using $\underline{\theta}_i$ model parameters and a random sample from the prediction error term.

In order to reduce the noise of the sampling estimate it is possible to evaluate the objective function multiple times with same model samples but different error samples and take the mean of the results. This is not computationally expensive since the model does not have to be evaluated again, only the likelihood evaluations are increased to account for the new error samples.

4.1.3 Monitoring an area of the network

In order to be able to monitor an area of the network for leakage or in other words to find sensor configurations which are optimal for locating leakage on a set of pipes, the following methodology is used. The objective function is calculated for every pipe in the set of pipes $S_P \in \mathbb{R}^{N_P}$ conditioned on the sensor configurations $\underline{\delta}$. Then, the average of the results is calculated and it is used as the new objective function to find the optimal sensor configurations to monitor an area of the network, which can be written as

$$U(\underline{\delta}; N_P) = \frac{1}{N_P}\sum_{i \in N_P} U(\underline{\delta}, M_i) \tag{137}$$

where M_i is the model of the system which has a leakage on pipe i and the objective function is evaluated for that model

4.2 Optimal sensor placement methodology

Now that the objective functions that can be used are formulated the methodology to find the optimal sensor configurations will be developed. The optimal sensor configuration involves selecting the type of sensors, the number of sensors and the locations to be placed. In the context of WDN presented here the sensors that will be used are flow rate sensors in pipes and pressure sensors in nodes. This OSP problem is an optimization problem with design variables being the sensor configurations $\underline{\delta}$, which take discrete values. The objective is to find configurations that result to the most informative measurements to identify the parameters of interest or equivalently to minimize the posterior uncertainty of the parameters. The formulated objective functions quantify the posterior uncertainty and the optimization problem is formulated as the following minimization problem.

$$\underline{\delta}_{opt} = argmin_{\underline{\delta}} U(\underline{\delta}) \quad (138)$$

It is noted that the worst sensor locations can also be found by solving for sensor configurations which maximize the objective function instead. This will be done in the next chapters in order to show the effectiveness of the resulting optimal sensor configurations compared to the worst sensor configurations. Since this is a discreate optimization problem, for relatively small and large number of sensors it is possible to solve exhaustively by evaluating the objective function for all possible combinations of sensors. However, for big networks this is computationally prohibitive due to the sheer number of combinations that must be evaluated. To account for this problem two heuristic sequential placing algorithms are used.

4.2.1 FSSP-BSSP Heuristic algorithms

The heuristic algorithms employed to solve the OSP problem are the FSSP (Forward Sequential Sensor Placement) and BSSP (Backward Sequential Sensor Placement) [11]. The way the FSSP algorithm works, is to find exhaustively the location of the first sensor and then given the location of the first sensor an exhaustive search is employed to find the location of the second sensor, then given the location of the first two sensors an exhaustive search is used to find the location of the third sensor, and the methodology is repeated until all sensor numbers are covered. The resulting sensor configurations are approximate but can be achieved in a very small fraction of the time compared to when an exhaustive search approach is used. It has been demonstrated in a number of applications that the approximate solution provided by the FSSP is very close to the exact one [11].

The BSSP is a similar algorithm but instead of starting with the first sensor, one starts with the last sensor. Given that sensors are placed everywhere, an exhaustive search is conducted to find which sensor, if removed, results to the smallest increase in the posterior uncertainty of the parameters. Given that one sensor is removed an exhaustive search is conducted to find the next sensor that can be removed with the smallest increase in the uncertainty. This methodology is repeated until the desired number of sensors is left.

After these algorithms are applied, the optimal sensor configuration for a given number of sensors is known. In order to choose what is the optimal number of sensors, the value of the objective function is going to be compared for each number of sensors. If a significant reduction in the value of the objective function can be accomplished with the addition of one more sensor, then the additional sensor should be considered. If the addition of a sensor does not reduce significantly the value of the objective function, then the current number of sensors is the optimal number to be used.

5. Presentation of network

The WDN that is used to evaluate the aforementioned methodologies is shown in Figure 5. It is an artificial network consisted of 2 reservoirs, 52 pipes and 31 junctions, all of which are active (have positive demand).

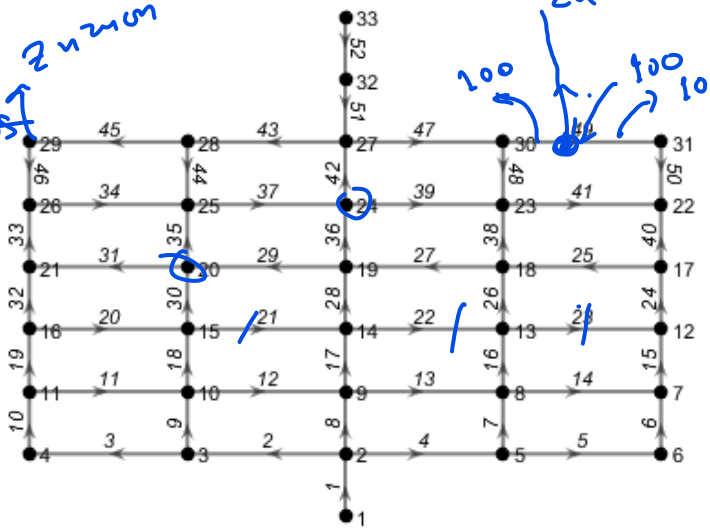


Figure 5 Water Distribution Network

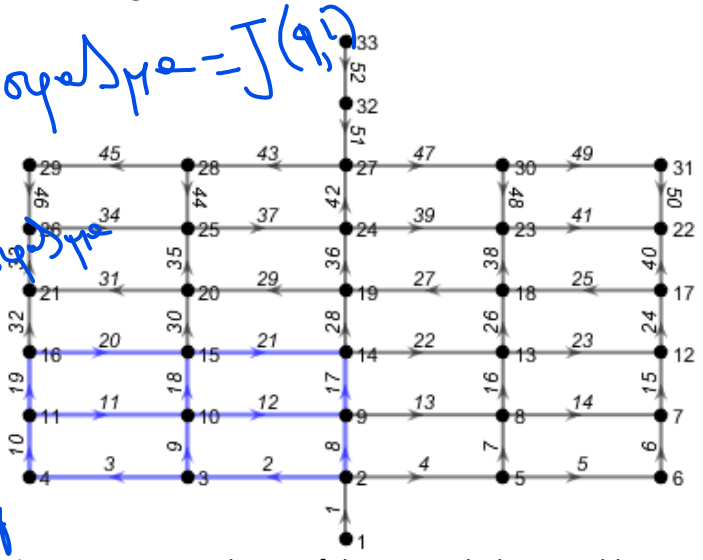


Figure 6 Monitored area of the network shown in blue

The reservoirs are in nodes 1 and 33 with the water level being in an altitude of 50m. All the pipes have a length (l) of 200m, diameter (d) of 250mm and roughness (ε) of 0.26mm with the addition of 100kW constant power (p) pumps with 0.75 efficiency (η) in pipes 1 and 52. All the junctions have an altitude of 0m and nominal demand of 10L/s. No minor losses (K) were introduced in the pipes. The properties of water used are density $\rho = 998.19\ kg/m^3$ and viscosity $\mu = 10^{-3}\ Pa \cdot s$

The area of the network that is monitored for leakage is the area on the bottom left of the network consisted of pipes 2, 3, 8, 9, 10, 11, 12, 17, 18, 19, 20 and 21 as shown in Figure 6 in blue.

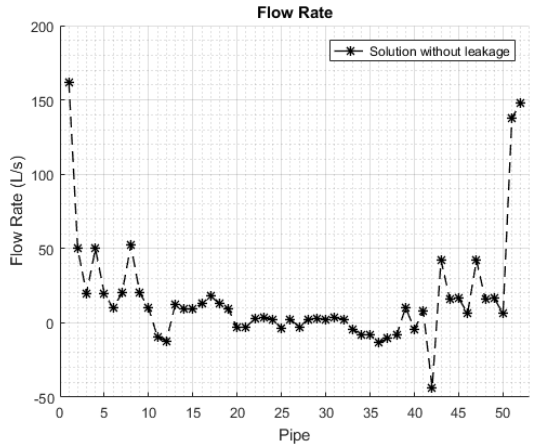


Figure 7 Flow rate

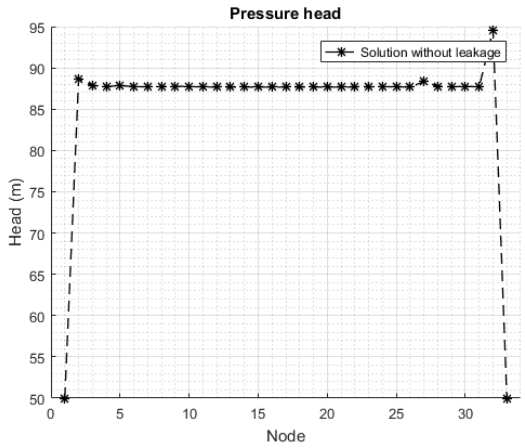


Figure 8 Pressure head

The solution of the model (flow-rate at pipes and pressure at nodes) for the nominal parameter values and without the presence of a leakage can be seen on Figure 7 and Figure 8.

6. Optimal sensor placement results

The methodology for OSP developed in chapter 4 will now be applied to the presented WDN. In total 10 cases are explored and each one of the cases is tested with pressure and flow rate sensors individually. These cases will explore the results of the two approximations of the objective function, the two different covariance matrices with two values for the prediction error parameter and the effect on the objective function of introducing uncertainty on the amount of leakage and on the nodal demands. For each case three diagrams are presented. The first one shows the value of the objective function as a function of the number of sensors placed in the best and in the worst locations. The second shows the best locations to place the sensors, for example if 5 sensors are to be placed in the optimal locations, they will be placed on locations 1 to 5. The third diagram is similar to the second but instead shows the worst sensor locations. All the following cases will use a Gaussian prior for the amount of leakage with mean 10L/s and variance 4, i.e. the prior is a normal distribution $N(10,4)$. The cases are the following.

Case 1

Objective function: Asymptotic approximation (129) with the amount of leakage set to the nominal value of 10L/s. This is achieved by using a Dirac delta function for the prior distribution of the model parameters in equation (129) in order to correspond to equation (125).

Nodal demands: Deterministic with values set to their nominal values.

Covariance matrix: Since it is calculated only once for the nominal parameter values both matrix formulations yield the same result (131), (132). The prediction error parameter σ is set to value 0.01 as nominal value.

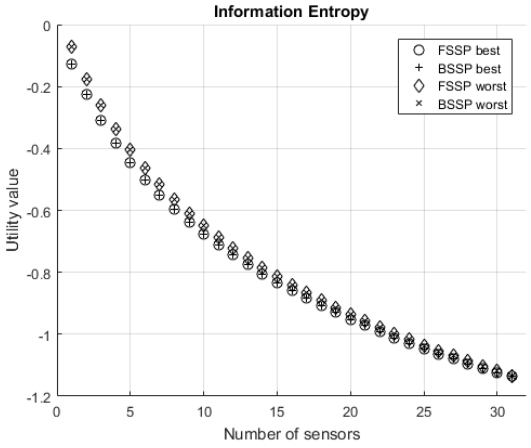


Figure 9 Objective function for pressure sensors

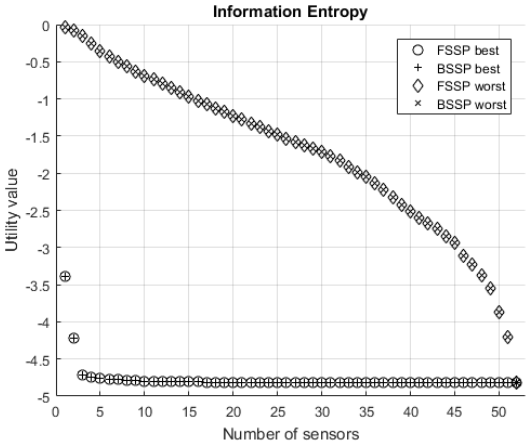


Figure 10 Objective function for flow-rate sensors

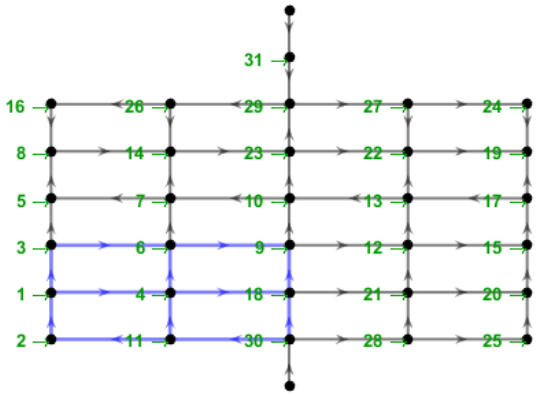


Figure 11 Best pressure sensor locations

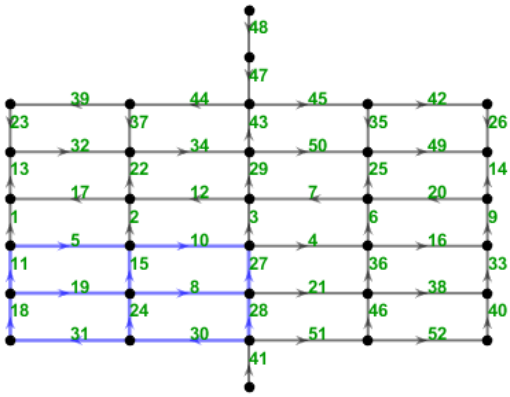


Figure 12 Best flow-rate sensor locations

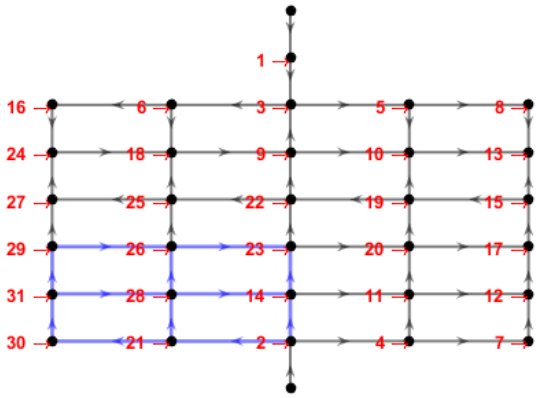


Figure 13 Worst pressure sensor locations

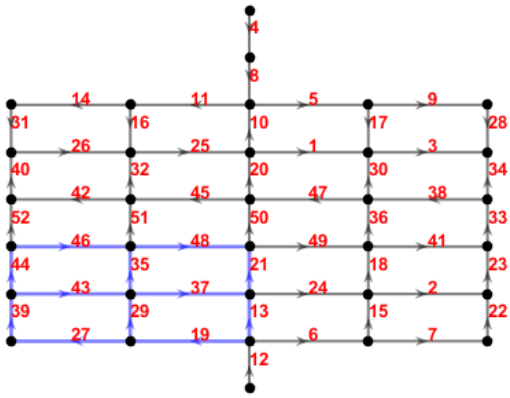


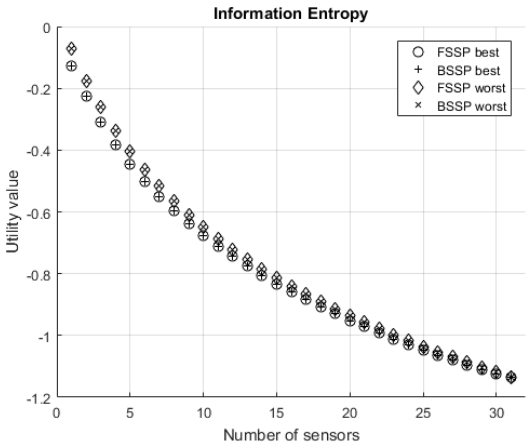
Figure 14 Worst flow-rate sensor locations

Case 2

Objective function: Asymptotic approximation (129) with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done using the Monte Carlo method with 1000 samples drawn from the Gaussian prior for the amount of leakage.

Nodal demands: Deterministic with values set to their nominal values.

Covariance matrix: Independent model predictions errors (Equation (132)) with prediction error parameter σ set to 0.01 as nominal value.



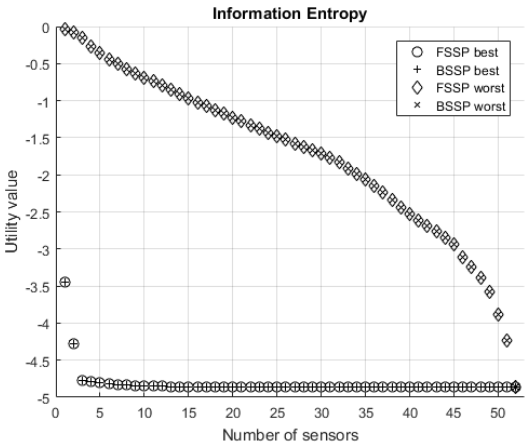


Figure 15 Objective function for pressure sensors **Figure 16** Objective function for flow-rate sensors

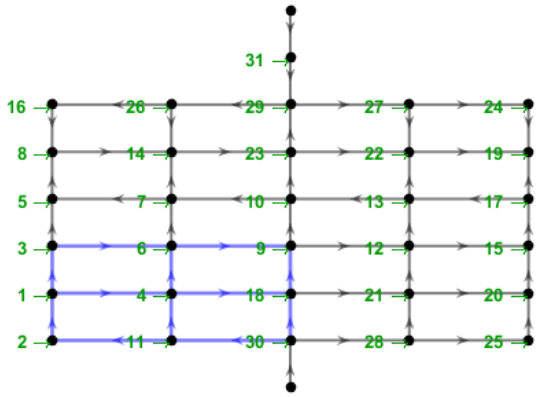


Figure 17 Best pressure sensor locations

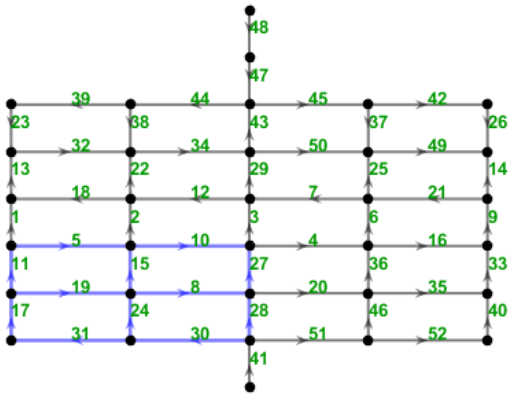


Figure 18 Best flow-rate sensor locations

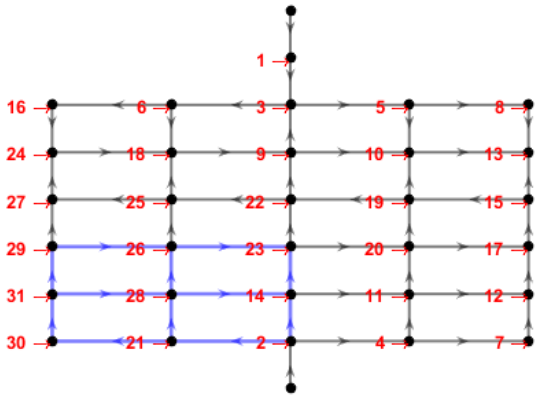


Figure 19 Worst pressure sensor locations

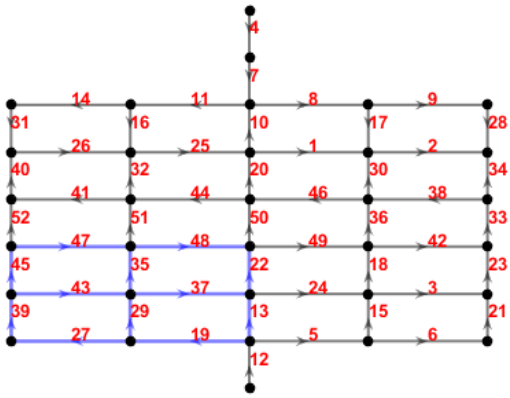


Figure 20 Worst flow-rate sensor locations

Case 3

Objective function: Sampling approximation (136), with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done using the Monte Carlo method with 1000

samples drawn from the Gaussian prior for the amount of leakage for flow-rate sensors and 5000 samples for pressure sensors.

Nodal demands: Deterministic with values set to their nominal values.

Covariance matrix: Independent of the model predictions errors (Equation (132)). The prediction error parameter σ is set to 0.01 as nominal value.

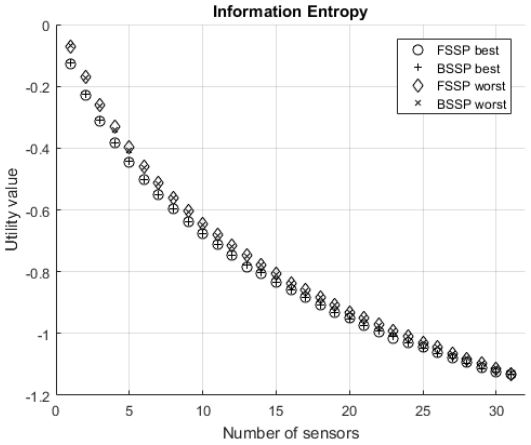


Figure 21 Objective function for pressure sensors

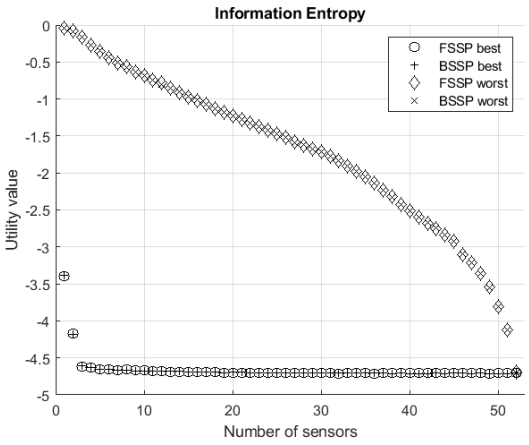


Figure 22 Objective function for flow-rate sensors

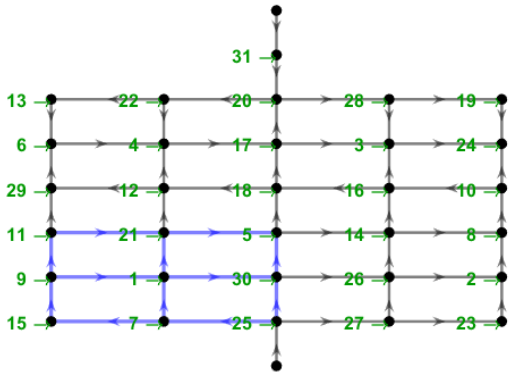


Figure 23 Best pressure sensor locations

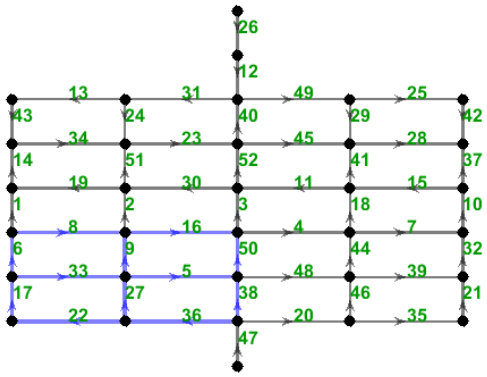


Figure 24 Best flow-rate sensor locations

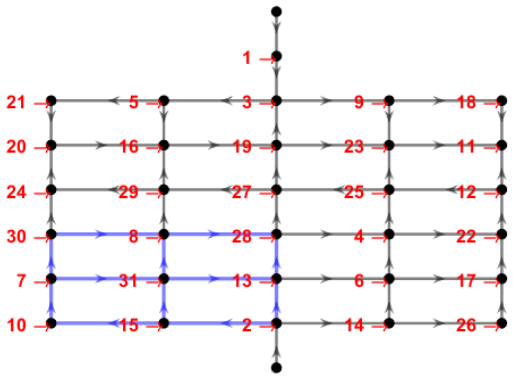


Figure 25 Worst pressure sensor locations

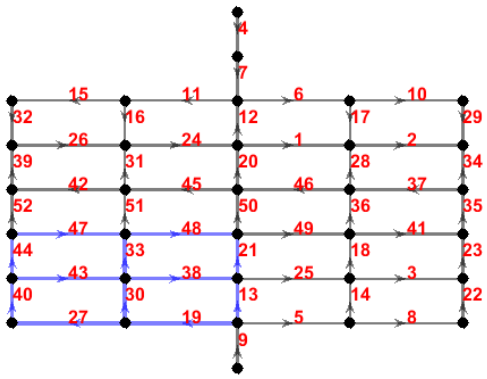


Figure 26 Worst flow-rate sensor locations

Case 4

Objective function: Asymptotic approximation (129), with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done using the Monte Carlo method with 1000 samples drawn from the Gaussian prior for the amount of leakage.

Nodal demands: Deterministic with values set to their nominal values.

Covariance matrix: Proportional model predictions errors (Equation (131)). The prediction error parameter σ is set to 0.01 as nominal value.

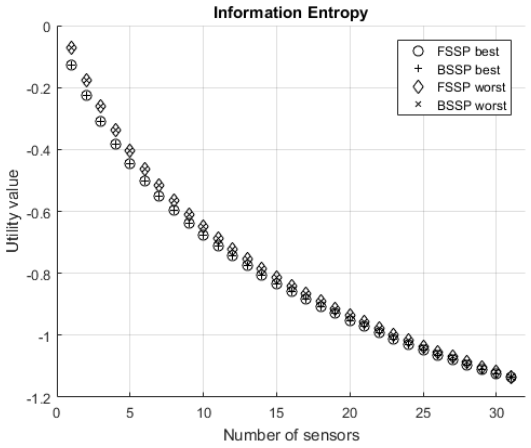


Figure 27 Objective function for pressure sensors

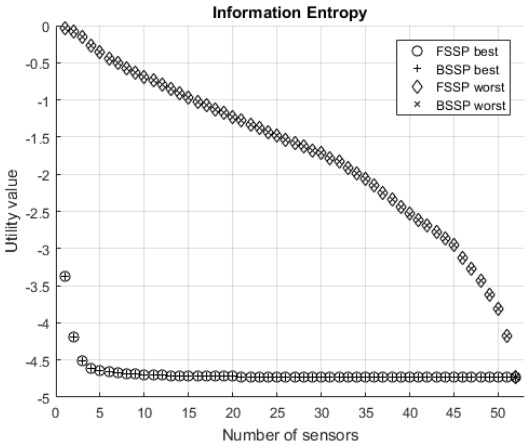


Figure 28 Objective function for flow-rate sensors

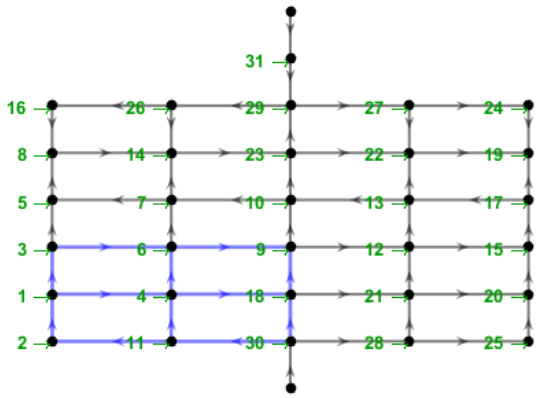


Figure 29 Best pressure sensor locations

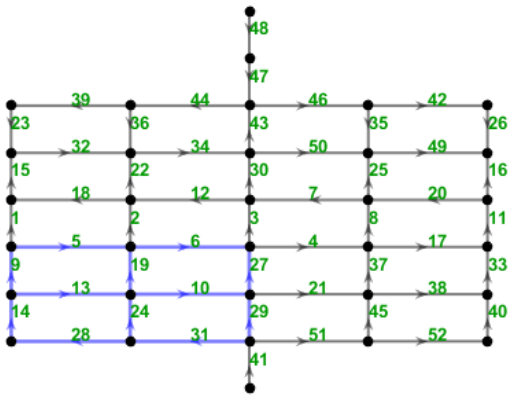


Figure 30 Best flow-rate sensor locations

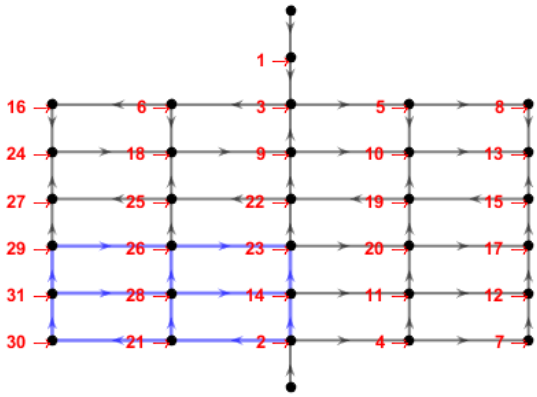


Figure 31 Worst pressure sensor locations

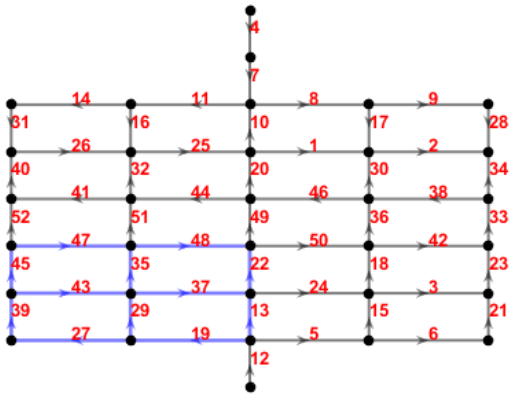


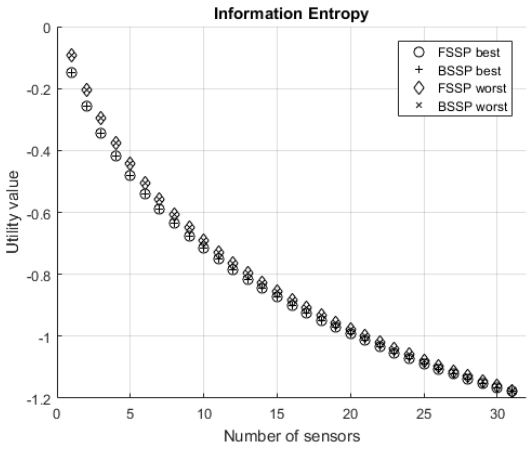
Figure 32 Worst flow-rate sensor locations

Case 5a

Objective function: Asymptotic approximation (129) with the amount of leakage set to the nominal value of 10L/s. This is achieved by using a Dirac delta function for the prior distribution of the model parameters in equation (129) in order to correspond to equation (125).

Nodal demands: Uncertain with all junctions constituting one group with the demand of the junctions in the group following a Gaussian distribution with mean equal to the nominal values of the nodal demands (equal to 10L/s) and standard deviation equal to 20% of the nominal value of the nodal demands, i.e. a Gaussian PDF $\sim N(10,4)$. The numerical integration is performed using the Monte Carlo method with 1000 samples drawn from the Gaussian prior.

Covariance matrix: Independent of the model predictions errors (Equation (132)). The prediction error parameter σ is set to 0.01 as nominal value.



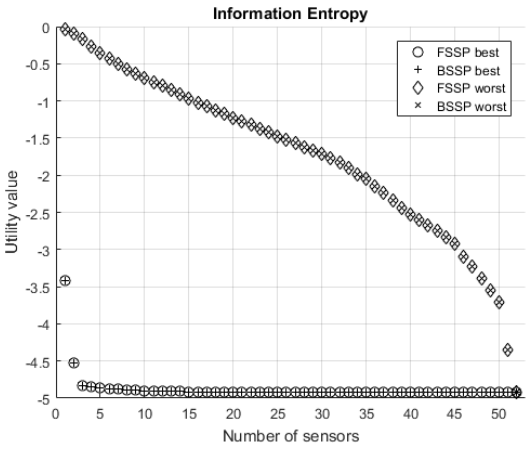


Figure 33 Objective function for pressure sensors **Figure 34** Objective function for flow-rate sensors

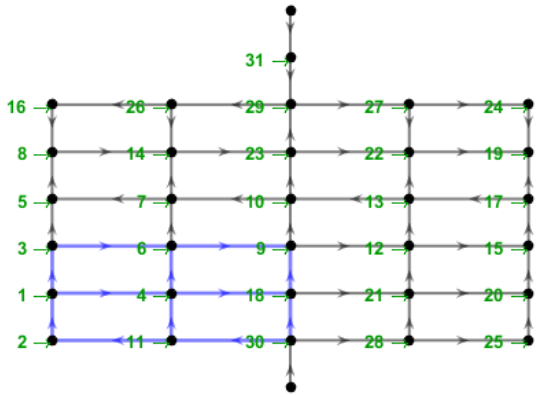


Figure 37 Best pressure sensor locations

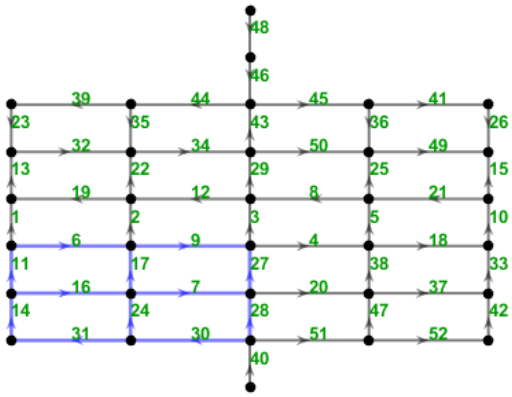


Figure 38 Best flow-rate sensor locations

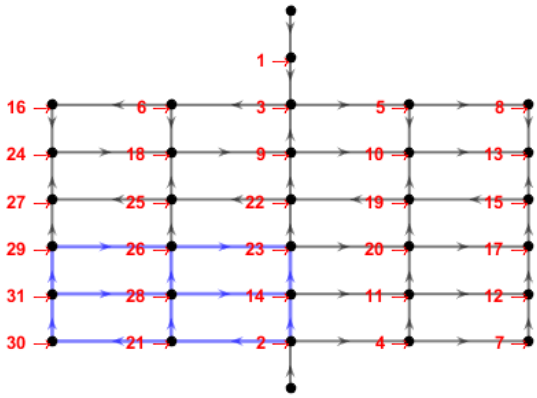


Figure 39 Worst pressure sensor locations

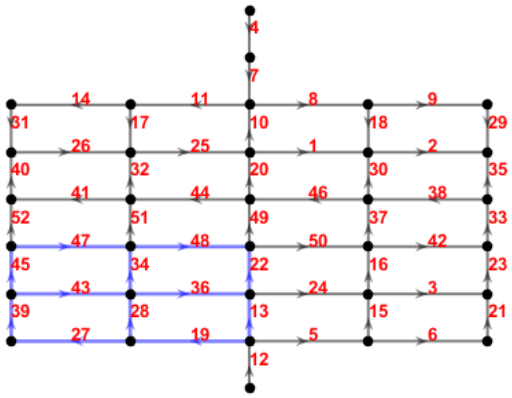


Figure 40 Worst flow-rate sensor locations

Case 5b

Similar to Case 5a apart from the objective function being an asymptotic approximation (129), with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done using the Monte Carlo method with 1000 samples drawn from the Gaussian prior for the amount of leakage.

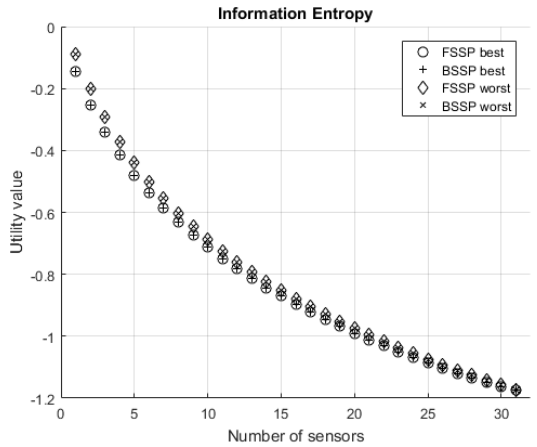


Figure 35 Objective function for pressure sensors

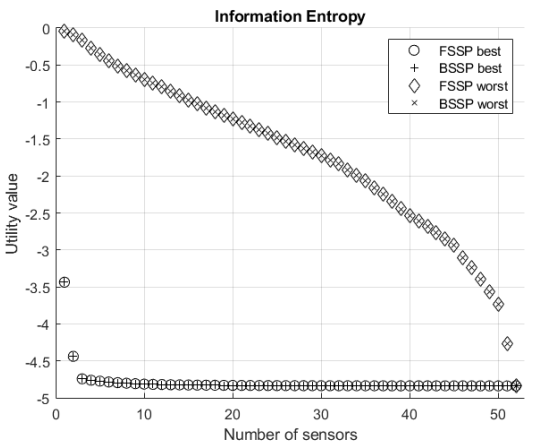


Figure 36 Objective function for flow-rate sensors

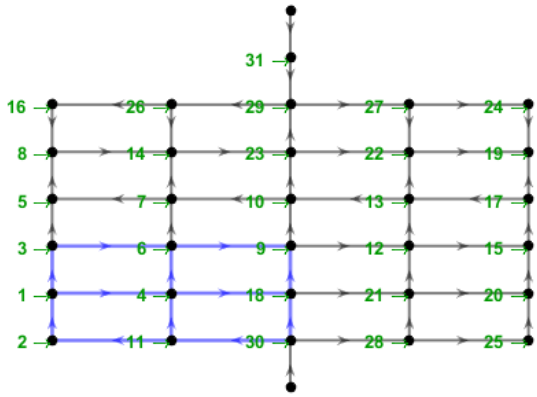


Figure 44 Best pressure sensor locations

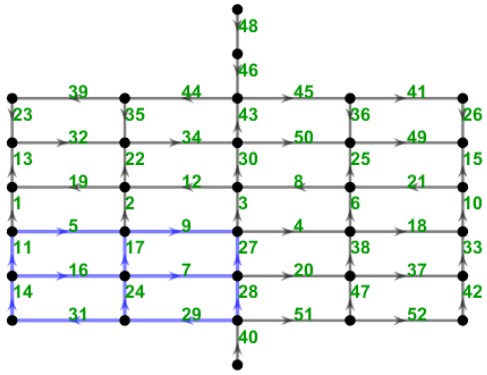


Figure 43 Best flow-rate sensor locations

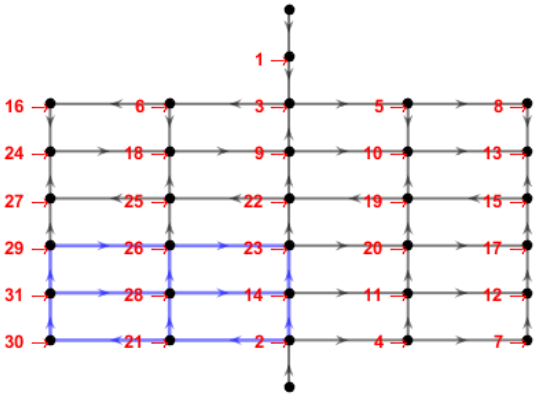


Figure 41 Worst pressure sensor locations

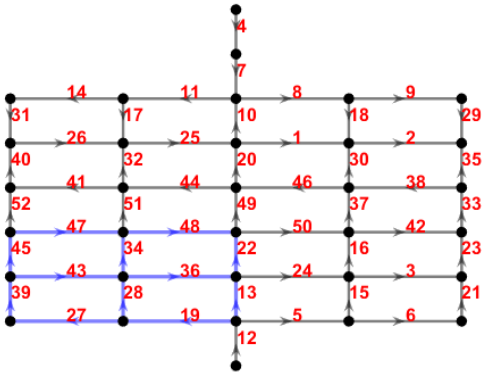


Figure 42 Worst flow-rate sensor locations

Case 6a

Objective function: Asymptotic approximation (129) with the amount of leakage set to the nominal value of 10L/s. This is achieved by using a Dirac delta function for the prior distribution of the model parameters in equation (129) in order to correspond to equation (125).

Nodal demands: Uncertain with all junctions constituting one group with the demand of the junctions in the group following a Gaussian distribution with mean equal to the nominal values of the nodal demands (equal to 10L/s) and standard deviation equal to 20% of the nominal value of the nodal demands, i.e. a Gaussian PDF $\sim N(10,4)$. The numerical integration is performed using the Monte Carlo method with 1000 samples drawn from the Gaussian prior.

Covariance matrix: Proportional model predictions errors (Equation (131)). The prediction error parameter σ is set to 0.01 as nominal value.

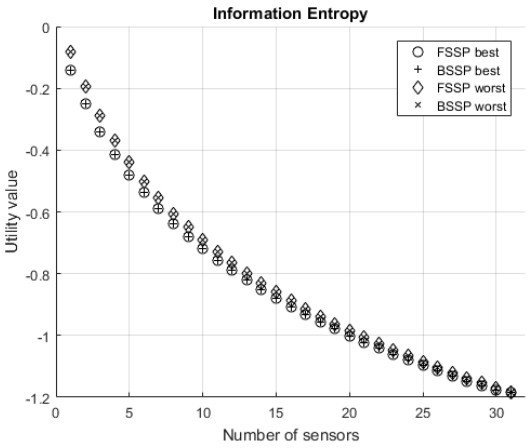


Figure 45 Objective function for pressure sensors

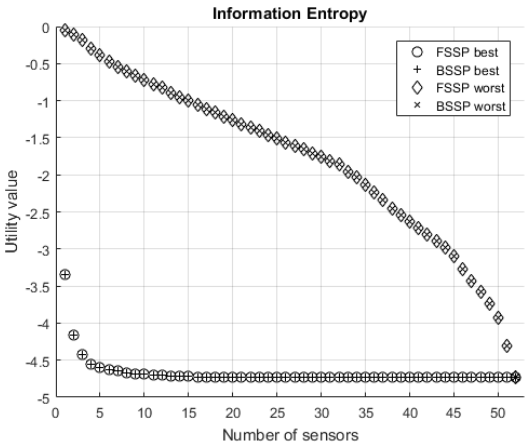


Figure 46 Objective function for flow-rate sensors

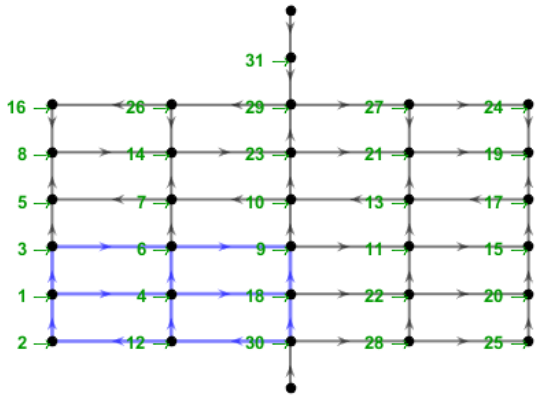


Figure 47 Best pressure sensor locations

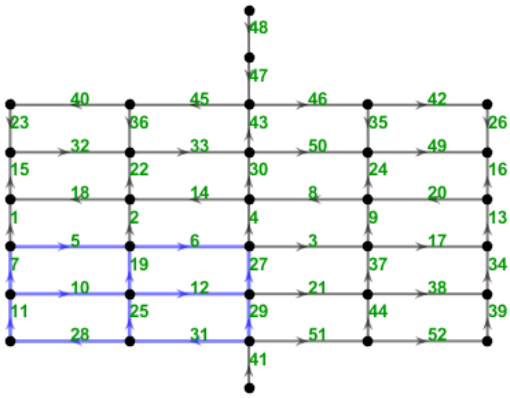


Figure 48 Best flow-rate sensor locations

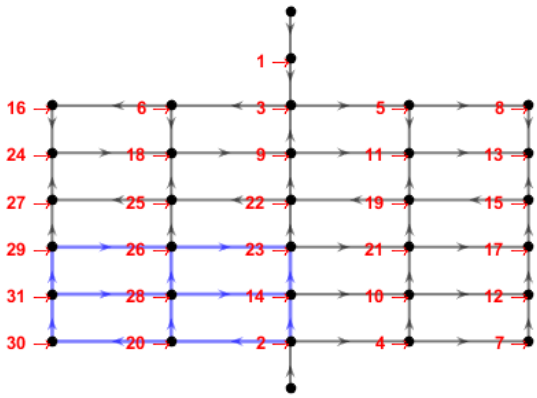


Figure 49 Worst pressure sensor locations

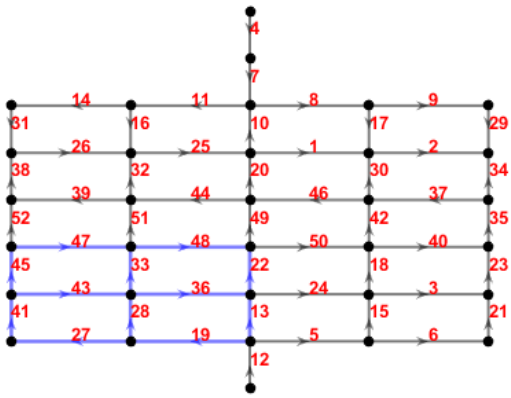


Figure 50 Worst flow-rate sensor locations

Case 6b

Similar to Case 6a apart from the objective function being an asymptotic approximation (129), with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done

using the Monte Carlo method with 1000 samples drawn from the Gaussian prior for the amount of leakage.

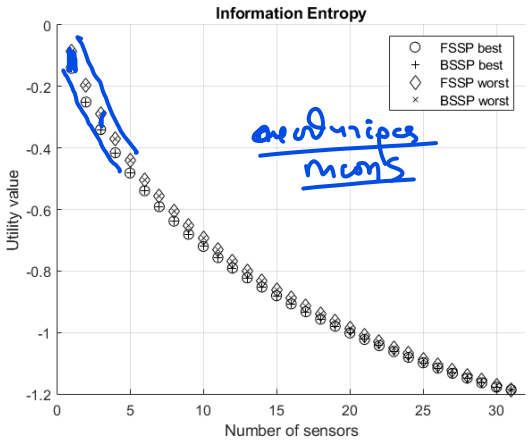


Figure 51 Objective function for pressure sensors

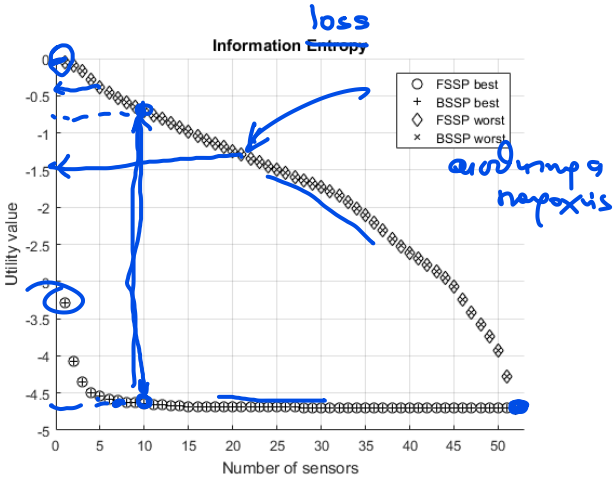


Figure 52 Objective function for flow-rate sensors

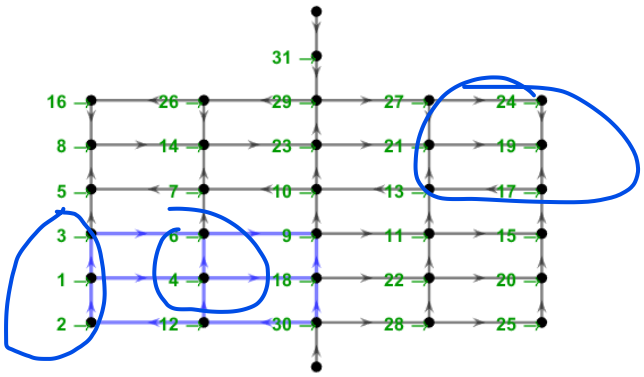


Figure 53 Best pressure sensor locations

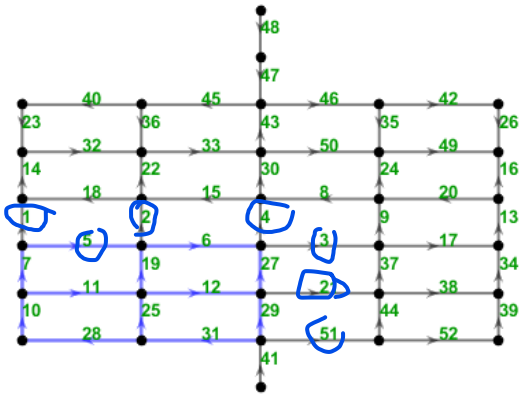


Figure 54 Best flow-rate sensor locations

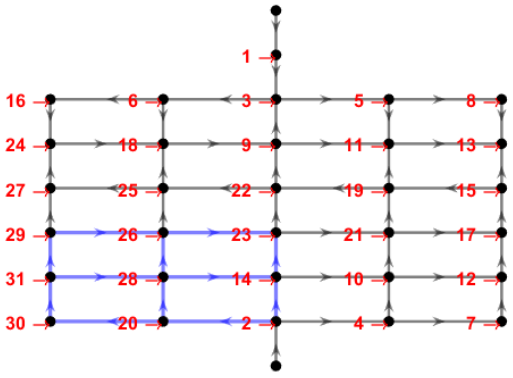


Figure 55 Worst pressure sensor locations

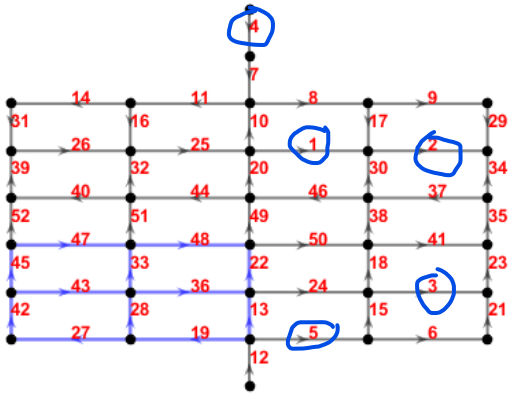


Figure 56 Worst flow-rate sensor locations

Case 7a

Objective function: Asymptotic approximation (129) with the amount of leakage set to the nominal value of 10L/s. This is achieved by using a Dirac delta function for the prior distribution of the model parameters in equation (129) in order to correspond to equation (125).

Nodal demands: Uncertain with all junctions divided into 4 groups with the demand of the junctions in each group following a Gaussian distribution with mean equal to the nominal values of the nodal demands (equal to 10L/s) and standard deviation equal to 20% of the nominal value of the nodal demands, i.e. a Gaussian PDF ~N(10,4). The numerical integration is performed using the Monte Carlo method with 1000 samples drawn from the Gaussian prior. The groups are the following and are also illustrated in the following figures with different colors.

Group 1 comprised of nodes 2, 3, 4, 10, 11, 14, 15 and 16.

Group 2 comprised of nodes 5, 6, 7, 8, 9, 12 and 13.

Group 3 comprised of nodes 20, 21, 24, 25, 26, 28, 29 and 32.

Group 4 comprised of nodes 17, 18, 19, 22, 23, 27, 30 and 31.

Covariance matrix: Independent of the model predictions errors (Equation (132)). The prediction error parameter σ is set to 0.01 as nominal value.

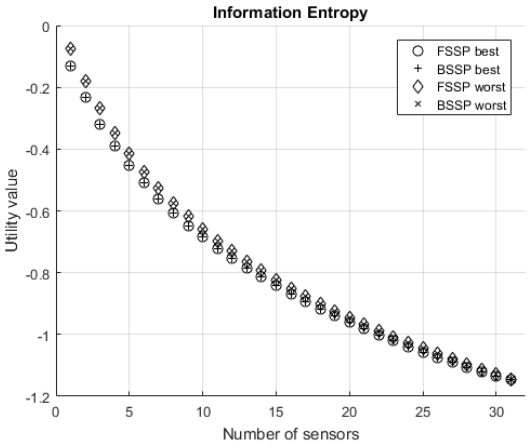


Figure 57 Objective function for pressure sensors

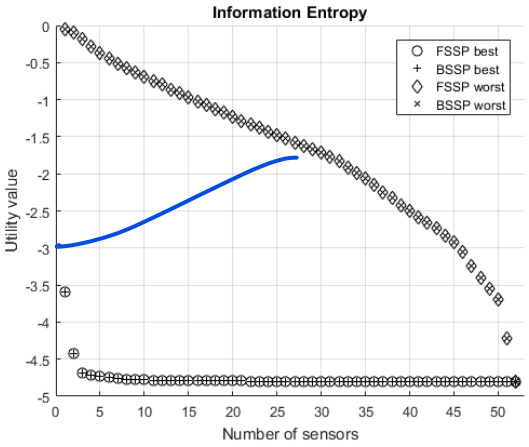


Figure 58 Objective function for flow-rate sensors

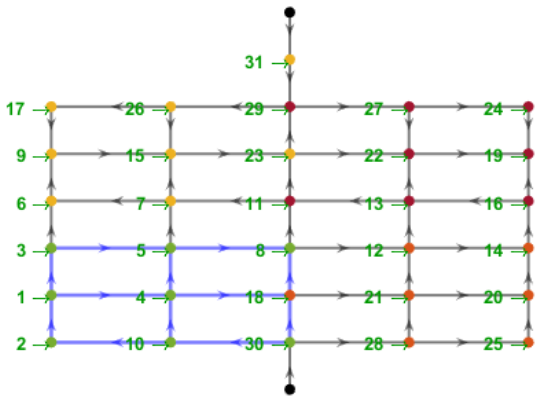


Figure 59 Best pressure sensor locations

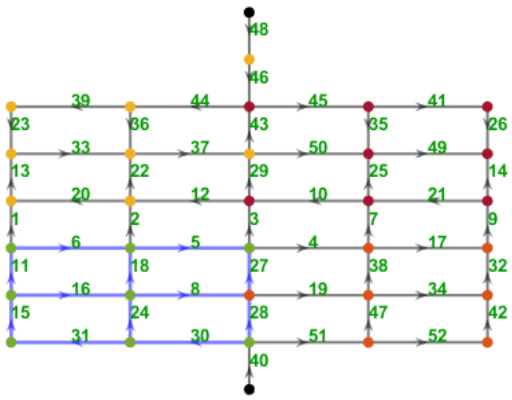


Figure 60 Best flow-rate sensor locations

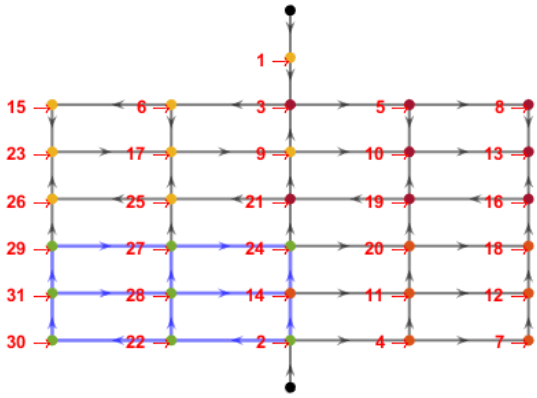


Figure 62 Worst pressure sensor locations

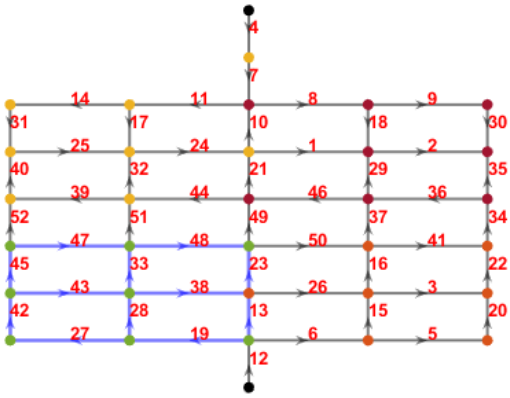


Figure 61 Worst flow-rate sensor locations

Case 7b

Similar to Case 7a apart from the objective function being an asymptotic approximation (129), with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done using the Monte Carlo method with 1000 samples drawn from the Gaussian prior for the amount of leakage.

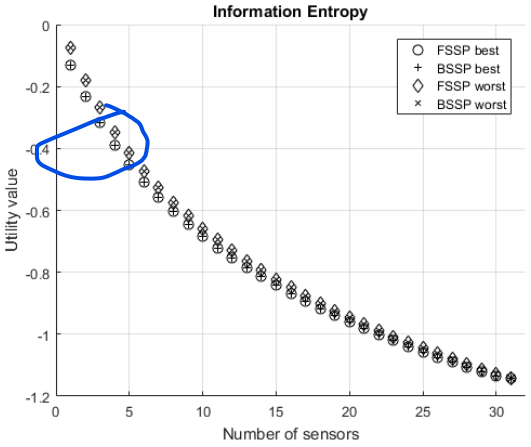


Figure 63 Objective function for pressure sensors

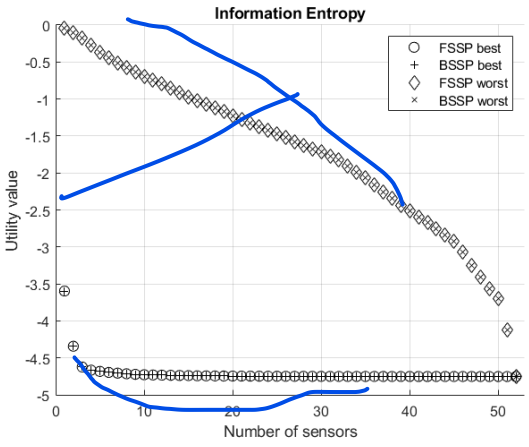


Figure 64 Objective function for flow-rate sensors

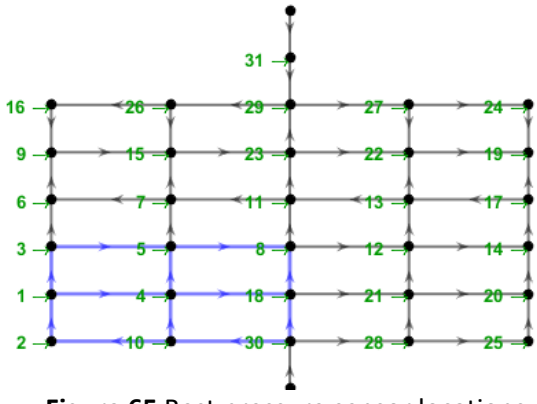


Figure 65 Best pressure sensor locations

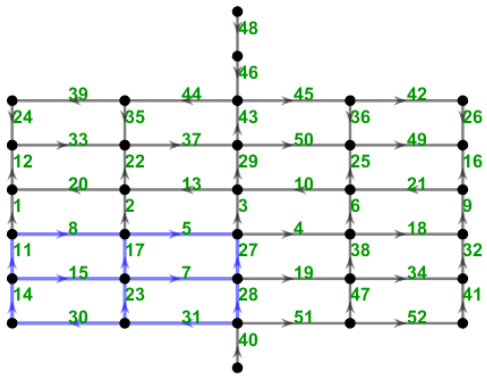


Figure 66 Best flow-rate sensor locations

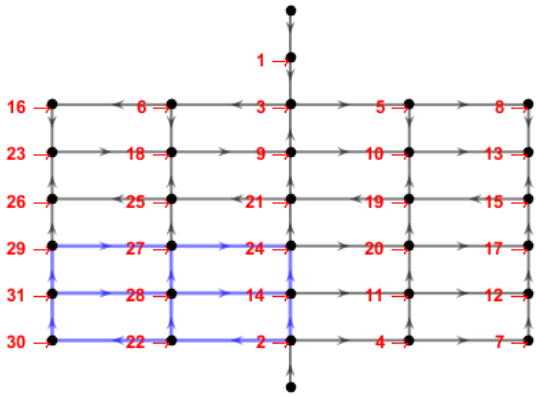


Figure 67 Worst pressure sensor locations

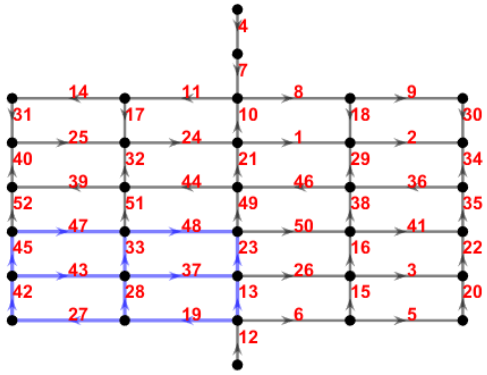


Figure 68 Worst flow-rate sensor locations

Case 7c

Similar to Case 7b apart from the covariance matrix prediction error parameter σ which is set to 0.05 as nominal value.

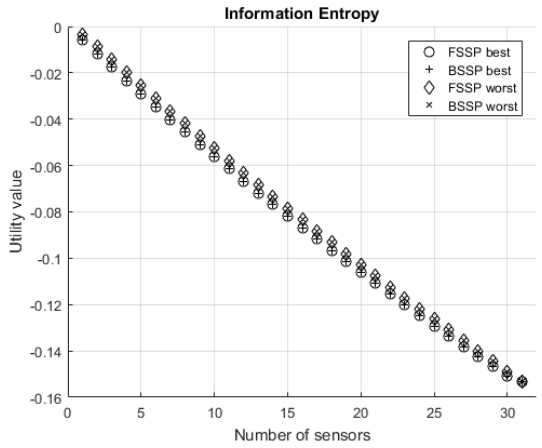


Figure 69 Objective function for pressure sensors

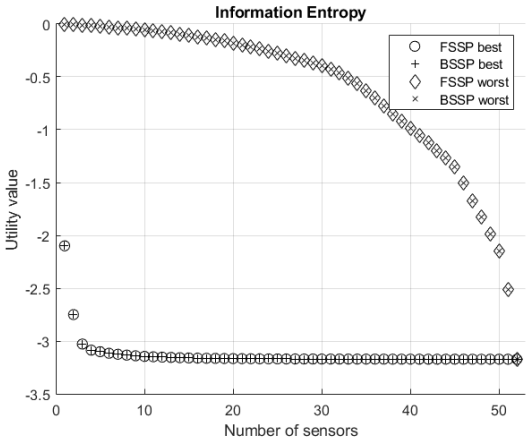


Figure 70 Objective function for flow-rate sensors

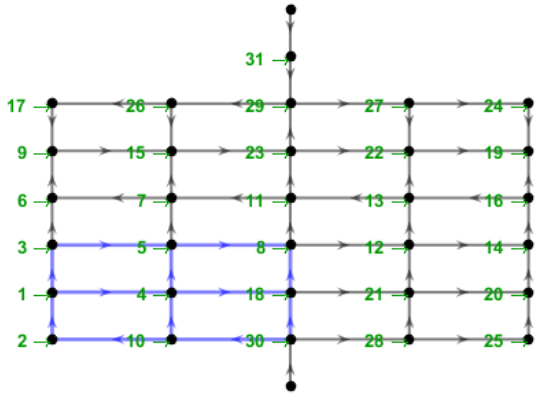


Figure 71 Best pressure sensor locations

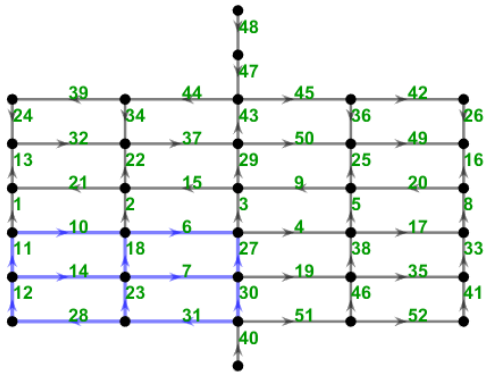


Figure 72 Best flow-rate sensor locations

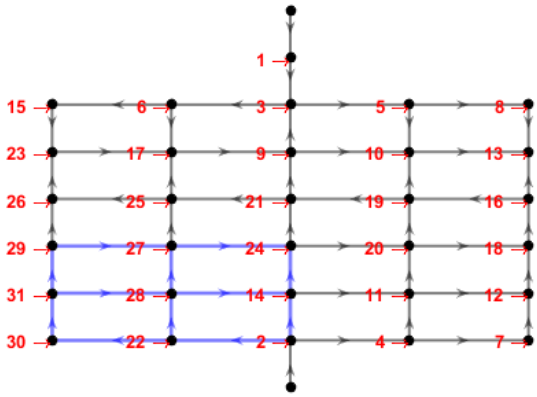


Figure 73 Worst pressure sensor locations

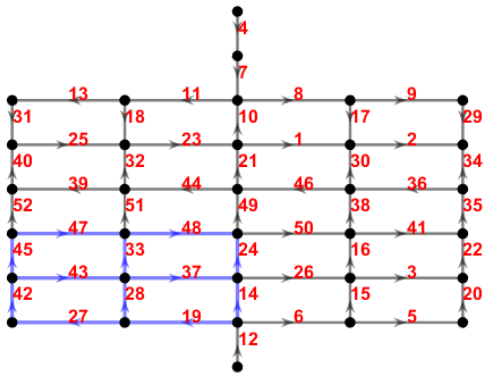


Figure 74 Worst flow-rate sensor locations

Case 8a

Objective function: Asymptotic approximation (129) with the amount of leakage set to the nominal value of 10L/s. This is achieved by using a Dirac delta function for the prior distribution of the model parameters in equation (129) in order to correspond to equation (125).

Nodal demands: Uncertain with all junctions divided into 4 groups with the demand of the junctions in each group following a Gaussian distribution with mean equal to the nominal values of the nodal demands (equal to 10L/s) and standard deviation equal to 20% of the nominal value of the nodal demands, i.e. a Gaussian PDF ~N(10,4). The numerical integration is performed using the Monte Carlo method with 1000 samples drawn from the Gaussian prior. The groups are the following and are also illustrated in the following figures with different colors.

Group 1 comprised of nodes 2, 3, 4, 10, 11, 14, 15 and 16.

Group 2 comprised of nodes 5, 6, 7, 8, 9, 12 and 13.

Group 3 comprised of nodes 20, 21, 24, 25, 26, 28, 29 and 32.

Group 4 comprised of nodes 17, 18, 19, 22, 23, 27, 30 and 31.

Covariance matrix: Proportional model predictions errors (Equation (131)). The prediction error parameter σ is set to 0.01 as nominal value.

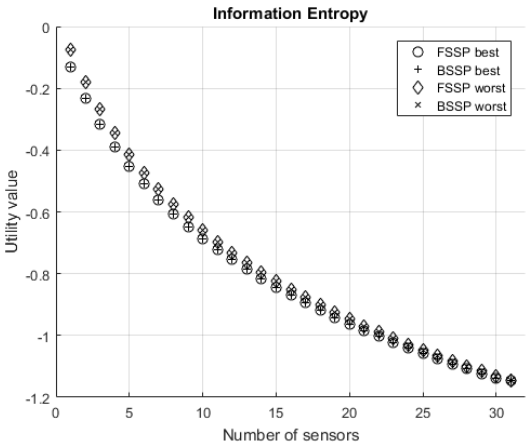


Figure 75 Objective function for pressure sensors

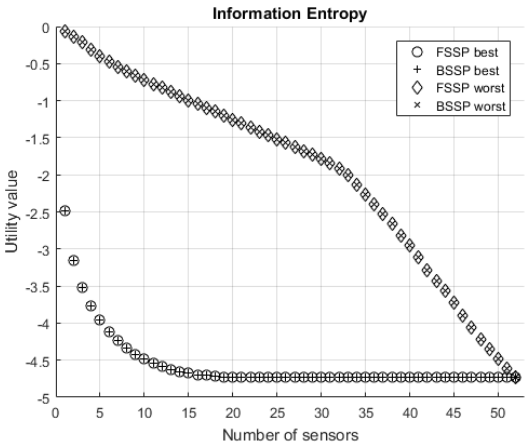


Figure 76 Objective function for flow-rate sensors

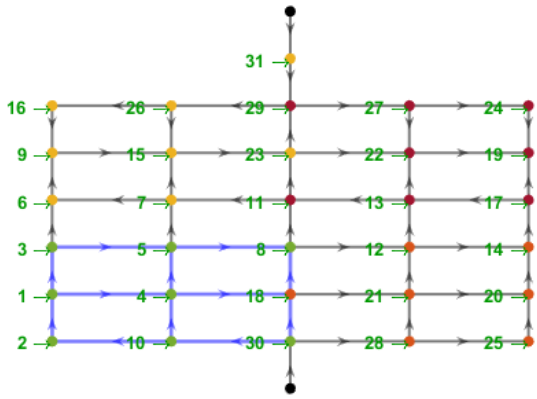


Figure 77 Best pressure sensor locations

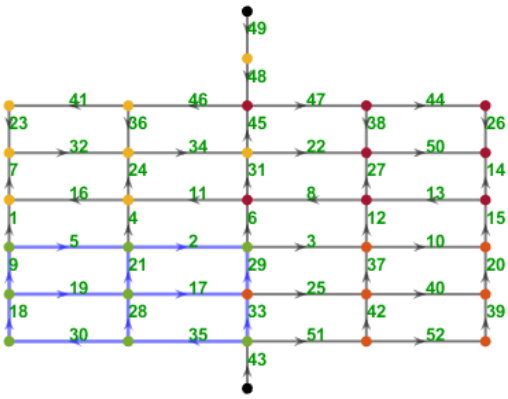


Figure 78 Best flow-rate sensor locations

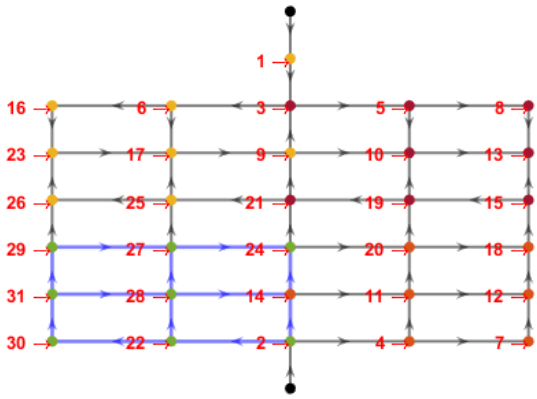


Figure 79 Worst pressure sensor locations

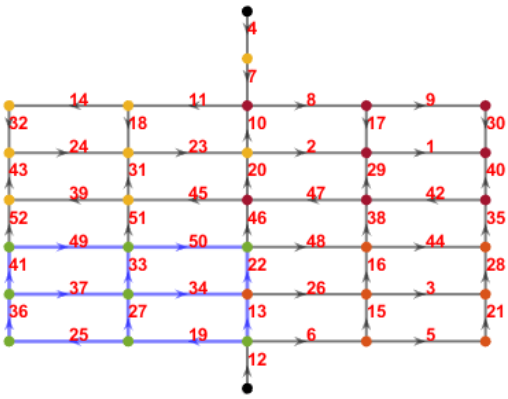


Figure 80 Worst flow-rate sensor locations

Case 8b

Similar to Case 8a apart from the objective function being an asymptotic approximation (129), with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done

using the Monte Carlo method with 1000 samples drawn from the Gaussian prior for the amount of leakage.

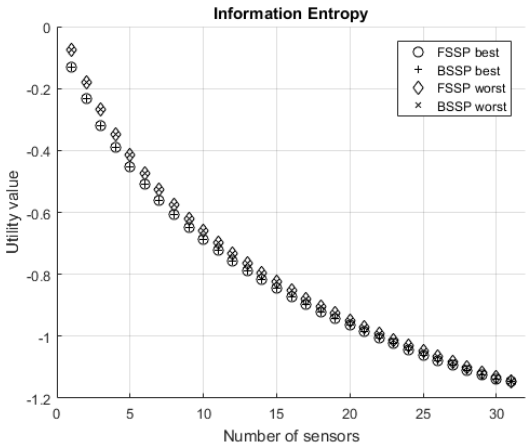


Figure 81 Objective function for pressure sensors

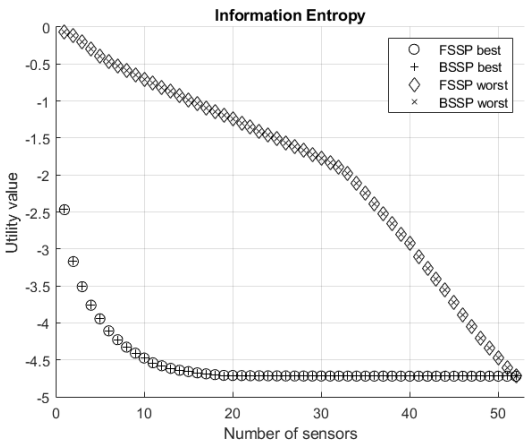


Figure 82 Objective function for flow-rate sensors

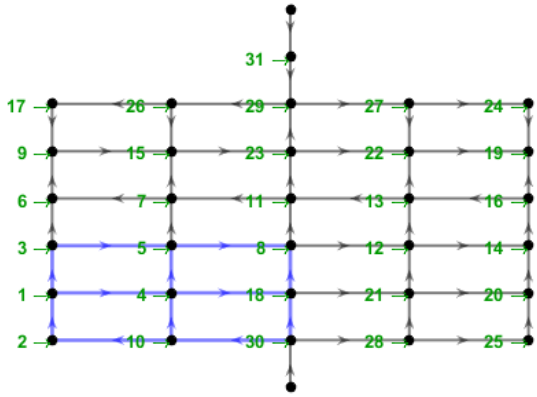


Figure 83 Best pressure sensor locations

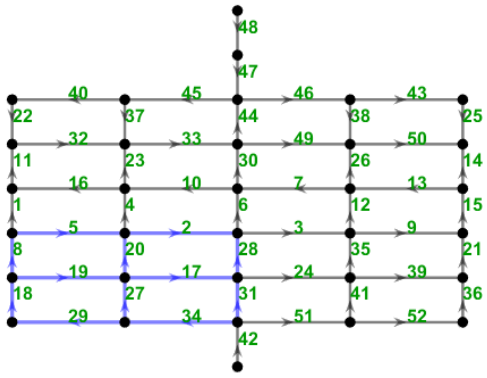


Figure 84 Best flow-rate sensor locations

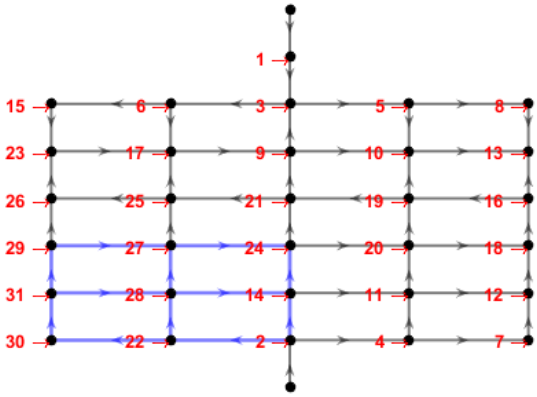


Figure 85 Worst pressure sensor locations

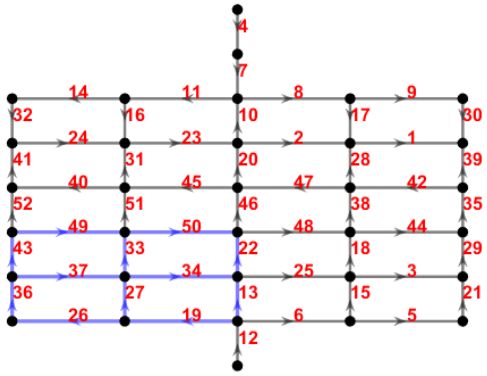


Figure 86 Worst flow-rate sensor locations

Case 8c

Similar to Case 8b apart from the covariance matrix prediction error parameter σ which is set to 0.05 as nominal value.

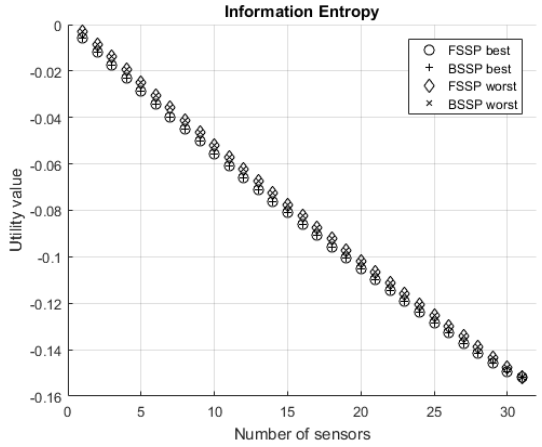


Figure 87 Objective function for pressure sensors

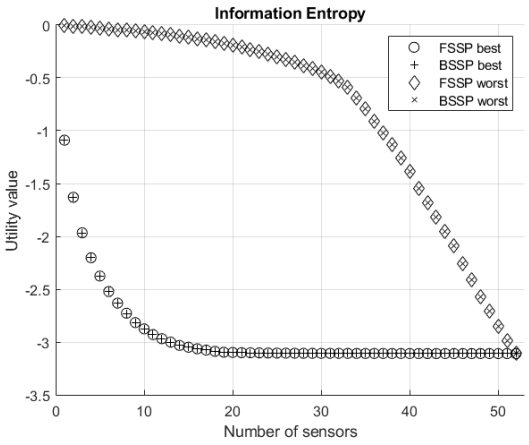


Figure 88 Objective function for flow-rate sensors

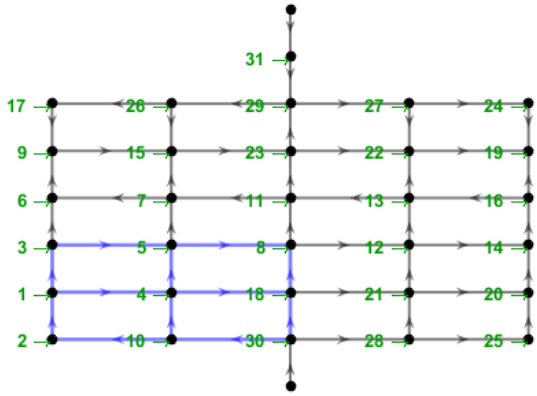


Figure 89 Best pressure sensor locations

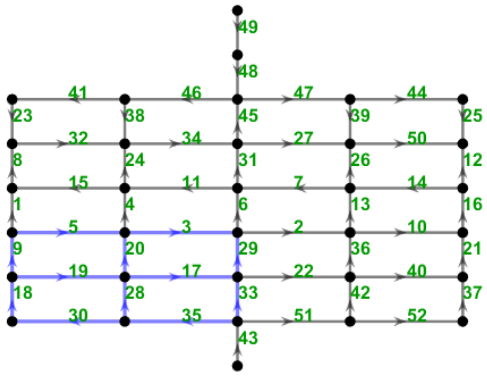


Figure 90 Best flow-rate sensor locations

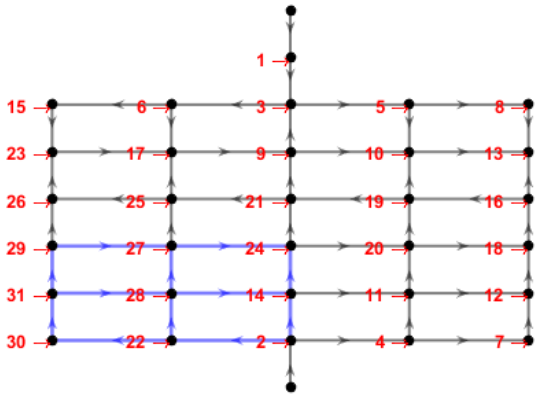


Figure 91 Worst pressure sensor locations

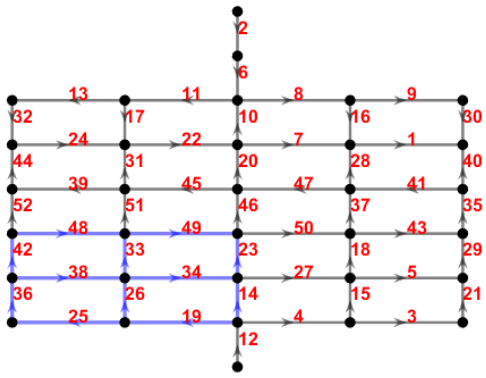


Figure 92 Worst flow-rate sensor locations

Case 9

Objective function: Asymptotic approximation (129) with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done using the Monte Carlo method with 10000 samples drawn from the Gaussian prior for the amount of leakage.

Nodal demands: Uncertain with the demand of all junctions being independent and identically distributed (i.i.d) following a Gaussian distribution with mean equal to the nominal values of the nodal demands (equal to 10L/s) and standard deviation equal to 20% of the nominal value of the nodal demands, i.e. a Gaussian PDF ~N(10,4). The numerical integration is performed using the Monte Carlo method with 10000 samples drawn from the Gaussian prior.

Covariance matrix: Independent of the model predictions errors (Equation (132)). The prediction error parameter σ is set to 0.01 as nominal value.

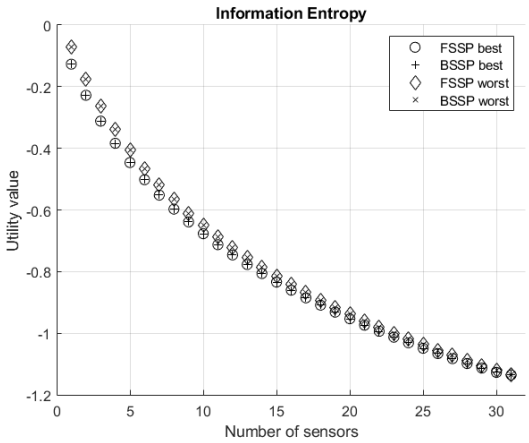


Figure 93 Objective function for pressure sensors

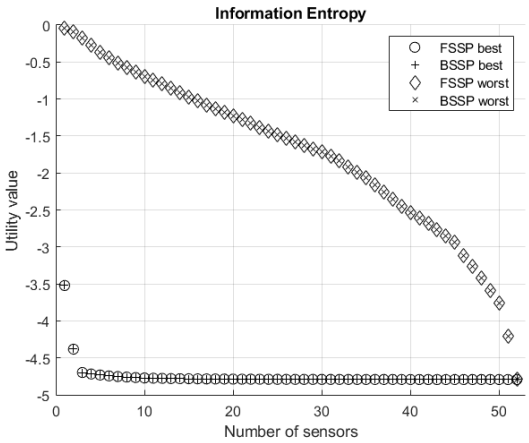


Figure 94 Objective function for flow-rate sensors

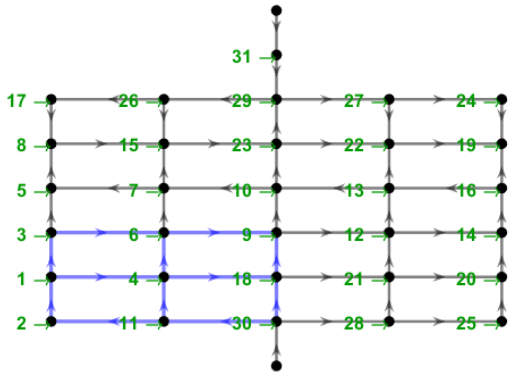


Figure 95 Best pressure sensor locations

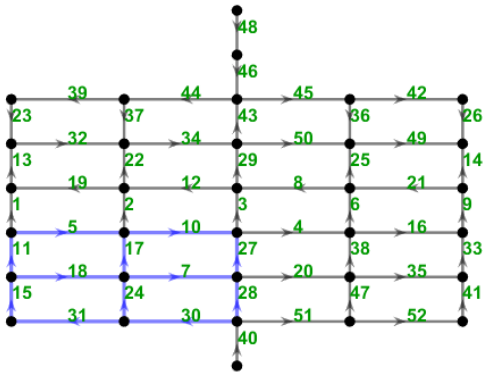


Figure 96 Best flow-rate sensor locations

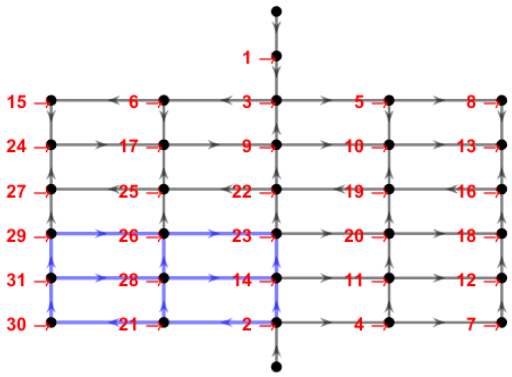


Figure 97 Worst pressure sensor locations

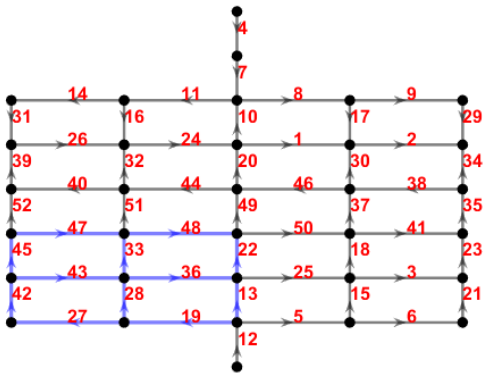


Figure 98 Worst flow-rate sensor locations

Case 10

Objective function: Asymptotic approximation (129) with uncertain prior selected to be Gaussian with mean equal to the nominal amount of leakage 10L/s and standard deviation equal to 20% of the nominal amount of leakage. The integration is done using the Monte Carlo method with 10000 samples drawn from the Gaussian prior for the amount of leakage.

Nodal demands: Uncertain with the demand of all junctions being independent and identically distributed (i.i.d) following a Gaussian distribution with mean equal to the nominal values of the nodal demands (equal to 10L/s) and standard deviation equal to 20% of the nominal value of the nodal demands, i.e. a Gaussian PDF ~N(10,4). The numerical integration is performed using the Monte Carlo method with 10000 samples drawn from the Gaussian prior.

Covariance matrix: Proportional model predictions errors (Equation (131)). The prediction error parameter σ is set to 0.01 as nominal value.

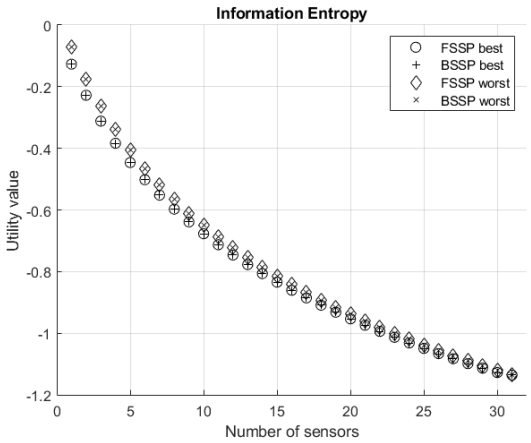


Figure 99 Objective function for pressure sensors

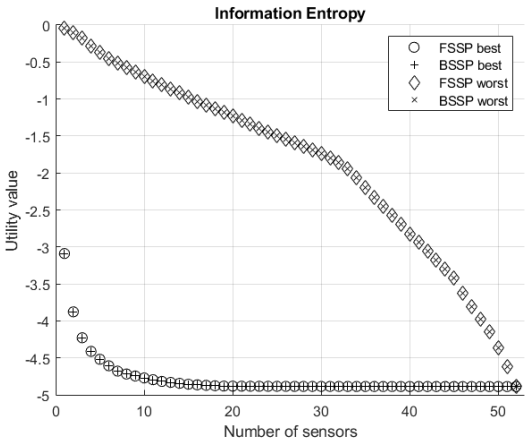


Figure 100 Objective function for flow-rate sensors

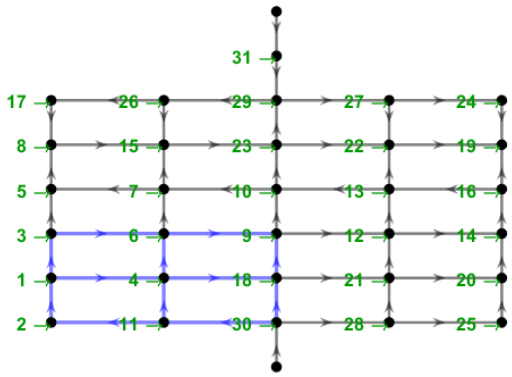


Figure 101 Best pressure sensor locations

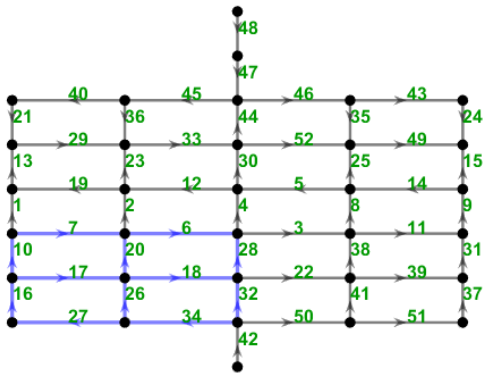


Figure 102 Best flow-rate sensor locations

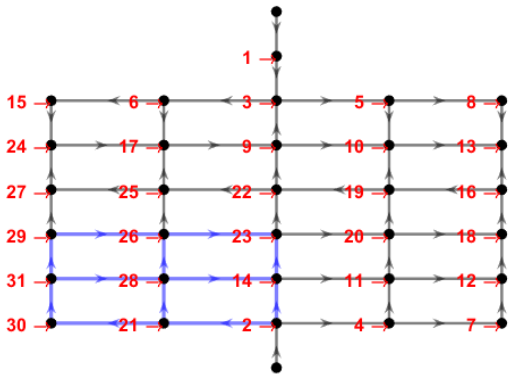


Figure 103 Worst pressure sensor locations

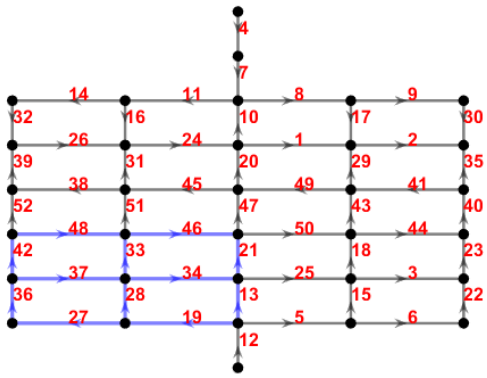


Figure 104 Worst flow-rate sensor locations

6.1 Discussion of OSP results

The general behaviour that can be observed in the objective function plots of all the cases presented in the optimal sensor placement results is that, for pressure sensors, placing an additional sensor will always lead to a reduction in the objective function, or in other words more information. Moreover, the difference in the objective function for a given number of sensors placed in the optimal and in the worst positions is very small compared to the difference observed with the flow-rate sensors. What this means is that the flow-rate sensors are expected to provide more informative measurements for inferring the location and magnitude of leakage. For the flow-rate sensors it seems that the maximum amount of information can be obtained by using only 5-15 sensors optimally placed, depending on the uncertainty that is introduced in the formulation. The difference in the optimal and worst locations is also significant so as in order to get the same amount of information as 5 optimally placed flow-rate sensors, by placing flow-rate sensors in the worst locations, more than 45 sensors must be used. Regarding the FSSP and BSSP heuristic algorithms used to solve the optimization problem, it seems

that both tend to produce the same results for all the cases and number of sensors strengthening the assumption that the solution that is provided is either the optimal or very close to the optimal solution.

Comparing the results of case 1, which uses nominal parameter values to calculate the objective function, and case 2, which is robust by integrating over the prior of the nominal parameters, it can be seen that the objective function results are very similar with exactly the same optimal configurations for pressure sensors and same solution for the first few more informative locations for flow-rate sensors. It seems that the nominal amount of leakage that will be chosen does not significantly affect the optimization results. This can be utilized by avoiding the integration over the prior of the nominal parameters and saving computational time without compromising the accuracy of the results. However, more testing should be conducted to verify this claim.

Comparing case 2 with case 3, with case 2 using the asymptotic approximation and case 3 the sampling approximation, the objective function results are very similar, as are the first 4 more informative optimal locations for the flow-rate sensors. The optimal locations for pressure sensors, however, are different with the asymptotic approximation choosing to place the first few sensors close to the monitored area, while the sampling choosing to spread the sensors more. Verifying which one is more accurate is difficult in this case since the pressure sensors seem to not be able to infer correctly the leaking pipe regardless of the number of sensors. This will become apparent in the next chapter.

Comparing case 2 with case 4, with the difference being that case 2 uses a fixed prediction error covariance matrix and case 4 uses a covariance matrix that is proportional to the model predictions, it can be seen that the objective function results are very similar with the exception being a small increase in the objective function in the first more informative locations for the flow-rate sensors for the case 4. In other words, using a proportional to model predictions covariance matrix introduces some uncertainty to the system and in the case presented requires 1 or 2 more sensors to be placed to compensate for this. The optimal locations are the same for the pressure sensors and same for the first few more informative sensor locations for flow-rate sensors with small differences in the rest less informative sensor locations.

Comparing case 1 with case 5, with the difference being that case 5 has uncertainty introduced in the nodal demands with all the nodes in the same group, the objective function is very similar for both with the same optimal locations for the pressure sensors and same first few more informative flow-rate sensor locations with small differences in the next less informative locations. It seems that introducing uncertainty in the nodal demands while using a fixed prediction error covariance matrix does not significantly affect the results. Cases 5a and 5b are very similar supporting the claim that

using nominal parameter values for the amount of leakage provide the same results as using the robust formulation.

Comparing case 5 with case 6, with the difference being that case 5 uses an independent prediction error covariance matrix and case 6 uses a proportional to model predictions covariance matrix, the objective functions are similar with a small difference in the first few flow-rate sensors where 1 or 2 more optimally placed sensors are needed to get the maximum information in case 6. The optimal locations for pressure sensors are very similar while the flow-rate sensors have some differences. Cases 6a and 6b have very similar results.

Comparing case 5 with case 7a, with the difference being that case 5 has all the junctions in one group with uncertain demand while case 7a splits the junctions into 4 groups, as was also the case 1 – case 5 comparison the objective function and optimal locations are very similar strengthening the assumption that the additional uncertainty in the demands with an independent prediction error covariance matrix does not significantly affect the results. Cases 7a and 7b have very similar results.

Comparing case 7a with case 7c, with the difference being that case 7a has prediction error parameter $\sigma = 0.01$ and case 7c has $\sigma = 0.05$, it can be seen that the total information that can be obtained from case 7c is less that the 7a since the objective function has higher value when all sensors are placed in 7c. This is intuitive since adding more noise to the measurements reduces the amount of usable information that can be extracted from the data. Using flow-rate sensors 5-10 sensors optimally placed in both cases can provide almost the maximum amount of information. However, the sensors placed in the worst locations are less informative when more noise is added in the measurements. The optimal locations are the same for the first few more informative sensor locations.

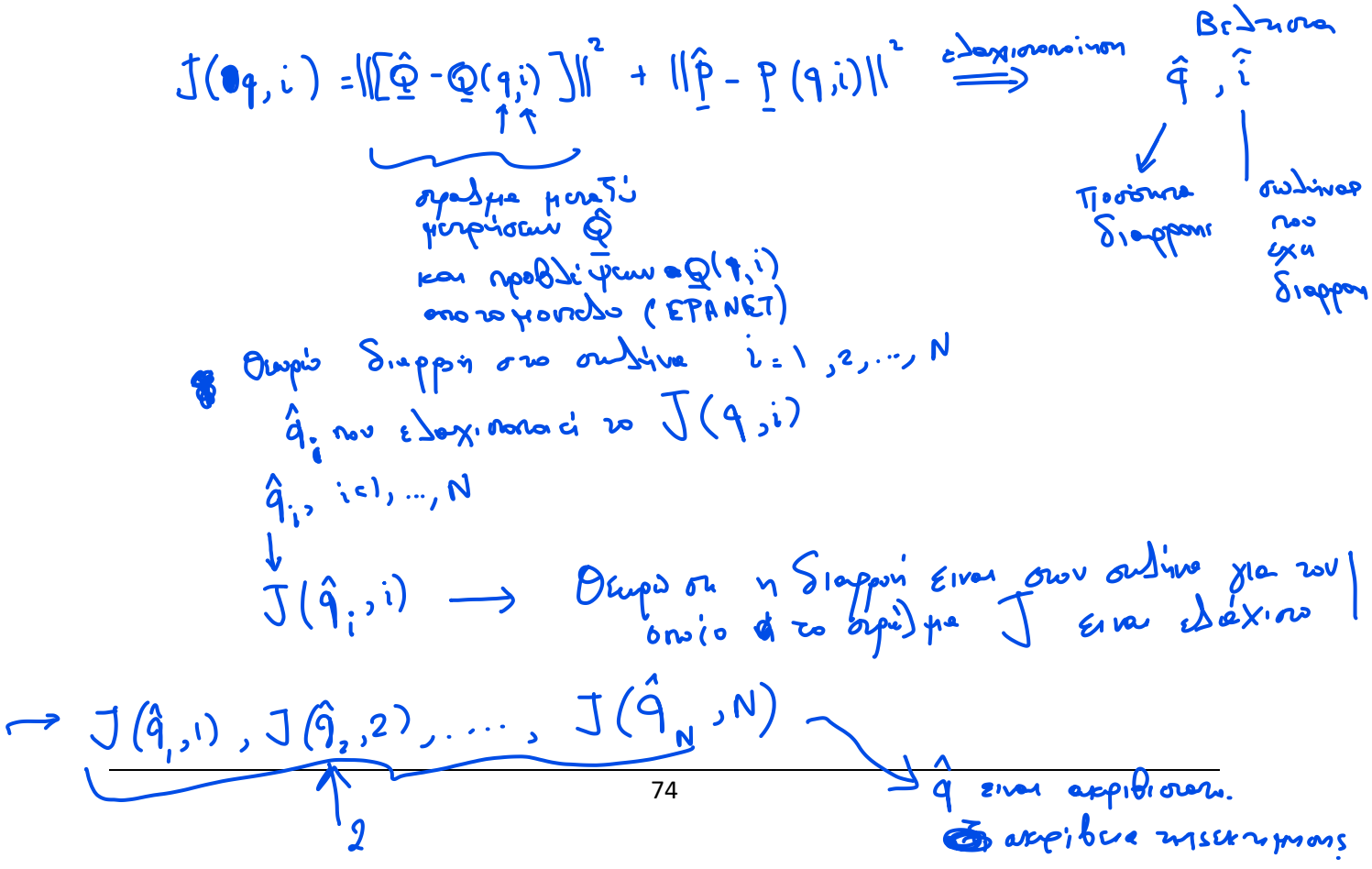
Comparing cases 7a,b and 8a,b, with the difference being that case 8 uses a prediction error covariance matrix that is proportional to the model predictions, the objective function and optimal locations for pressure sensors are very similar. However, for the flow-rate sensors the difference in the objective function is more prominent with case 8 needing around 15 optimally placed sensors to gain the maximum amount of information, while case 7 needs only 5 sensors. The uncertainty in case 8 is more significant requiring more sensors in order to gain the same amount of information. Cases 8a and 8b are very similar.

Comparing case 8a to case 8c, with the difference being that case 8a has prediction error parameter $\sigma = 0.01$ and case 8c has $\sigma = 0.05$, as was also the case 7a – case 7c comparison, adding more noise to the measurements reduces the amount of usable information that can be extracted from the data.

It can also be seen that for case 8c the sensors placed in the worst locations provide less information with the added noise.

Comparing cases 7a,b to case 9, with the difference being that case 7 has the nodes split into 4 groups with the nodal demands being independent for each group and case 9 has all nodal demands being independent, the objective functions are very similar requiring the same number of sensors. For the optimal sensor locations there are minor differences for the pressure sensors and for the flow-rate sensors the first few more informative locations are the same with minor differences on the rest of the sensors. Cases 7a and 7b are very similar.

Lastly, comparing cases 8a,b to case 10, with the difference being that case 8 has the nodes split into 4 groups with the nodal demands being independent for each group and case 10 has all nodal demands being independent, case 8 requires about 2 more sensors than case 10 to achieve the same fraction of the total information i.e. when all flow-rate sensors are placed. This is counterintuitive since case 10 has more uncertainty presented as each nodal demand is independent of the rest. A possible explanation to this is that one of the groups used in case 8 contains completely the monitored area. This makes the total demand of the area to vary much more with each sample that case 10 where the independent demands of the nodes average out for the total demand of the area. The optimal locations for pressure sensors are the same for the two cases but the flow-rate locations have differences even in the first few more informative sensors locations. Cases 8a and 8b are very similar.



7. Evaluation of optimal sensor configurations

In this chapter the optimal sensor configurations presented in chapter 6 will be evaluated by using these sensor configurations to locate simulated leaks. A single leakage will be added to a pipe in the monitored area and then a small number of sensors placed in optimal locations will be used to infer the location and magnitude of the leakage. The procedure for identifying leakage will be repeated assuming leakage at different pipes, specifically pipe 3, or pipe 12, or pipe 17, or pipe 19 in the monitored area. For each leaking pipe, results will be repeated 20 times assuming different noise in the measurements. Specifically, the measurements will be simulated from the nominal WDN model and noise will be added to the predictions at the sensor locations. Different error samples will be added each time to the model predictions to account for the measurement and model error in the Bayesian inference framework. Two values of the prediction error covariance matrix parameter σ will be used, $\sigma = 0.01$ and $\sigma = 0.05$. The effectiveness of the sensor configurations will be evaluated by counting the number of times the correct pipe was inferred and if this is not the case, it will be checked if the inferred pipe is one pipe away or two pipes away from the actual leaking pipe. Also, the accuracy of inferring the amount of leakage on the actual leaking pipe, regardless if it was correctly inferred or not, will be evaluated by calculating the mean and the standard deviation of the values inferred. Lastly, another measure that will be used is, in the cases where the leaking pipe is correctly inferred, the average difference in the objective function (123) between the most probable pipe and the second most probable pipe. If this difference has a high value, it will indicate that the measurements show that the correct pipe is much more probable to be the leaking pipe than the rest of the pipes.

In order to check the effectiveness of the optimal sensor configuration in identifying the location and the magnitude of leakage, the leakage identification results will also be obtained from sub-optimal sensor configurations, which will be compared to the leakage identification results obtained from the optimal sensor configuration. The suboptimal sensor configurations consist of one that involves the same number of sensors placed in the worst sensor locations and also with a second sensor configuration with even more sensors placed in the worst locations, as long as this configuration has higher information entropy that the optimal configuration which is used.

The amount of leakage used to simulate the measurements in order to perform leakage detection in all the following cases is 10L/s. The covariance matrix that is used is equation (106), which is proportional to the measurements. In order to infer the amount of leakage the equation (112) is used and in order to choose the most probable of the models the equation (123) is used.

Case A

The first case that is explored corresponds to the cases 1 to 4 presented in the previous chapter with nominal values used for the nodal demands, using flow-rate sensors. It can be seen from the objective function diagrams (Figure 10, 16, 22 and 28) that by using 5 flow-rate sensors optimally placed almost the maximum amount of information can be obtained. The best sensor locations in these cases are the same for the first 4 flow-rate sensors with small differences in the 5^{th} sensor, however the objective function reduction resulting from adding the 5^{th} sensor is very small so the difference shouldn't be significant. The 5 optimal locations that will be used are pipes 32, 30, 28, 22 and 20 which can be seen in Figure 5 or equivalently the first 5 locations in Figure 12. Then, for the first suboptimal sensor configuration 25 sensors are placed in the worst locations which are pipes 39, 14, 41, 52, 47, 4, 5 ,51, 49, 42, 43, 1, 8, 45, 7, 44, 48, 16, 2, 36, 17, 6, 15, 13 and 37 and last the second suboptimal sensor configuration with 5 sensors in the worst locations in pipes 39, 14, 41, 52 and 47.

The following tables contain all the information that will be used to evaluate the optimal sensor configurations by assuming leakage in pipes 3, 12, 17 and 19 respectively. For each leaking pipe the first table corresponds to prediction error parameter $\sigma = 0.01$ and the second table corresponds to $\sigma = 0.05$.

Table 1 Results of flow-rate sensors for leaking pipe 3 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 3	5 sensors at optimal location	25 sensors at worst location	5 sensors at worst location
Correct pipe inferred	20	17	2
1 pipe away inferred	-	2	11
2 pipes away inferred	-	1	7
Wrong pipe inferred	-	-	-
Mean of leak (L/s)	9.999	9.974	10.276
Standard deviation of leak	5.65E-03	5.71E-01	2.09E+00
Average diff. in the objective function	4.18E+00	2.84E+00	1.37E-01

Table 2 Results of flow-rate sensors for leaking pipe 3 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 3	5 sensors at optimal location	25 sensors at worst location	5 sensors at worst location
Correct pipe inferred	15	7	1
1 pipe away inferred	-	5	2
2 pipes away inferred	5	5	9
Wrong pipe inferred	-	3	8
Mean of leak (L/s)	10.011	9.949	13.297
Standard deviation of leak	4.87E-02	1.94E+00	8.40E+00
Average diff. in the objective function	1.21E+00	2.70E-01	1.93E-01

Table 3 Results of flow-rate sensors for leaking pipe 12 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 12	5 sensors at optimal location	25 sensors at worst location	5 sensors at worst location
Correct pipe inferred	20	20	15
1 pipe away inferred	-	-	5
2 pipes away inferred	-	-	-
Wrong pipe inferred	-	-	-
Mean of leak (L/s)	9.991	9.882	9.826
Standard deviation of leak	5.88E-02	4.48E-01	1.23E+00
Average diff. in the objective function	8.90E+00	2.18E+01	1.06E+00

Table 4 Results of flow-rate sensors for leaking pipe 12 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 12	5 sensors at optimal location	25 sensors at worst location	5 sensors at worst location
Correct pipe inferred	20	16	4
1 pipe away inferred	-	4	8
2 pipes away inferred	-	-	4

Wrong pipe inferred	-	-	4
Mean of leak (L/s)	10.082	9.518	10.781
Standard deviation of leak	3.09E-01	1.94E+00	8.55E+00
Average diff. in the objective function	3.94E+00	1.89E+00	2.50E-01

Table 5 Results of flow-rate sensors for leaking pipe 17 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 17	5 sensors at optimal location	25 sensors at worst location	5 sensors at worst location
Correct pipe inferred	20	20	5
1 pipe away inferred	-	-	2
2 pipes away inferred	-	-	12
Wrong pipe inferred	-	-	1
Mean of leak (L/s)	10.000	9.936	9.540
Standard deviation of leak	1.00E-03	3.84E-01	1.54E+00
Average diff. in the objective function	1.26E+01	3.11E+01	6.80E-01

Table 6 Results of flow-rate sensors for leaking pipe 17 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 17	5 sensors at optimal location	25 sensors at worst location	5 sensors at worst location
Correct pipe inferred	20	20	2
1 pipe away inferred	-	-	4
2 pipes away inferred	-	-	7
Wrong pipe inferred	-	-	7
Mean of leak (L/s)	9.998	8.882	12.501
Standard deviation of leak	6.04E-03	1.25E+00	5.34E+00
Average diff. in the objective function	7.93E+00	4.29E+00	2.81E-01

Table 7 Results of flow-rate sensors for leaking pipe 19 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 19	5 sensors at optimal location	25 sensors at worst location	5 sensors at worst location
Correct pipe inferred	20	16	8
1 pipe away inferred	-	3	8
2 pipes away inferred	-	1	2
Wrong pipe inferred	-	-	2
Mean of leak (L/s)	9.993	9.847	9.473
Standard deviation of leak	2.27E-02	4.73E-01	2.07E+00
Average diff. in the objective function	9.40E+00	2.65E+00	4.24E-01

Table 8 Results of flow-rate sensors for leaking pipe 19 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 19	5 sensors at optimal location	25 sensors at worst location	5 sensors at worst location
Correct pipe inferred	20	1	1
1 pipe away inferred	-	4	3
2 pipes away inferred	-	6	7
Wrong pipe inferred	-	9	9
Mean of leak (L/s)	10.026	8.683	7.833
Standard deviation of leak	1.27E-01	2.84E+00	7.93E+00
Average diff. in the objective function	4.19E+00	4.91E-02	9.31E-02

Case B

This case is the same as case A but will use pressure sensors instead of flow-rate sensors. As can be seen from the objective function diagrams (Figure 9, 15, 21 and 27) adding one more sensor will always result in more information gained. Also, the difference between sensors placed in optimal and in worst locations is not significant so only sensors in optimal locations will be considered. In this case the sensor configurations that will be used are 10, 20 and 31 sensors (maximum number of sensors) at

optimal locations which can be seen in Figure 11. The prediction error parameter that is used is $\sigma = 0.01$.

Table 9 Results of pressure sensors for leaking pipe 3

Leaking pipe: 3	10 sensors at optimal location	20 sensors at optimal location	31 sensors at optimal location
Correct pipe inferred	-	-	-
1 pipe away inferred	-	-	-
2 pipes away inferred	-	-	10
Wrong pipe inferred	20	20	10
Mean of leak (L/s)	10.224	10.041	10.228
Standard deviation of leak	1.12E+00	9.60E-01	5.85E-01
Average diff. in the objective function	-	-	-

Table 10 Results of pressure sensors for leaking pipe 12

Leaking pipe: 12	10 sensors at optimal location	20 sensors at optimal location	31 sensors at optimal location
Correct pipe inferred	-	-	-
1 pipe away inferred	-	-	-
2 pipes away inferred	-	-	11
Wrong pipe inferred	20	20	9
Mean of leak (L/s)	10.199	9.978	10.080
Standard deviation of leak	1.76E+00	9.04E-01	6.21E-01
Average diff. in the objective function	-	-	-

Table 11 Results of pressure sensors for leaking pipe 17

Leaking pipe: 17	10 sensors at optimal location	20 sensors at optimal location	31 sensors at optimal location
Correct pipe inferred	-	-	-

1 pipe away inferred	-	-	-
2 pipes away inferred	-	-	9
Wrong pipe inferred	20	20	11
Mean of leak (L/s)	10.239	10.029	10.029
Standard deviation of leak	9.90E-01	1.21E+00	8.93E-01
Average diff. in the objective function	-	-	-

Table 12 Results of pressure sensors for leaking pipe 19

Leaking pipe: 19	10 sensors at optimal location	20 sensors at optimal location	31 sensors at optimal location
Correct pipe inferred	-	-	-
1 pipe away inferred	-	-	-
2 pipes away inferred	-	-	-
Wrong pipe inferred	20	20	20
Mean of leak (L/s)	10.416	10.190	10.399
Standard deviation of leak	1.04E+00	1.02E+00	6.15E-01
Average diff. in the objective function	-	-	-

Case C

In this case the model used to simulate the measurements will use perturbed nodal demands. All the junctions will be split into 4 groups, as was also the case on cases 7 and 8 on the optimal sensor placement results. The groups will be the same as before and can also be seen below this paragraph. The demands of junctions in the same group will be the same and the demand of each group will follow a Gaussian distribution with 10L/s mean and standard deviation 20% the mean value, i.e. the Gaussian PDF is $\sim N(10\,,4)$. The pipes that the leakage will be introduced will again be pipes 3, 12, 17 and 19. The sensors that will be used here are, 10 flow-rate sensors at optimal locations, 25 flow-rate sensors at worst locations and 10 flow-rate sensors at worst locations. The sensor configurations that will be used are the results of case 8 and can be seen in Figure 78 and Figure 80 and are also mentioned

below this paragraph. The uncertainty in nodal demands will not be tested with pressure sensors since they struggled to locate correctly the leaking pipe in case B even without the extra uncertainty.

Group 1 comprised of nodes 2, 3, 4, 10, 11, 14, 15 and 16.
Group 2 comprised of nodes 5, 6, 7, 8, 9, 12 and 13.
Group 3 comprised of nodes 20, 21, 24, 25, 26, 28, 29 and 32.
Group 4 comprised of nodes 17, 18, 19, 22, 23, 27, 30 and 31.

10 sensors at optimal locations placed in pipes: 32, 21, 22, 30, 20, 28, 33, 27, 19 and 23
25 sensors at worst locations placed in pipes: 41, 39, 14, 52, 5, 4, 51, 47, 49, 42, 43, 1, 8, 45, 7, 16, 48, 44, 2, 36, 6, 17, 37, 34 and 3
10 sensors at worst locations placed in pipes: 41, 39, 14, 52, 5, 4, 51, 47, 49 and 42

Table 13 Results of flow-rate sensors for leaking pipe 3 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 3	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	20	9
1 pipe away inferred	-	-	6
2 pipes away inferred	-	-	4
Wrong pipe inferred	-	-	1
Mean of leak (L/s)	10.004	9.791	9.704
Standard deviation of leak	4.98E-02	3.07E-01	1.15E+00
Average diff. in the objective function	1.02E+01	2.12E+01	4.75E-01

Table 14 Results of flow-rate sensors for leaking pipe 3 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 3	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	18	18	3
1 pipe away inferred	1	2	4
2 pipes away inferred	1	-	8
Wrong pipe inferred	-	-	5
Mean of leak (L/s)	9.989	9.841	10.240
Standard deviation of leak	2.71E-01	1.75E+00	6.11E+00

Average diff. in the objective function	3.06E+00	2.83E+00	1.28E-01

Table 15 Results of flow-rate sensors for leaking pipe 12 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 12	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	20	17
1 pipe away inferred	-	-	2
2 pipes away inferred	-	-	1
Wrong pipe inferred	-	-	-
Mean of leak (L/s)	10.013	9.835	9.760
Standard deviation of leak	8.50E-02	6.66E-01	9.79E-01
Average diff. in the objective function	1.98E+01	1.78E+01	2.43E+00

Table 16 Results of flow-rate sensors for leaking pipe 12 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 12	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	12	6
1 pipe away inferred	-	8	4
2 pipes away inferred	-	-	8
Wrong pipe inferred	-	-	2
Mean of leak (L/s)	10.049	9.519	8.549
Standard deviation of leak	3.57E-01	1.72E+00	4.03E+00
Average diff. in the objective function	7.97E+00	1.29E+00	5.20E-01

Table 17 Results of flow-rate sensors for leaking pipe 17 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 17	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location

Correct pipe inferred	20	20	13
1 pipe away inferred	-	-	3
2 pipes away inferred	-	-	3
Wrong pipe inferred	-	-	1
Mean of leak (L/s)	9.988	10.159	10.271
Standard deviation of leak	5.22E-02	3.74E-01	1.17E+00
Average diff. in the objective function	2.81E+01	2.71E+01	2.91E+00

Table 18 Results of flow-rate sensors for leaking pipe 17 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 17	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	19	1
1 pipe away inferred	-	-	5
2 pipes away inferred	-	1	9
Wrong pipe inferred	-	-	5
Mean of leak (L/s)	9.984	9.247	10.444
Standard deviation of leak	2.02E-01	1.46E+00	4.22E+00
Average diff. in the objective function	1.63E+01	3.09E+00	4.69E-02

Table 19 Results of flow-rate sensors for leaking pipe 19 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 19	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	18	7
1 pipe away inferred	-	2	6
2 pipes away inferred	-	-	7
Wrong pipe inferred	-	-	-
Mean of leak (L/s)	9.995	9.922	9.933

Standard deviation of leak	4.44E-02	5.17E-01	1.26E+00
Average diff. in the objective function	2.67E+01	4.24E+00	6.66E-01

Table 20 Results of flow-rate sensors for leaking pipe 19 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 19	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	2	1
1 pipe away inferred	-	9	1
2 pipes away inferred	-	8	5
Wrong pipe inferred	-	1	13
Mean of leak (L/s)	9.911	8.958	7.832
Standard deviation of leak	2.61E-01	2.36E+00	6.65E+00
Average diff. in the objective function	1.38E+01	1.22E-01	4.33E-02

Case D

In this case the model used to simulate the measurements will use perturbed nodal demands. The demand of each junction will be uncertain and independent of the rest, as was the case on cases 9 and 10 of the optimal sensor placement results. The demands of the junctions follow a Gaussian distribution with 10L/s mean and standard deviation 20% the mean value, i.e. the Gaussian PDF is $\sim N(10\,,4)$. The pipes that the leakage is introduced are pipes 3, 12, 17 and 19. The sensors that are used here are, 10 flow-rate sensors at optimal locations, 25 flow-rate sensors at worst locations and 10 flow-rate sensors at worst locations. The sensor configurations that will be used are the results of case 10 and can be seen in Figure 102 and are also mentioned below this paragraph.

10 sensors at optimal locations placed in pipes: 32, 30, 22, 28, 27, 21, 20, 26, 24 and 19
25 sensors at worst locations placed in pipes: 39, 41, 14, 52, 4, 5, 51, 47, 49, 42, 43, 1, 8, 45, 7 ,44, 48, 16, 2, 36, 17, 6, 15, 37 and 13
10 sensors at worst locations placed in pipes: 39, 41, 14, 52, 4, 5, 51, 47, 49 and 42

Table 21 Results of flow-rate sensors for leaking pipe 3 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 3	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	15	11
1 pipe away inferred	-	1	3
2 pipes away inferred	-	4	3
Wrong pipe inferred	-	-	3
Mean of leak (L/s)	10.001	10.083	9.801
Standard deviation of leak	2.64E-02	4.01E-01	1.22E+00
Average diff. in the objective function	1.33E+01	2.00E+00	3.14E-01

Table 22 Results of flow-rate sensors for leaking pipe 3 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 3	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	7	3
1 pipe away inferred	-	5	2
2 pipes away inferred	-	5	5
Wrong pipe inferred	-	3	10
Mean of leak (L/s)	9.978	9.439	10.611
Standard deviation of leak	1.07E-01	1.91E+00	6.22E+00
Average diff. in the objective function	3.33E+00	2.84E-01	6.82E-02

Table 23 Results of flow-rate sensors for leaking pipe 12 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 12	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	20	19
1 pipe away inferred	-	-	1
2 pipes away inferred	-	-	-

Wrong pipe inferred	-	-	-
Mean of leak (L/s)	10.003	10.043	10.461
Standard deviation of leak	3.71E-02	4.66E-01	1.09E+00
Average diff. in the objective function	2.36E+01	2.03E+01	2.84E+00

Table 24 Results of flow-rate sensors for leaking pipe 12 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 12	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	16	3
1 pipe away inferred	-	3	2
2 pipes away inferred	-	1	9
Wrong pipe inferred	-	-	6
Mean of leak (L/s)	9.986	9.016	9.222
Standard deviation of leak	1.92E-01	2.23E+00	5.51E+00
Average diff. in the objective function	9.78E+00	1.20E+00	5.02E-01

Table 25 Results of flow-rate sensors for leaking pipe 17 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 17	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	20	16
1 pipe away inferred	-	-	1
2 pipes away inferred	-	-	2
Wrong pipe inferred	-	-	1
Mean of leak (L/s)	10.001	9.954	10.068
Standard deviation of leak	7.49E-03	3.02E-01	8.79E-01
Average diff. in the objective function	3.35E+01	3.28E+01	1.57E+00

Table 26 Results of flow-rate sensors for leaking pipe 17 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 17	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	20	1
1 pipe away inferred	-	-	8
2 pipes away inferred	-	-	7
Wrong pipe inferred	-	-	4
Mean of leak (L/s)	9.974	9.333	9.568
Standard deviation of leak	1.39E-01	1.35E+00	4.64E+00
Average diff. in the objective function	2.17E+01	5.29E+00	1.22E-02

Table 27 Results of flow-rate sensors for leaking pipe 19 and $\boldsymbol{\sigma = 0.01}$

Leaking pipe: 19	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	14	6
1 pipe away inferred	-	4	6
2 pipes away inferred	-	2	7
Wrong pipe inferred	-	-	1
Mean of leak (L/s)	10.004	9.936	9.672
Standard deviation of leak	2.80E-02	4.48E-01	1.01E+00
Average diff. in the objective function	2.87E+01	2.58E+00	3.69E-01

Table 28 Results of flow-rate sensors for leaking pipe 19 and $\boldsymbol{\sigma = 0.05}$

Leaking pipe: 19	10 sensors at optimal location	25 sensors at worst location	10 sensors at worst location
Correct pipe inferred	20	-	-

1 pipe away inferred	-	4	-
2 pipes away inferred	-	6	10
Wrong pipe inferred	-	10	10
Mean of leak (L/s)	10.017	8.475	8.753
Standard deviation of leak	1.22E-01	2.74E+00	4.46E+00
Average diff. in the objective function	1.54E+01	-	-

7.1 Discussion of leakage detection results

In this chapter the optimal sensor placement results are applied to detect simulated leaks in order to verify the effectiveness of the optimal sensor configurations. In case A where flow-rate sensors are used on a model with nominal value deterministic demands, it seems that by using 5 optimally placed sensors the leaking pipe is consistently inferred correctly for $\sigma = 0.01$ and almost all of the times for $\sigma = 0.05$. Furthermore, the mean value of the magnitude of leakage is very close to the actual amount of leakage, 10L/s, with very small variance, so the magnitude of leakage is also consistently inferred accurately. Using now 5 sensors on the worst locations, most of the times the pipe that is inferred is not the actual leaking pipe. However, the inferred pipe is most of the times close to the leaking pipe with a few exceptions where the inferred pipe is further away. The mean of the magnitude of leakage is less accurate and the standard deviation of it is 2 to 3 orders of magnitude larger. The results for 25 sensors placed at the worst locations are, as expected, between the 5 sensors best and worst configurations. With this configuration most of the times the correct pipe is inferred, while the rest of the times the pipe that is inferred is very close to the leaking pipe. The mean of the magnitude of leakage is less accurate than the 5 sensors best configuration and more accurate that then 5 sensor worst configuration, with the standard deviation of it being 1 to 2 orders of magnitude larger than the 5 sensor best configuration.

The average difference of the objective function between the leaking pipe and the second most probable pipe usually is larger for the 5 optimally placed sensors compared to the 2 worst configurations, however, this measure does not seem to be very consistent, especially when the number of times that the correct pipe is inferred is very small. Comparing the results of the two values for the prediction error parameter σ, as expected, increasing the noise in the measurements reduces the accuracy of inferring the magnitude of leakage and reduces the number of times that the leaking pipe is inferred correctly. The optimal sensor configuration, even with the increased noise, can still

reliably locate the leaking pipe and has high accuracy on inferring the magnitude of leakage. The two suboptimal sensor configurations are affected more by the increased noise as the number of times that the correct pipe is inferred is reduced.

In case B where pressure sensors are used, even when every junction has a pressure sensor, 31 junctions with 31 pressure sensors, the inferred pipe is consistently wrong. The best result that was obtained is when 31 sensors were used and the inferred pipe was 2 pipes away from the leaking pipe. It is noted that adding more sensors increases the accuracy of inferring the magnitude of leakage and leads to smaller variance. The results indicate that flow-rate sensors are much more informative that pressure sensors which corroborates with the objective function results that were presented in chapter 6. A possible explanation to this is that the pressure of the system is not very sensitive to the location of the leakage. When a leak is introduced into the network the pressure of the whole system drops instead of only to an area close to the leaking pipe. This can be seen in the pressure head results of the network without leakage in Figure 8. The pressure of all the junctions is roughly the same, and when a leakage is introduced the pressure head of every junction reduces by approximately the same amount.

In case C where flow-rate sensors are used with uncertain demands in junctions, which are split into 4 groups, the results are very similar to the results of case A. Using 10 flow-rate sensors optimally placed is enough to consistently find the location of the leaking pipe and accurately infer the magnitude of leakage with small variance for both σ values. Using 10 sensors at the worst locations the leaking pipe cannot be located consistently and when the noise is increased the leaking pipe is rarely located correctly. The standard deviation of the inferred magnitude of leakage is 1 to 2 orders of magnitude larger than when using 10 optimally placed sensors. The results for 25 sensors in the worst locations are also not as reliable as 10 optimally placed sensors with the difference being more significant when more noise is added to the measurements. The variance of the magnitude of leakage is also larger. Case C uses 10 flow-rate sensors optimally placed instead of 5 to compensate for the additional uncertainty introduced to the nodal demands and it should be noted that the standard deviation in the inferred magnitude of leakage in case C is higher than case A which has less sensors. This corroborates with the objective function results for these cases since the 5 sensors configuration with deterministic demands has lower value in the objective function than the 10 sensors configuration with uncertain nodal demands.

In case D the nodal demand of each junction is uncertain and independent of the rest of the nodal demands. As is the trend with the previous cases the optimal sensor configuration locates correctly the leaking pipe even when the prediction error parameter has the highest of the two values that are

used. The magnitude of leakage is accurately inferred with small variance and it should be noted here that the variance for the optimal sensor configuration in case D is smaller than the variance of the optimal sensor configuration in case C which corroborates with the objective function values of the respective OSP results. The suboptimal configuration with 10 sensors is the least accurate of the three configurations and seems to struggle to locate the correct pipe when more noise is added to the measurements. The standard deviation of this configuration is 1 to 2 orders of magnitude larger than the optimal configuration. The suboptimal configuration with 25 sensors gives results which are worse than the optimal configuration and better than the 10 sensor suboptimal configuration.

8. Conclusions

An entropy based optimal experimental design methodology for locating leaks in water distribution networks using pressure and flow-rate sensors is developed. The hydraulic model of the system is formulated to aid in providing the necessary hydraulic quantities of interest, as is the Bayesian system identification technique used to locate leaks. The proposed OSP methodology is augmented to also take into consideration the unavoidable uncertainties present in model parameters to be able to give robust optimal sensor configurations. The effectiveness of the resulting optimal sensor configurations is demonstrated by using these sensor configurations to provide simulated measurements to the system identification technique and measure the accuracy of the inferred results.

Several conclusions were drawn through this study which are the following. First, pressure sensors are not sensitive enough to be able to locate the leaking pipe consistently and the difference in the information provided by optimally placed pressure sensors and sensors placed in the worst locations is not significant. Second, using a relatively small number of optimally placed flow-rate sensors is enough to be able to locate consistently and accurately the location and magnitude of leakage. This kind of sensor also exhibited significant difference in the information gained through optimally placed sensors and sensors placed in the worst locations, requiring significantly more sensors placed in the worst locations to obtain the same amount of information, giving more incentive in investing into finding the optimal sensor locations in the first place. Third, by introducing more uncertainties into the system the minimum number of required sensors increases with the modelling error being mainly responsible for this.

The proposed OSP methodology can be implemented into a smart water distribution network capable of online monitoring the health of the system. Finding the optimal locations to place the sensors can potentially reduce the cost of the monitoring system by using a smaller number of sensors, while simultaneously being a strong asset in the decision making of the operation by providing very informative measurements capable of locating leaks early.

Potential future developments in this study are first, to test the methodology in a more complex realistic network. Second to introduce uncertainty to more model parameters and study the effects of these uncertainties, as well as, explore the effect of using different magnitudes of uncertainty. Third, explore the effects of introducing more than 1 leakage simultaneously into the network and lastly, use a hydraulic model of the system that is dynamic and can adapt into the more realistic time varying model parameters.

9. References

[1] O. Hunaidi, A. Wang, M. Bracken, T. Gambino, and C. Fricke, “Acoustic methods for locating leaks in municipal water pipe networks,” 2004.

[2] M. Nakhkash and M. R. Mahmood-Zadeh, “Water leak detection using ground penetrating radar,” 2004.

[3] A. F. Colombo, P. Lee, and B. W. Karney, “A selective literature review of transient-based leak detection methods,” *Journal of Hydro-Environment Research*. 2009, doi: 10.1016/j.jher.2009.02.003.

[4] J. A. Goulet, S. Coutu, and I. F. C. Smith, “Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks,” *Adv. Eng. Informatics*, 2013, doi: 10.1016/j.aei.2013.01.001.

[5] Z. Poulakis, D. Valougeorgis, and C. Papadimitriou, “Leakage detection in water pipe networks using a Bayesian probabilistic framework,” *Probabilistic Eng. Mech.*, 2003, doi: 10.1016/S0266-8920(03)00045-6.

[6] H. A. Jensen and D. J. Jerez, “A Bayesian model updating approach for detection-related problems in water distribution networks,” *Reliab. Eng. Syst. Saf.*, 2019, doi: 10.1016/j.ress.2018.12.014.

[7] M. C. Potter, D. C. Wiggert, and B. H. Ramadan, *Mechanics of Fluids*, 4th ed. 2010.

[8] Ζ. Πουλάκης, “Σχεδιασμός, Βελτιστοποίηση και Διάγνωση Βλαβών σε Δίκτυα Σωληνώσεων,” University of Thessaly, 2000.

[9] L. A. Rossman, “Epanet 2 User ’s Manual,” *National Risk Management Research Laboratory Office of Research and Development. U.S. Environmental Protection Agency Cincinnati*, 2000.

[10] S. Martínez and F. Bullo, “Optimal sensor placement and motion coordination for target tracking,” *Automatica*, 2006, doi: 10.1016/j.automatica.2005.12.018.

[11] C. Papadimitriou, “Optimal sensor placement methodology for parametric identification of structural systems,” *J. Sound Vib.*, 2004, doi: 10.1016/j.jsv.2003.10.063.

[12] C. Papadimitriou, J. L. Beck, and S. K. Au, “Entropy-based optimal sensor location for structural model updating,” *JVC/Journal Vib. Control*, 2000, doi: 10.1177/107754630000600508.

[13] N. S. Cheng, “Formulas for friction factor in transitional regimes,” *J. Hydraul. Eng.*, 2008, doi: 10.1061/(ASCE)0733-9429(2008)134:9(1357).

[14] R. D. Blevins, “Applied fluid dynamics handbook.,” 1984, doi: 10.1016/0894-1777(94)90012-4.

[15] N. Vasileiadis, G. Tatsios, S. Misdanitis, and D. Valougeorgis, “Modeling of complex gas distribution systems operating under any vacuum conditions: Simulations of the ITER divertor pumping system,” *Fusion Eng. Des.*, 2016, doi: 10.1016/j.fusengdes.2015.12.033.

[16] N. Deo, “GRAPH THEORY WITH APPLICATIONS TO ENGINEERING AND COMPUTER SCIENCE,” *Networks*, 1975, doi: 10.1002/net.1975.5.3.299.

[17] E. Simoen, C. Papadimitriou, and G. Lombaert, “On prediction error correlation in Bayesian

model updating," *J. Sound Vib.*, 2013, doi: 10.1016/j.jsv.2013.03.019.

[18] D. I. Papadimitriou and C. Papadimitriou, "Optimal sensor placement for the estimation of turbulence model parameters in CFD," *Int. J. Uncertain. Quantif.*, 2015, doi: 10.1615/Int.J.UncertaintyQuantification.2015015239.

[19] J. Ching and Y. C. Chen, "Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging," *J. Eng. Mech.*, 2007, doi: 10.1061/(ASCE)0733-9399(2007)133:7(816).

[20] S. Wu, P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, "Bayesian annealed sequential importance sampling: An unbiased version of transitional Markov chain Monte Carlo," *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.*, 2018, doi: 10.1115/1.4037450.

[21] E. T. Jaynes, "Where Do We Stand on Maximum Entropy?," *The Maximum Entropy Formalism*. 1978, doi: 10.1007/BF01008275.

[22] C. Argyris, "Bayesian Uncertainty Quantification and Optimal Experimental Design in Data-Driven Simulations of Engineering Systems," University of Thessaly, 2017.

[23] X. Huan and Y. M. Marzouk, "Gradient-based stochastic optimization methods in Bayesian experimental design," *Int. J. Uncertain. Quantif.*, 2014, doi: 10.1615/Int.J.UncertaintyQuantification.2014006730.