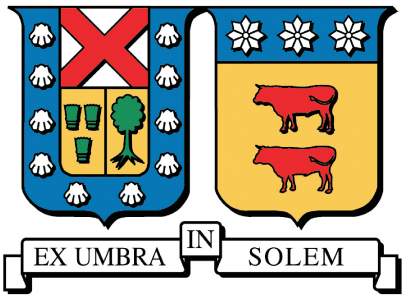
2019

A PROBABILISTIC FRAMEWORK FOR HYDRAULIC PERFORMANCE ASSESSMENT OF COMPLEX WATER DISTRIBUTION NETWORKS

JEREZ URQUIETA, DANKO JOSÉ



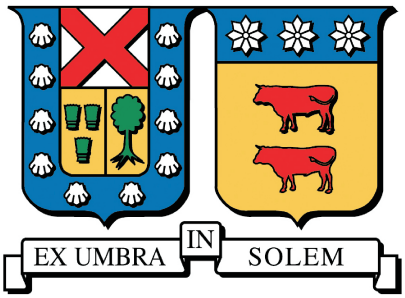
FEDERICO SANTA MARÍA TECHNICAL UNIVERSITY
CIVIL ENGINEERING DEPARTMENT

A PROBABILISTIC FRAMEWORK FOR HYDRAULIC PERFORMANCE ASSESSMENT OF COMPLEX WATER DISTRIBUTION NETWORKS

DANKO JOSÉ JEREZ URQUIETA

Civil Engineer
Master of Science in Civil Engineering

March, 2019



FEDERICO SANTA MARÍA TECHNICAL UNIVERSITY
CIVIL ENGINEERING DEPARTMENT

A PROBABILISTIC FRAMEWORK FOR HYDRAULIC PERFORMANCE ASSESSMENT OF COMPLEX WATER DISTRIBUTION NETWORKS

A thesis submitted by
DANKO JOSÉ JEREZ URQUIETA

in fulfillment of the requirements for the degree of
Civil Engineer

and the degree of
Master of Science in Civil Engineering

Advisor
Dr. Hector Jensen Velasco

March, 2019

A PROBABILISTIC FRAMEWORK FOR HYDRAULIC PERFORMANCE ASSESSMENT OF COMPLEX WATER DISTRIBUTION NETWORKS

AUTHOR

DANKO JOSÉ JEREZ URQUIETA

A thesis submitted to the Federico Santa María Technical University in fulfillment of the requirements for the degree of Civil Engineer and the degree of Master of Science in Civil Engineering.

EVALUATION COMMITTEE

Dr. Héctor A. Jensen _____

Dr. Marcos A. Valdebenito _____

Dr. André T. Beck _____

Valparaíso, Chile, March 2019.

Memento mori

ACKNOWLEDGMENTS

I would first like to thank my thesis advisor Dr. Hector Jensen. His guidance, patience, motivation and knowledge were fundamental during the learning process of this thesis. I want to also express my gratitude for his permanent support during my studies, and the trust placed in me to develop this work.

I would also like to thank Dr. Marcos Valdebenito for his continuous help during my studies, the insightful comments during the development of this thesis, and for the opportunity to participate as his teaching assistant all these years.

This work would have been impossible without the advice of Mauricio Correa. His invaluable experience and help were fundamental for the development of the study cases.

The unfailing support and continuous encouragement of my parents, Abraham and Paola, and my beloved partner Jocelyn, were essential to complete my studies. Special thanks also to my brother Viktor and my sister Millaray, for the nice moments we have shared together during our entire life.

I would like to express the deepest appreciation to my grandmother Sonia, my grandfather Alberto, and my great-uncle Pepe. Every single step in my life finds its foundations on their life principles, the good moments we shared, and the profound sense of responsibility they taught me.

Finally, my graduate studies were partially funded by CONICYT (National Comission for Scientific and Technological Research) under CONICYT-PFCHA/MagísterNacional-22171881 program, and my undergraduate studies were fully funded by ESO (European Southern Observatory) through the ESO-Paranal program. I want to gratefully acknowledge these supports.

ABSTRACT

A probabilistic framework for evaluating the hydraulic performance of complex water distribution networks is formulated and implemented in this work. Two classes of problems are addressed: hydraulic reliability assessment, and detection-related problems. On one hand, a stochastic approach is developed and implemented in order to handle the hydraulic reliability assessment of large-scale water distribution systems. Hydraulic reliability is a probabilistic measure of the hydraulic performance of a water distribution system, involving high-dimensional problems from the probabilistic point of view. The effectiveness of the approach is evaluated with the study of a real-life water distribution network consisting of thousands of nodes and pipes. A number of analyses within the proposed framework are demonstrated, including reliability and sensitivity analysis, uncertainty propagation and failure analysis. On the other hand, a simulation-based Bayesian model updating methodology is developed to deal with detection-related problems in complex water utility networks, such as leakage detection. Detection-related problems are expressed as model class selection problems, where a number of model classes are defined as potential characterizations of the actual network condition. Revised information based on available data is then used to select the most probable model class, which is identified as the actual network condition. Two application examples are studied in order to evaluate the capabilities of the proposed approach. These examples correspond to leakage detection and connectivity detection problems, involving real-life water distribution networks and synthetic measurement data. Several factors affecting the effectiveness of the approach are evaluated, including modeling uncertainties, measurement errors and sensors configuration. In order to set and formulate the proposed approaches, advanced simulation techniques are properly adapted and integrated with a commercial level hydraulic simulator. Results demonstrate that the framework provides functional tools that give a valuable insight into the hydraulic performance of a class of complex utility networks.

RESUMEN

En este trabajo se formula e implementa un marco probabilístico para la evaluación del desempeño hidráulico de redes complejas de distribución de agua. Se consideran dos tipos de problemas en este contexto: evaluación de la confiabilidad hidráulica, y problemas relacionados a detección. Por un lado, se desarrolla e implementa un enfoque estocástico para llevar a cabo la evaluación de confiabilidad hidráulica de sistemas complejos de distribución de agua. La confiabilidad hidráulica es una medida probabilística del nivel de desempeño hidráulico de un sistema de distribución de agua, que involucra problemas altamente dimensionales desde el punto de vista probabilístico. La efectividad del método propuesto es evaluado mediante el estudio de una red de distribución real con miles de nodos y tuberías. Se demuestran varios tipos de análisis posibles en el marco propuesto, incluyendo propagación de incertidumbre, análisis de confiabilidad y sensibilidad, y análisis de falla. Por otro lado, se desarrolla una metodología de actualización Bayesiana de modelos basada en simulación para enfrentar problemas relacionados a detección, tales como el localizar fugas. Este tipo de problemas se formula como un problema de selección de clase de modelo probabilístico, en donde un cierto número de clases de modelos se definen como potenciales caracterizaciones de la condición real de la red. La clase más probable se determina en base a las mediciones disponibles y el conocimiento previo sobre el sistema. Dicha clase se interpreta como la condición real de la red. Con el fin de demostrar la aplicabilidad del enfoque propuesto, se formulan y estudian dos ejemplos de aplicación. Estos corresponden a los problemas de detección de fugas y detección de conectividad, involucrando redes reales de distribución de agua y mediciones sintéticas. Se estudia la incidencia de varios factores en la efectividad de la metodología propuesta, incluyendo errores de medición y modelación, así como la configuración de los dispositivos de monitoreo. Con el fin de formular e implementar las metodologías propuestas, dos técnicas de simulación estocástica se han adaptado apropiadamente e integrado con un software hidráulico de nivel comercial. Los resultados obtenidos demuestran que el marco probabilístico desarrollado en este trabajo provee herramientas eficientes y efectivas que permiten un entendimiento valioso del desempeño hidráulico de redes complejas de distribución de agua.

GLOSSARY

1. **Water distribution network:** Engineered system composed of a collection of sources, pipes and hydraulic control elements (e.g., pumps, valves, regulators, tanks) delivering to consumers prescribed water quantities at desired pressures and water qualities.

2. **Advanced simulation techniques:** Simulation algorithms developed to obtain probabilistic information of a system based on an efficient sampling scheme, in order to reduce the computational effort. These methods decouple the sampling process from the physics-based modeling of the system under consideration.

3. **Probability of failure:** Probabilistic measure of the plausibility that a given system is not operating as expected. The probability of failure is related to the reliability of the system through the following equality: Reliability $= 1 -$ Probability of failure.

4. **Reliability sensitivity:** Variation of the reliability of a system when certain parameters defining it are disturbed. In a local approach, it involves the computation of partial derivatives of the reliability with respect to the parameters of interest.

5. **Bayesian model updating:** General framework to determine the most plausible values for the parameters of a given mathematical model, based on available measurement data.

6. **Probabilistic model class:** Mathematical object defined in the Bayesian model updating framework for formulating the problem in a probabilistic setting. It comprises physical assumptions about the system, parametrization of computational models, prior knowledge, and a probabilistic characterization of the system response.

7. **Computational cost:** Time required to complete a certain process in a computer.

8. **Post-processing:** Use of the samples already generated by a simulation procedure to obtain new information about a certain system, without the need for solving any additional computational model.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Modern industry and society rely on the correct performance of several critical infrastructure systems, such as water distribution networks. These increasingly interwoven, large-scale and complex hydraulic systems are pivotal engineering assets, dispensing one of the most essential resources for human life: water. In this context, water utility networks work as a backbone of industrial societies, and their performance significantly affects diverse human behaviour with critical influence [1]. Thus, the development of efficient, reliable and robust water distribution networks is an essential task to ensure public welfare and the progress of modern society [2].

Current real-world systems and their environment are characterized by a rapid growth in complexity and scale. Moreover, water distribution networks constantly operate under uncertain conditions (e.g. unknown demands) and are prone to uncontrolled external events (e.g. natural hazards). Then, water utility managers are faced to difficult decision-making processes regarding the repair, maintenance and enhancement of such hydraulic systems. These decisions under uncertain conditions have a significant economical and societal impact. In this regard, there is a current need for developing suitable numerical tools to quantitatively assess the hydraulic performance of real-life water distribution systems, taking explicitly into account all the involved uncertainties.

The novel aspect of this thesis is the development and implementation of a framework based on probability theory that allows the treatment of two relevant classes of problems in the context of complex water distribution networks: hydraulic reliability assessment,

and detection-related problems. In the first part of this work, a stochastic approach is formulated for hydraulic reliability assessment, which is a probabilistic measure of the hydraulic performance of hydraulic systems. The outstanding feature of this approach, in comparison with previous contributions, is the ability to handle high-dimensional reliability problems arising in real-life water utility systems. In the second part of this document, a simulation-based Bayesian model updating approach is proposed for detection-related problems directly related to the hydraulic performance of water utility systems. The approach can handle a variety of detection-related problems in complex water distribution networks, including identifiable and strictly unidentifiable systems. Both methodologies involve the use of advanced simulation techniques [3,4], which are properly adapted as tools to set and formulate the required analyses. The capabilities and generality of the proposed framework, as well as the type of information that can be obtained, are demonstrated with three application examples involving real-life water distribution networks.

1.1 Thesis structure

This thesis consists of eight chapters, structured as follows.

- **Chapter 1:** Description of the general problem under consideration and its relevance. General and specific objectives are also formulated here.
- **Chapter 2:** Evaluation of the state of the art regarding reliability analysis and detection-related problems in water distribution networks.
- **Chapter 3:** Formulation of the stochastic approach proposed for hydraulic reliability analysis of complex water utility systems.
- **Chapter 4:** Evaluation and discussion of the results obtained in the hydraulic reliability analysis of a real-life water distribution network.
- **Chapter 5:** Formulation of the simulation-based Bayesian model updating approach proposed for detection-related problems in complex water utility systems.
- **Chapters 6 and 7:** Evaluation and discussion of the results obtained in two application examples (leakage detection and connectivity detection, respectively).
- **Chapter 8:** Discussion of the main conclusions and future research efforts.

1.2 General objectives

This work develops a probabilistic framework to handle two problems associated with the hydraulic performance assessment of complex water distribution networks. Thus, the following two general objectives are formulated:

1. To formulate, implement and evaluate a stochastic simulation approach for hydraulic reliability assessment of complex water distribution networks.

2. To formulate, implement and evaluate a simulation-based Bayesian model updating approach for detection-related problems in complex water distribution networks.

1.3 Specific objectives

The following specific objectives have been formulated to achieve the general objectives:

1. To review and evaluate the state of the art regarding reliability analysis and detection-related problems in water distribution networks.

2. To formulate hydraulic reliability analysis and detection-related problems (emphasizing leakage location) in the context of probability theory, adapting advanced simulation techniques to this end.

3. To implement advanced simulation techniques adapted to the context of water distribution networks.

4. To model real-life water distribution networks in a commercial-level software, in order to evaluate the capabilities of the proposed framework.

5. To evaluate the strengths, limitations and further capabilities of the proposed methodologies, based on the results obtained from application examples involving real-life water distribution networks.

Chapter 2

STATE OF THE ART

Two important types of problems associated to the hydraulic performance of real-life water distribution networks are addressed in this work: (1) hydraulic reliability analysis, and (2) detection-related problems, with emphasis on leakage location. This chapter discusses the current state of the art in these subjects, identifying current challenges and introducing the main ideas to develop the proposed probabilistic framework.

2.1 Reliability assessment of water distribution networks

Water is an essential resource for public welfare and industrial activities. Thus, water distribution networks must be highly reliable in their operations [5–8]. As already pointed out, one of the intrinsic characteristics of these systems is that they are subjected to uncertainty, that is, the magnitude of a number of parameters describing the system is not completely known in practice. Examples of this situation are chemical characteristics of inflow water, deterioration level of network components, users requirements, etc. Uncertainties arising in this context can be modeled by means of, for example, fuzzy sets [9] or probability theory [10]. The last approach is the one considered in this work, where the uncertainty in the system parameters is characterized by a joint probability density function. Then, the degree to which a water utility network is able to provide some desired level of service can be quantitatively assessed during its design and operation. In this regard, reliability is a suitable probabilistic measure to assess the performance of an engineering system when uncertain conditions are taken into account. Water distribution systems are

concerned with two classes of reliability measures: mechanical or structural reliability, and operational reliability.

Mechanical reliability is related to the ability of the distribution system to provide its function under some mechanical or structural failure, such as pipe breakage, pump failure, power outage and equipment failure in general. Thus, this type of reliability is associated with the capacity of the network to continue operating without the need for repairs, modifications or replacement of components and subcomponents [11]. Network reliability analyses based on mechanical, structural or operational failure are well established and have been developed for many fields, such as mechanical, electrical and chemical engineering systems. The reader is referred to the following representative works for further information [12–16].

Operational reliability quantifies the ability of the system to continuously provide a prescribed level of service under normal operational conditions. This work is concerned with the hydraulic reliability assessment of complex water distribution networks, a class of operational reliability that is directly related to the hydraulic performance of the system. Specifically, hydraulic reliability is associated to the ocurrence of hydraulic failure, which takes place when users receive insufficient delivered flow or inadequate pressure head [11]. Such behavior may be due to several reasons, such as deterioration of network hydraulic capacity, insufficient water inflow, changes in consumer demands or pressure head requirements, inadequate maintenance and operation of control elements, etc.

The literature presents several studies developed to assess the hydraulic reliability of water distribution networks. Traditional simulation approaches, such as Monte Carlo and Latin Hypercube sampling techniques, have been reported in the past to perform reliability and uncertainty propagation analyses [11,17,18]. The combination of surrogate models with Monte Carlo simulation has also been used to address the reliability analysis of these systems [19]. Lately, first-order reliability methods in connection with response surface methodologies have been implemented to address the capacity reliability problem [20, 21]. Additional developments in this context can be found in [15, 22–24]. The previous approaches target small scale water distribution networks, proving to be quite useful in those systems. Nonetheless, such small scale models are unable to represent real-life water distribution systems, which are characterized by thousands or hundreds of thousands of network elements.

Real-life water distribution networks are intrinsically defined by their very large size [8].

This leads to hydraulic models involving hundreds or thousands of nodes and links, highly intricate topology, and non-trivial interactions between model parameters. As a result, a high number of uncertain parameters (of the order of thousands or hundreds of thousands) needs to be considered for an appropriate characterization of the engineering system. Thus, the study of the reliability, sensitivity and robustness of these highly dimensional systems becomes extremely complex, where traditional methodologies such as direct integration or first/second order reliability methods are not appropriate. On the other hand, these systems must be highly reliable, that is, small failure probabilities need to be assessed (below 10^{-3}). Although standard Monte Carlo simulation techniques can be theoretically applied, they are unfeasible due to the very large number of hydraulic simulations required to carry out the failure probability estimation. From the previous challenges, it becomes apparent that the development of suitable methods and techniques for the reliability assessment of real-life water distribution systems is needed.

This work proposes an efficient and effective stochastic approach for assessing the hydraulic reliability and sensitivity of complex water distribution networks. The formulation of the approach is based on the following conceptual aspects. First, complex utility networks can be understood as structural skeletons of complex dynamical systems. Thus, advanced simulation techniques [3, 25] can be adapted as a tool to address hydraulic reliability problems. Second, a number of accurate and efficient computational tools exists for performing physics-based analyses of water utility networks [26], which can be integrated with advanced simulation techniques. Finally, current advances of parallelization techniques allow to implement stochastic simulation methods and hydraulic analyses in a parallel environment, increasing the computational efficiency for real-life applications. Figure 2.1 summarises both the challenges of the hydraulic reliability analysis of complex water distribution networks and the conceptual features that lead to the implementation of the proposed stochastic framework.

2.2 Detection-related problems in water utility networks

Water utility networks are constantly exposed to unexpected events that can critically affect their hydraulic performance and, therefore, public welfare. Examples of such problems include background and burst-related leakage, equipment failure, unauthorized demand, incorrect connectivity, sabotage, pipe breakage, unacceptable water quality, pump failure, etc. [13, 15, 27–29] A crucial task is to characterize, based on available data about

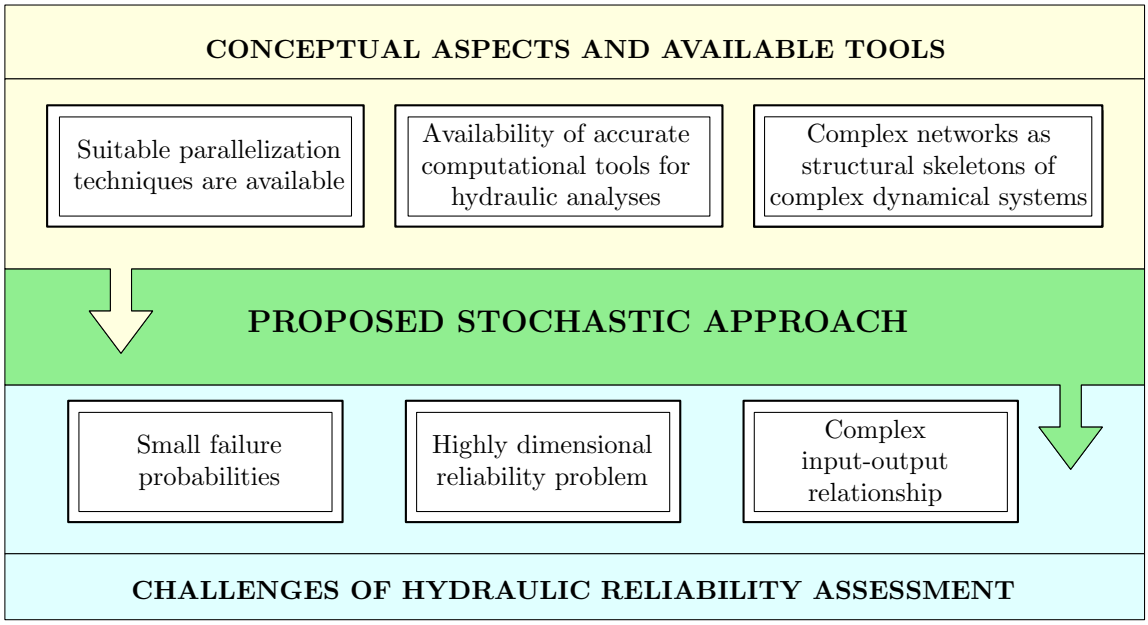


Figure 2.1: Summary of the approach proposed for hydraulic reliability analysis

the behavior of the system, the specific event affecting the performance of the network. In this work, such problems are called *detection-related problems*, being leakage detection one of the major concerns in this context.

Leakage involves water escaping from the pipeline system at a certain location. A number of factors can lead to the development of this phenomenon in a water distribution system, such as excessive loads, pipe ageing, poor quality workmanship, insufficient maintenance, soil liquefaction, etc. Leaks usually result in service disruption and huge losses for water utility companies [30]. In fact, the average losses in these systems range between 15% and 35% of their hydraulic inflow [31]. The UK Environment Agency states that sufficient water to meet the needs of 20 million people is lost through leakage every day (3 billion liter/day), which shows the global impact of this problem. On the other hand, the presence of leakage can affect public health due to infiltration of contaminants into the pipeline system. Moreover, water is an increasingly scarce resource that should not be wasted. As a consequence, the development of effective techniques and methodologies for leakage detection is an important task for modern society, considering practical, economical, environmental and health issues.

The leakage management process involves three main stages: assessment, detection and control. Leakage assessment is related to the quantification of total water losses in a cer-

tain network, such as top-down and bottom-up assessment [30,32], while leakage control is concerned with strategies, policies and operations directed to reduce leakages [33]. Leakage detection methods, on the other hand, address the issue of locating and characterizing the leak. The most accurate methodologies in this context are pin-pointing methods, such as leak noise correlators, gas injection, and pig-mounted acoustic sensing [34–36]. However, their application is time-demanding and expensive. Then, it is necessary to first identify and prioritize areas of leakage, a task usually known as *leakage location*.

The most well-known technique regarding leakage location is step-testing [37], although it cannot address real-time analyses and undesired phenomena can occur during its application, such as backsiphonage or infiltration of ground water into the network. Some sensor-driven approaches have been developed in order to overcome these issues, such as acoustic logging, ground motion sensors and ground penetrating radar [38, 39]. These methodologies have been developed and tested with different levels of effectiveness and applicability. Another class of techniques is based on the similarity between measurements and hydraulic models, where the goal is to minimize the differences between predicted hydraulic responses from computational models and field measurements. Representative works in this direction are [40–43]. Some of these methodologies include explicit representations of the unavoidable uncertainties arising in the modeling and monitoring processes of water distribution systems.

A different group of probability-based techniques that have been used in the context of leakage detection are Bayesian system identification methodologies [44–47], which have also been reported in the context of contaminant source detection [48] and network model calibration [49]. These methodologies are based on the Bayesian interpretation of probability theory, where probability is defined as a measure of the plausibility of a given hypothesis [50]. In this way, such methodologies allow to deal with usual situations in real-life water distribution networks where a large amount of data is not available. Nonetheless, it is believed that there is still room for further developments in model-based techniques with applications to involved water distribution systems.

This work presents a general Bayesian model updating framework for solving detection-related problems in complex water distribution networks. The approach is based on the following aspects. First, Bayesian probabilities provide a rational representation of uncertainties that are inherent to detection processes, e.g. those coming from modeling and measurement errors [51–53]. Hence, the same theoretical framework can be used to handle

detection-related problems in complex water distribution networks. Second, advanced simulation techniques for Bayesian model updating are well developed and tested [4, 54, 55]. These techniques allow to handle a variety of problems including identifiable and strictly unidentifiable systems, providing valuable information about the interaction of different model parameters. Thus, based on the interpretation of utility networks as structural skeletons of complex dynamical systems, simulation methods can be suitably integrated as general tools for solving detection-related problems. In particular, a multi-level Markov chain Monte Carlo algorithm is adopted here [4, 54]. Third, as previously pointed out, available hydraulic simulators can be integrated with advanced simulation techniques to perform the required physics-based analyses. Finally, different features of the proposed framework allow the implementation in a parallel environment to enhance its efficiency. Figure 2.2 summarises both the challenges of detection-related problems in complex water distribution networks and the conceptual aspects that allow the implementation of the proposed Bayesian model updating framework to address them.

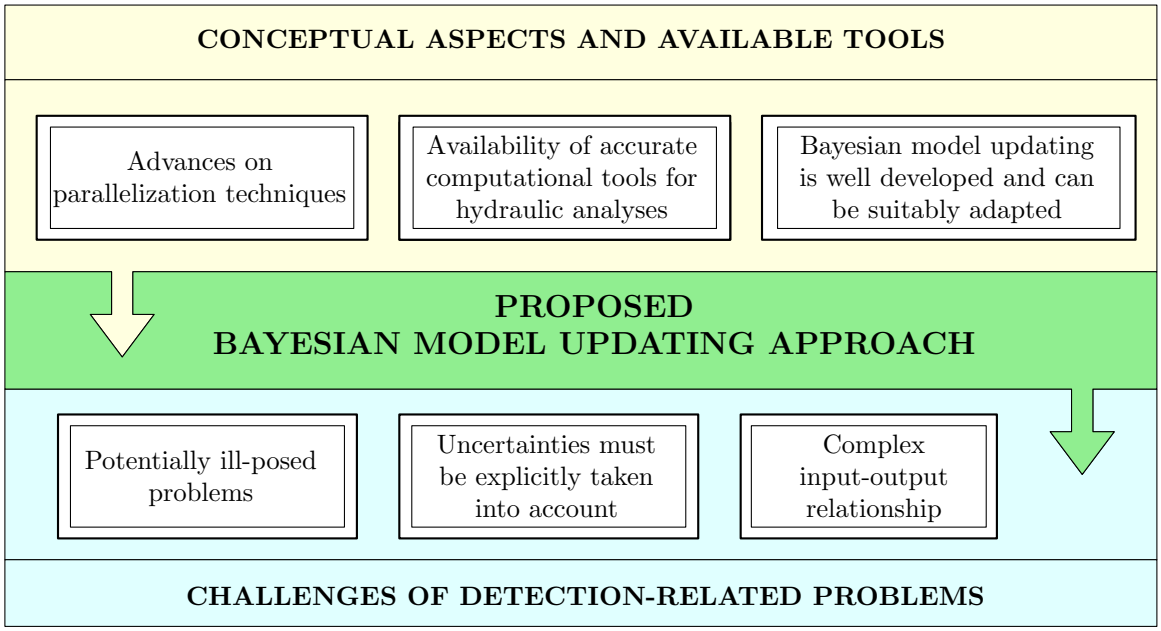


Figure 2.2: Summary of the approach proposed for detection-related problems

Chapter 3

STOCHASTIC APPROACH FOR RELIABILITY ASSESSMENT

The hydraulic performance of water distribution networks operating under uncertain conditions must be quantitatively assessed during their design and operation. However, the very large size and complexity of these systems make the required reliability analysis a very challenging problem. This chapter formulates the hydraulic reliability problem in a probabilistic context and introduces a stochastic approach to carry out the required calculations. The approach can handle complex water distribution networks involving thousands of nodes and links. Subset simulation is used as a tool for carrying out the reliability assessment, allowing several additional analyses that include reliability sensitivity and failure analysis. The approach provides a general means to study and evaluate the hydraulic performance of water distribution networks operating under uncertain conditions.

3.1 Network hydraulic reliability problem

Consider a water distribution network consisting of a large number of elements, such as pipes, nodes, pumps, valves, regulation tanks, etc. The corresponding hydraulic model is parametrized by n_θ uncertain parameters comprised in the *network state vector* $\boldsymbol{\theta} \in \Omega_\theta \subset \mathbb{R}^{n_\theta}$. These parameters represent features of the system affecting its hydraulic performance. Examples of such parameters include roughness coefficients of pipes, nodal demands, prescribed nodal heads, pipe diameters, etc.

The network state vector $\boldsymbol{\theta}$ is defined in a probabilistic manner by means of a joint probability density function $p(\boldsymbol{\theta})$. This function is defined explicitly through a number of distribution parameters $\boldsymbol{\tau}$, that is, $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}; \boldsymbol{\tau})$. The distribution parameters $\boldsymbol{\tau}$ can be, for example, the first two statistical moments, upper and lower bounds, or other representative values of $\boldsymbol{\theta}$. The usual case is when the mean values represent nominal values of the parameters and the standard deviation specifies a degree of variability in the network parameters, based on available data or engineering judgment.

The uncertainty in the network state parameters $\boldsymbol{\theta}$ will affect the behavior of the water distribution system. In this context, a utility function $\mu(\boldsymbol{\theta})$ is introduced to quantitatively assess the performance level of such system, where higher values of $\mu(\boldsymbol{\theta})$ correspond to better system performance. This function is defined, in general, by means of a complex input-output relation between the network state parameters $\boldsymbol{\theta}$ and some quantity of interest. The input-output relation can be provided by available hydraulic simulation packages, such as WaterCAD, InfoWater or EPANET [26]. Some examples of quantities of interest are delivered flow and pressure heads. In this setting, a *system failure* is defined when the utility function is *below* some critical performance level. Thus, the failure domain $F \subset \Omega_\theta \subset \mathbb{R}^{n_\theta}$ is defined as

$$F = \{\boldsymbol{\theta} \in \Omega_\theta : \mu(\boldsymbol{\theta}) < \mu^*\} \tag{3.1.1}$$

where μ^* is the critical threshold. For instance, if the utility function corresponds to total delivered flow in the network, then μ^* represents the minimum delivered flow required in the network, and the corresponding failure domain F is the set of all network state parameters that lead to unacceptable delivered flow in the water distribution system (unacceptable level of service).

The hydraulic reliability of the water distribution system (R_F) is defined as the probability that the network provides an acceptable performance level for a given configuration. The corresponding network hydraulic reliability problem is formulated in terms of the probability of failure P_F, since $R_F = 1 - P_F$. The failure probability is defined by a multi-dimensional probability integral over the failure domain F as

$$P_F = \int_F p(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{3.1.2}$$

The above expression can also be written in terms of the indicator function $I_F(\boldsymbol{\theta})$, where $I_F(\boldsymbol{\theta}) = 1$ if $\boldsymbol{\theta} \in F$ and $I_F(\boldsymbol{\theta}) = 0$ otherwise. Then,

$$P_F = \int_{\Omega_\theta} I_F(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{3.1.3}$$

Equivalently, the same integral can be written as

$$P_F = E_{p(\boldsymbol{\theta})}[I_F(\boldsymbol{\theta})] \tag{3.1.4}$$

where $E_{p(\boldsymbol{\theta})}[\,\cdot\,]$ is the expectation operator with respect to the distribution $p(\boldsymbol{\theta})$. Note that the expressions in equations (3.1.2), (3.1.3) and (3.1.4) are similar to the ones involved in reliability estimation of dynamical systems [3].

It is remarked that the formulation presented in this chapter does not include mechanical or structural failure of system components. The integration of the proposed framework with these additional reliability considerations, which can provide a measure of overall system performance, is left for future research efforts.

3.2 Reliability estimation

As introduced in Section 2.1, the computation of the hydraulic reliability of real-life water distribution networks is a very challenging problem. First, due to the high complexity of this type of hydraulic systems, a very large number of uncertain parameters is required to properly characterize them (n_θ of the order of thousands or hundreds of thousands). Therefore, the probability integral is high-dimensional. Second, real-life water distribution networks involve very complex input-output relations $\boldsymbol{\theta} \rightarrow \mu(\boldsymbol{\theta})$ to assess their performance and, hence, the computational effort required to solve a single hydraulic model is significant. Finally, current water distribution systems must be highly reliable, which means that failure is a rare event; in other words, the probability of failure is small ($P_F < 10^{-3}$).

Based on the previously discussed issues, it becomes clear that traditional techniques based on Equation (3.1.2), such as numerical integration or standard first and second-order reliability methods, are not suitable to handle the hydraulic reliability problem. This favours the use of simulation techniques based on Equation (3.1.4). However, since

failure is a rare event (P_F is small), the use of traditional simulation techniques such as Monte Carlo simulation is inefficient. Therefore, considering these issues and understanding that water distribution networks are structural skeletons of complex dynamical systems, advanced simulation techniques arise as a natural tool to handle the hydraulic reliability problem.

3.2.1 Advanced simulation techniques

Advanced simulation techniques are numerical tools specially developed to deal with challenging reliability problems. The main idea of these algorithms is to formulate specially designed sampling schemes to reduce the number of model realizations, while maintaining a good quality of the failure probability estimator. Some examples of advanced simulation techniques are line sampling [56, 57], importance sampling [58], auxiliary domain method [59], horseracing simulation [60], subset simulation [3, 8, 25] and subset simulation based on hidden variables [61].

The approach proposed in this chapter considers the implementation of subset simulation as a tool to handle the hydraulic reliability problem. Several works have shown that this method can be applied efficiently to a wide range of dynamical systems [5, 25, 62, 63]. The simulation method does not make any assumption about the topology of the failure domain, hence its performance is not affected by the dimensionality of the reliability problem. Subset simulation is selected for handling the hydraulic reliability problem because it provides a good balance between efficiency and robustness. Additionally, simulated samples can be directly used to perform failure analysis [25] and reliability sensitivity can be estimated as a post-process of the simulation procedure [64,65]. Appendix A provides a detailed description of the method, including a pseudo-code with the actual implementation. For completeness, the main ideas of the algorithm are discussed here.

3.2.2 Basic ideas of subset simulation

Subset simulation is based on understanding the failure domain F as the intersection of a sequence of m nested intermediate failure domains, that is, $F = F_m \subset F_{m-1} \subset \ldots \subset F_1$, so that $F = \bigcap_{\kappa=1}^{m} F_\kappa$. Then, the definition of conditional probability gives

$$P_F = P(F_m) = P(F_1) \times \prod_{\kappa=1}^{m} P(F_\kappa | F_{\kappa-1}) \tag{3.2.1}$$

In this setting, it is required to compute the unconditional probability $P(F_1)$ and the

conditional probabilities $P(F_\kappa|F_{\kappa-1}), \kappa = 1, \ldots, m$. Observe that, although P_F can be small, the previous quantities can be efficiently estimated by direct simulation under an appropriate choice of m and F_κ. The intermediate failure domains are defined as

$$F_\kappa = \{\boldsymbol{\theta} \in \Omega_\theta : \mu(\boldsymbol{\theta}) < \mu_\kappa^*\}, \quad \kappa = 1, \ldots, m \tag{3.2.2}$$

where $\mu_1^* > \cdots > \mu_{m-1}^* > \mu_m^* = \mu^*$ are the intermediate thresholds. The unconditional samples at the first stage ($\kappa = 1$) are generated by direct Monte Carlo simulation, whereas the modified Metropolis-Hasting algorithm [3, 66, 67] is used to generate the conditional samples $\{\boldsymbol{\theta}_{\kappa-1,i} : i = 1, \ldots, N_\kappa\}, \kappa = 2, \ldots, m$. In the actual implementation, a constant sample size N_s is considered for all stages, that is $N_1 = \ldots = N_m = N_s$.

3.2.3 Failure probability estimation

In the actual implementation of subset simulation, the intermediate thresholds are adaptively selected from the simulated samples in such a way that the probabilities in Equation (3.2.1) are equal to a prescribed value p_0, which is called the *conditional failure probability*. In other words, the algorithm estimates the intermediate thresholds $\mu_1, \ldots, \mu_{m-1}$ for pre-specified failure probability levels $p_0, p_0^2, \ldots, p_0^{m-1}$, respectively. Therefore, all failure probabilities in Equation (3.2.1) are equal to p_0, except $P(F_m|F_{m-1})$. The failure probability is then computed as

$$P_F = p_0^{m-1} \int_{\boldsymbol{\theta} \in \Omega_\theta} I_F(\boldsymbol{\theta}) p(\boldsymbol{\theta}|F_{m-1}) d\boldsymbol{\theta} \tag{3.2.3}$$

or, equivalently,

$$P_F = p_0^{m-1} E_{p(\boldsymbol{\theta}|F_{m-1})} [I_F(\boldsymbol{\theta})] \tag{3.2.4}$$

where $E_{p(\boldsymbol{\theta}|F_{m-1})}[\,\cdot\,]$ is the expectation operator with respect to the conditional distribution $p(\boldsymbol{\theta}|F_{m-1})$. The expectation $E_{p(\boldsymbol{\theta}|F_{m-1})}$ can be estimated with the set of samples conditional to F_{m-1}, that is $\{\boldsymbol{\theta}_{m-1,i} : i = 1, \ldots, N_m\}$. Therefore

$$P_F \approx \hat{P}_F = p_0^{m-1} \frac{1}{N_m} \sum_{i=1}^{N_m} I_F(\boldsymbol{\theta}_{m-1,i}) \tag{3.2.5}$$

Note that subset simulation is actually a method to estimate the critical thresholds associated to pre-established values of failure probability. In other words, the algorithm provides a full characterization of the hydraulic reliability trend for different levels of acceptable performance without the need to explicitly define a target failure event. This

feature allows, for example, to gain an insight into the hydraulic performance of the system when different network configurations are taken into account, as demonstrated in Chapter 4.

3.2.4 Implementation issues

The highest computational cost of the simulation procedure comes from sampling the indicator function $I_F(\boldsymbol{\theta})$, which involves solving a hydraulic model of the water distribution network. The number of hydraulic simulation runs required during the analysis depends on the number of simulations needed by the proposed simulation approach. In this context, the numerical implementation of subset simulation can be improved by considering available parallelization strategies [68, 69]. For instance, the unconditional level (stage 0) is completely parallelizable. On the other hand, although samples from the same Markov chain are not parallelizable, different Markov chains in the same stage can be computed concurrently. Additionally, low-level parallelism can also be considered to accelerate the individual runs (network solution), improving even more the numerical implementation.

3.3 Reliability sensitivity analysis

From the previous formulation of the hydraulic reliability problem, it is clear that the probability of failure depends on several factors, among them, the distribution parameters $\boldsymbol{\tau}$ of the probability density function that characterizes the uncertain system parameters $\boldsymbol{\theta}$. Hence, the failure probability depends on the distribution parameters, i.e. $P_F = P_F(\boldsymbol{\tau})$. Clearly, variations in the value for the distribution parameters will affect the reliability of the water distribution system.

3.3.1 Sensitivity measure

A traditional sensitivity measure corresponds to the gradient of the quantity of interest. In this way, the reliability sensitivity is defined as the partial derivative of the failure probability with respect to distribution parameters $\boldsymbol{\tau}$. Formally, the sensitivity of the failure probability with respect to distribution parameter τ_j is given by

$$\left.\frac{\partial P_F(\boldsymbol{\tau})}{\partial \tau_j}\right|_{\boldsymbol{\tau}^0} = \int_{\boldsymbol{\theta}\in\Omega_\theta} I_F(\boldsymbol{\theta}) \left.\frac{\partial p(\boldsymbol{\theta};\boldsymbol{\tau})}{\partial \tau_j}\right|_{\boldsymbol{\tau}^0} d\boldsymbol{\theta} \tag{3.3.1}$$

where $\boldsymbol{\tau}^0$ is the value of the distribution parameter vector where the partial derivative is evaluated. Note that, in the above equation, it has been assumed that the integration domain does not depend on τ_j and that $p(\boldsymbol{\theta};\boldsymbol{\tau})$ is differentiable with respect to τ_j. An approach for estimating the above quantity as a post-process of subset simulation has been recently introduced in [64], which is adopted in this work.

A related sensitivity measure is the so-called *elasticity coefficient*. It is based on the partial derivative in Equation (3.3.1) and formally defined as

$$e_{\tau_j} = \left.\frac{\partial P_F}{\partial \tau_j}\right|_{\boldsymbol{\tau}^0} \frac{\tau_j^0}{P_F} \tag{3.3.2}$$

where τ_j^0 is assumed to be non-zero. Observe that this quantity is dimensionless, which makes it a more objective sensitivity measure when considering parameters that are diverse in dimension [10]. This feature allows to rank the importance of the distribution parameters on the system reliability. The usefulness of this sensitivity measure is demonstrated in Chapter 4.

3.3.2 Sensitivity estimation

One single run of subset simulation allows to obtain reliability sensitivity estimates by post-processing the simulation results [64, 65]. In other words, the sensitivity estimation does not require to solve any additional hydraulic model. In this context, the partial derivative of the failure probability with respect to distribution parameter τ_j can be estimated using simulated samples as

$$\left.\frac{\partial P_F(\boldsymbol{\tau})}{\partial \tau_j}\right|_{\boldsymbol{\tau}^0} \approx p_0^{m-1} \frac{1}{N_m} \sum_{i=1}^{N_m} I_F\left(\boldsymbol{\theta}_{m-1,i}^0\right) \frac{\frac{\partial p}{\partial \tau_j}(\boldsymbol{\theta}_{m-1,i}^0;\boldsymbol{\tau}^0)}{p(\boldsymbol{\theta}_{m-1,i}^0;\boldsymbol{\tau}^0)} \tag{3.3.3}$$

where $\{\boldsymbol{\theta}_{m-1,i}^0, i = 1,\ldots,N_m\}$ is the set of samples generated at the last stage of subset simulation under probability density function $p(\boldsymbol{\theta};\boldsymbol{\tau}^0)$. Appendix B presents a more detailed description of the approach, including the derivation of Equation (3.3.3) and explicit expressions for some common distributions.

Note that, since the reliability sensitivity estimation is based on subset simulation, a single run of the algorithm provides estimates of the reliability sensitivity for different thresholds. In other words, the approach provides a full characterization of the reliability sensitivity trend with respect to different threshold levels, giving a valuable insight about the effect of distribution parameters on the hydraulic performance of the system. Clearly, the same type of information is obtained for the elasticity of the distribution parameters.

Chapter 4

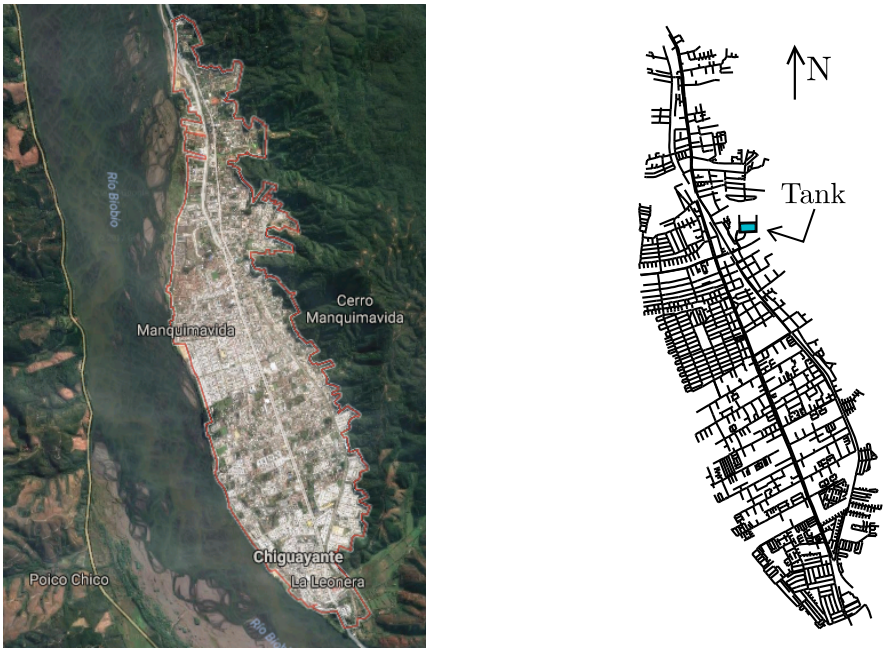
APPLICATION 1: RELIABILITY ANALYSIS OF A REAL SYSTEM

This chapter describes the hydraulic reliability assessment of a real-life, large-scale water distribution system using the approach presented in Chapter 3. The generality of the proposed framework is demonstrated with a number of analyses, including uncertainty propagation, reliability and reliability sensitivity assessment, robustness analysis with respect to local demand changes, and evaluation of the effect of pipe redundancy on the reliability of the system. Finally, the extension of the analyses to quasi-dynamic conditions (extended period simulation) is demonstrated and discussed. The results show that the stochastic framework can effectively and efficiently handle high-dimensional reliability problems arising from the analysis of complex water distribution networks.

4.1 Problem formulation

4.1.1 Description of the water distribution system

The water distribution system considered in this application problem supplies Chiguayante, a Chilean city of 140.000 inhabitants distributed over 72 [km^2] and located near Concepción. Figure 4.1-(a) shows a satellite image of the city, whereas Figure 4.1-(b) shows the layout defined in the hydraulic model of the water distribution network . For reference purposes, the North-South dimension of the system is about 8.3 [km].



(a) Satellite image of Chiguayante (Google Earth) (b) EPANET model

Figure 4.1: Water distribution system of Chiguayante

The hydraulic model of the water distribution system consists of 5978 nodes, 5655 pipes and one tank. Steady-state conditions are considered in this part of the application example.

The total demand of the network is 412.8 [l/s], distributed over 3130 nodes (in other word, 2848 nodes have zero demand). Figure 4.2 shows the nominal values of the nodal demands ($d_i, i = 1, \ldots, 3130$), which correspond to the worst-case scenario usually considered during the design stage of water distribution systems. The nominal demands range from 2.0×10^{-4} [l/s] to 3.8 [l/s].

The water distribution system comprises 190 [km] of pipelines, modeled with 5655 pipes. Figure 4.3 shows the lengths of the pipes in the network arranged in an increasing manner, ranging from 0.3 [m] to 1636 [m]. These indices are used to identify the pipes. The pipeline system consists of four different materials: high density polyethylene (HDPE), polyvinyl chloride (PVC), asbestos cement (AC) and steel. The corresponding roughness coefficients (Hazen-Williams coefficients) are 150, 140, 130 and 100, respectively. Table 4.1 shows information about the pipe diameters, which range from 57 [mm] to 502 [mm].

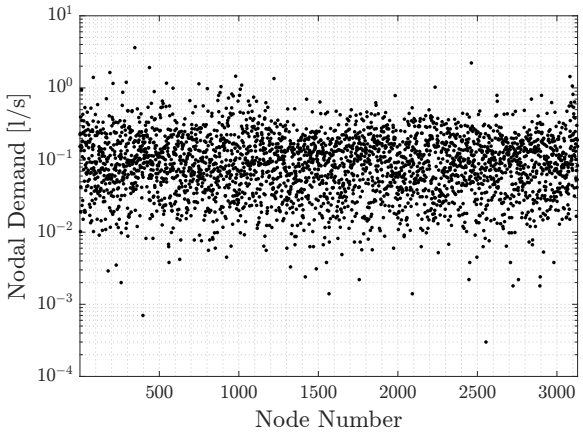


Figure 4.2: Nodal demands in hydraulic model of application example 1

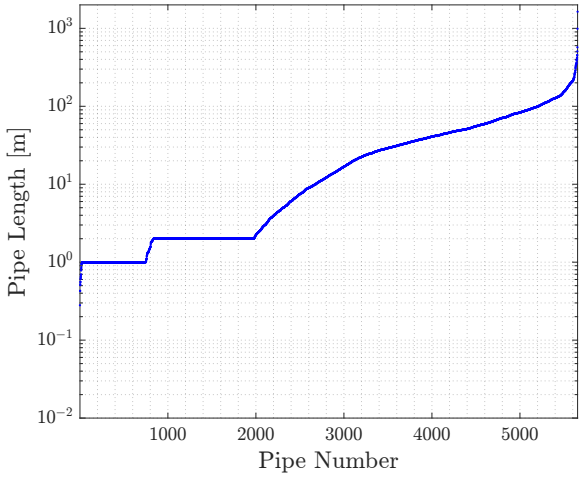


Figure 4.3: Lengths of pipes in hydraulic model of application example 1

4.1.2 Hydraulic simulator

The analysis of the system requires the use of some algorithm for performing hydraulic simulations. The multi-purpose hydraulic simulator EPANET [26] is considered in this work to deal with the physics-based analysis of water distribution networks. This software allows to perform extended-period simulation of hydraulic and water quality behavior within pressurized pipe networks. Validation calculations have shown the efficiency and flexibility of this modeling software in a number of water distribution networks, including large-scale systems.

The mathematical relationships in water distribution systems under steady-state conditions are based on flow continuity and energy conservation equations, which give a nonlinear

Table 4.1: Pipe diameters of Chiguayante's water distribution network

Pipe diameter [mm]	Number of pipes	Pipe diameter [mm]	Number of pipes
57.0	34	150.0	119
67.8	240	180.0	2
75.0	321	180.8	34
81.4	697	200.0	75
96.8	188	220.4	3
99.4	2343	226.0	5
100.0	501	250.0	9
110.0	3	300.0	12
113.0	203	321.0	6
125.0	361	350.0	20
126.6	148	361.8	16
141.0	3	400.0	3
144.0	1	500.0	23
144.6	282	502.0	3

system that EPANET iteratively solves using a Newton-type method. In this manner, the flow in each pipe and the energy head at each junction node are solved given all network characteristics. In other words, the system response at a given time is obtained. The reader is referred to Appendix D for a more detailed description about the algorithm implemented in EPANET to obtain the network solution under steady-state conditions.

Unsteady analyses can also be considered by EPANET, where the network requirements and supply are not constant but vary over time. In particular, a quasi-dynamic analysis known as extended period simulation can be performed directly in the context of the hydraulic simulator. In this type of analysis, the nodal demands and water inflow vary through the analysis period in a number of discrete time steps.

It is remarked that EPANET is implemented in all the application examples presented in this work (Chapters 4, 6 and 7). It is also remarked that the software is used as a black box and, therefore, the sampling process is decoupled from the available hydraulic simulator.

4.1.3 Uncertain parameters

The uncertainty in the water distribution system arises in both external conditions (nodal demands) and network hydraulic capacity (tank level and roughness coefficients). In this context, a total of 8786 network parameters are modeled as random variables: all

nodal demands ($d_i : i = 1, \ldots, 3130$), all pipe roughness coefficients ($RC_i : i = 1, \ldots, 5655$), and the prescribed nodal head at the storage tank (h_{tank}). Observe that this problem is clearly high-dimensional from the reliability point of view.

The nodal demands ($d_i : i = 1, \ldots, 3130$) are modeled as log-normal independent random variables with mean values equal to the corresponding nominal values (see Figure 4.2), whereas the coefficient of variation is taken as 20%. Figure 4.4-A shows the probability density function of a representative node of the network.

The prescribed water level above the discharge point of the storage tank (h_{tank}) is modeled as a log-normal random variable, with mean value equal to 3.2 [m] and coefficient of variation equal to 20%. This distribution is shown in Figure 4.4-B.

The pipe roughness coefficients are described using independent truncated normal random variables, with mean values equal to their corresponding nominal roughness coefficients, coefficient of variation of 10%, and ranging between 85% and 115% of the mean value. A compact support is chosen, since the variation of these parameters is feasible only within a certain range. Figure 4.4-C shows the distribution of a representative roughness coefficient.

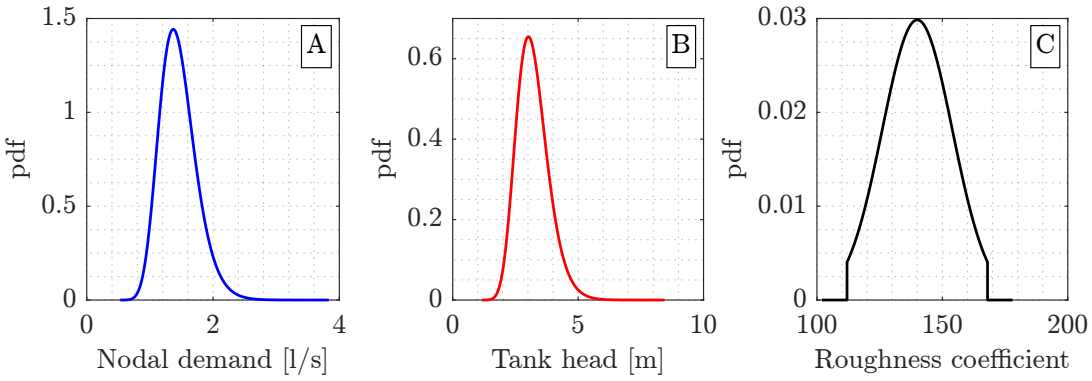


Figure 4.4: Probability density function of representative uncertain parameters. A: Nodal demand. B: Tank head. C: Roughness coefficient

4.1.4 Utility function

The utility function considers the minimum pressure head over a set of nodes Ω_{nodes} under steady-state conditions. In this application problem, Ω_{nodes} includes all nodes with

non-zero demand (3130 nodes). The utility function is then defined as

$$\mu(\boldsymbol{\theta}) = \min_{i \in \Omega_{\text{nodes}}} p_i(\boldsymbol{\theta}) \tag{4.1.1}$$

where $p_i(\boldsymbol{\theta})$ is the pressure head at node $i \in \Omega_{\text{nodes}}$. This utility function captures one of the most relevant system responses, which has a direct impact on the level of service provided to the users.

4.2 Uncertainty propagation

The results of an uncertainty propagation analysis are presented in this section. This type of analysis is crucial to gain a general insight about the hydraulic performance of the water distribution network under uncertain conditions.

As a previous step, it is relevant to study the response of the network characterized in a deterministic manner, i.e. when the values of the parameters are equal to the nominal ones. For reference purposes, Figure 4.5 shows the pressure head in all demand nodes of the network, sorted in a decreasing manner. These indices are the ones used in this chapter to identify the nodes. Note that the pressure heads vary from 17.18 [m] to 50.67 [m]. In other words, the utility function evaluated at to the nominal values of the parameters is $\mu(\boldsymbol{\theta}^{\text{nom}}) = 17.18$ [m].

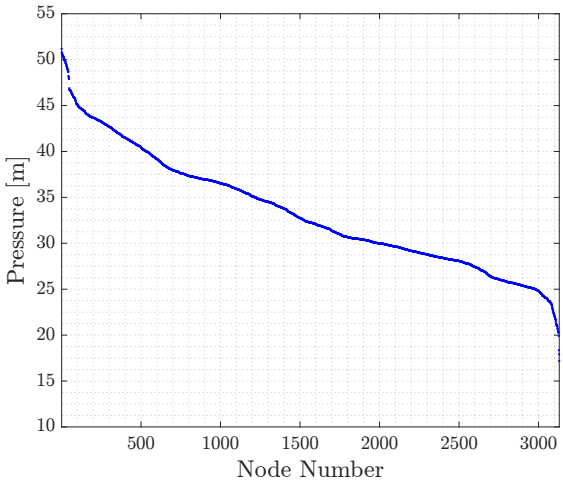


Figure 4.5: Deterministic pressure heads of all demand nodes sorted in a decreasing manner (nominal response)

The uncertainty propagation analysis is performed with a sample size equal to 5×10^4, obtaining the pressure head in all demand nodes of the network. As already discussed, the hydraulic simulations are performed using the commercial-level software EPANET. Figure 4.6 shows the 50% quantile and the 95% probability interval (i.e. both 97.5 percentile and 2.5 percentile) of all corresponding pressure heads in the demand nodes. During the analysis, it was observed that the mean pressure head in each node is very close to the 50% quantile. Observe that this figure shows the effect of the prescribed uncertain conditions into the pressure head of all demand nodes in the network.

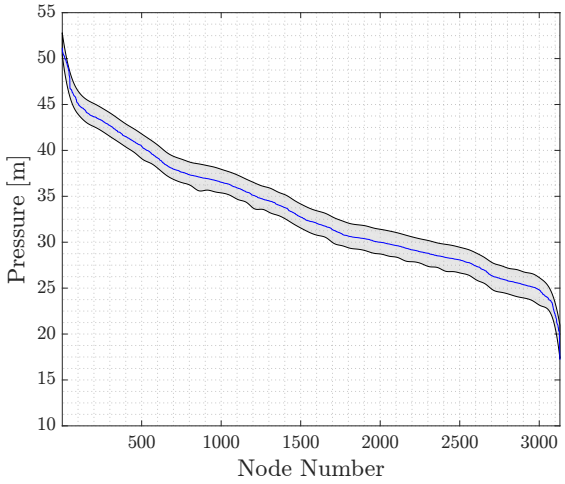


Figure 4.6: 50% quantile and 95% probability interval of the network pressure heads

Using the same set of samples, the minimum pressure head of the network for each hydraulic simulation can be obtained. Such a quantity corresponds to the utility function defined in Equation (4.1.1). Figure 4.7 shows the normalized histogram of the minimum pressure head over all demand nodes. For reference purposes, a dashed line in the same figure shows the minimum pressure associated to the deterministic response (i.e. the network characterized by the nominal values of the uncertain parameters), which is 17.18 [m]. Note that the uncertain parameters have a considerable effect on the minimum pressure head of the network. Actually, the histogram shows a variation of that response between 14 [m] and 21 [m]. Therefore, it is expected that the uncertain conditions will have a significant effect on the reliability of the system.

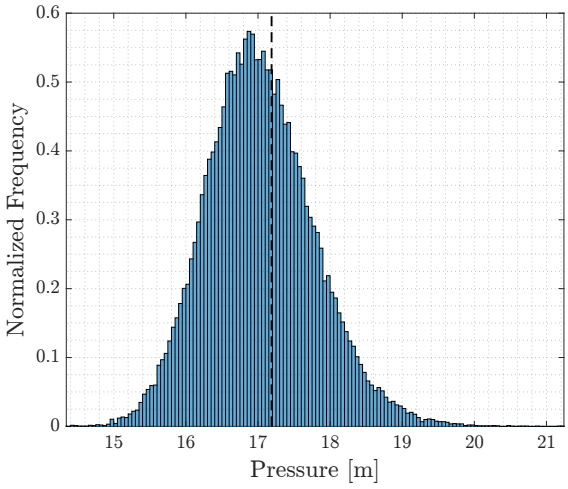


Figure 4.7: Normalized histogram of network minimum pressure head

4.3 Hydraulic performance: regular operational conditions

4.3.1 Reliability analysis

As already pointed out, a single run of subset simulation provides reliability estimates for all threshold levels up to the smallest considered in the analysis. In connection with Chapter 3, the algorithm is implemented with $p_0 = 0.1$ (conditional failure probability) and $N_s = 1000$ (number of samples at each stage). Figure 4.8 shows the trend of the failure probability versus critical threshold levels, considering a total number of 4 stages (failure probabilities from 0.90 to 10^{-4}). For illustration purposes, the average of five independent runs of subset simulation is presented here.

The information contained in the figure allows to understand the hydraulic performance of this complex utility network in terms of the minimum pressure head. It is seen that, for instance, the probability of the minimum pressure head being below 16.5 [m] is about 2×10^{-1}, while a failure probability of 10^{-3} is associated to a minimum allowable pressure equal to 14.7 [m]. An interesting result is that a critical threshold equal to 17.18 [m], which corresponds to the nominal minimum pressure head, is associated to a probability of failure equal to 60%. Thus, the network characterized in a deterministic manner is quite unreliable under the uncertain conditions considered here.

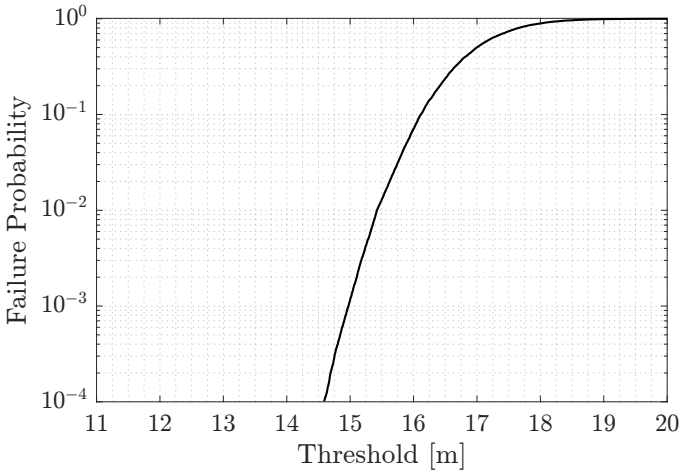


Figure 4.8: Failure probability in terms of critical threshold level

4.3.2 Reliability sensitivity analysis

The *elasticity coefficient* was introduced in Equation (3.3.2). This dimensionless reliability sensitivity measure can be used to rank the importance of distribution parameters on the system reliability when they are diverse in dimension. As indicated in Chapter 3, the proposed reliability sensitivity analysis is a post processing of subset simulation, thus allowing to obtain the entire trend of the sensitivity measure in terms of different threshold levels and not only for one specific value.

Figure 4.10 shows the elasticity coefficients of the failure probability, in terms of the threshold level, with respect to the mean value of the most influential parameters. The parameters considered in the figure are the mean values of the prescribed head of the storage tank and the roughness coefficients of pipe numbers 5558, 5648 and 5655 (see Figure 4.9). An average of ten independent runs is considered in the figure. It is seen that the prescribed head of the storage tank plays a major role in the probability of failure, as expected. Besides, note that all the elasticity coefficients are negative. This means that increasing the mean value of these parameters decreases the probability of failure, which is reasonable from the physical point of view: a higher tank level increases the overall pressure heads, while higher Hazen-Williams coefficients increase the hydraulic capacity of pipelines. Another interesting observation is that the elasticity coefficients increase in magnitude as the threshold level decreases. Therefore, the failure probability becomes more sensitive to these distribution parameters as the probability of failure becomes smaller.

Validation calculations have shown that the elasticity coefficients of the failure prob-

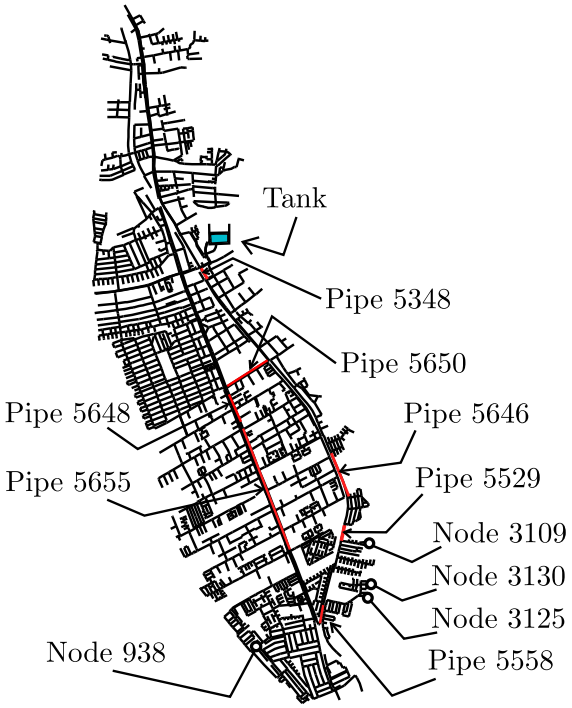


Figure 4.9: Identification of most influential pipes in terms of their roughness coefficients, failure modes, nonfunctional pipes and some illustrative nodes

ability with respect to the mean value of the roughness coefficients are either negative or very close to zero. On the other hand, it is found that the elasticity of the failure probability with respect to the mean value of individual nodal demands shows a mixed behavior: some of the elasticity coefficients are positive, while others are negative. In other words, the effect of one particular nodal demand can be beneficial or detrimental with respect to the network reliability. This behavior is due to the complex interaction between the nodal demands and their effect on the pressure heads of the network. In any case, the elasticity coefficients associated to the mean value of the nodal demands are relatively small.

A similar reliability sensitivity analysis can be performed with respect to the standard deviation of the network state parameters. This type of analysis provides information to identify the parameters whose uncertainty plays a major role in affecting the hydraulic reliability of the network. Figure 4.11 presents the elasticity coefficients in terms of the threshold level for the most influential parameters: the prescribed head of the regulation tank, and the roughness coefficients of pipe numbers 5558, 5648 and 5655 (see Figure 4.9). It is observed that the uncertainty of the prescribed tank head presents the most important effect on the hydraulic reliability of the system. Note also that the elasticity of the failure

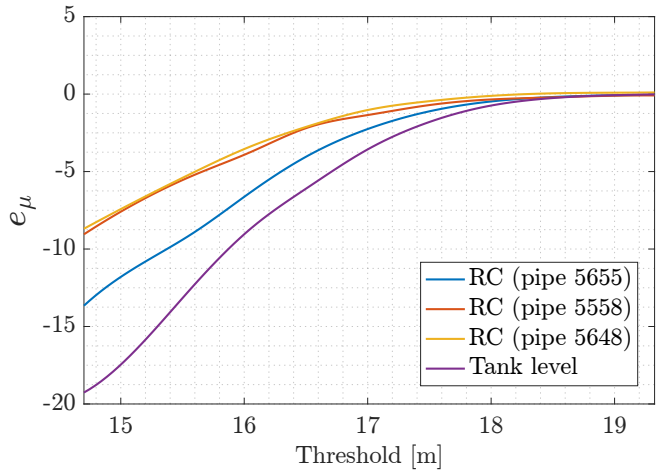


Figure 4.10: Elasticity coefficients versus threshold. Distribution parameter: mean value

probability with respect to the standard deviation of the different parameters is positive for threshold levels associated to small failure probabilities, i.e. minimum allowable pressure head below 17.0 [m]. Thus, the probability of failure decreases when reducing the variability of the system parameters. This information gives a valuable insight into the effect of uncertain system parameters on the reliability of the network.

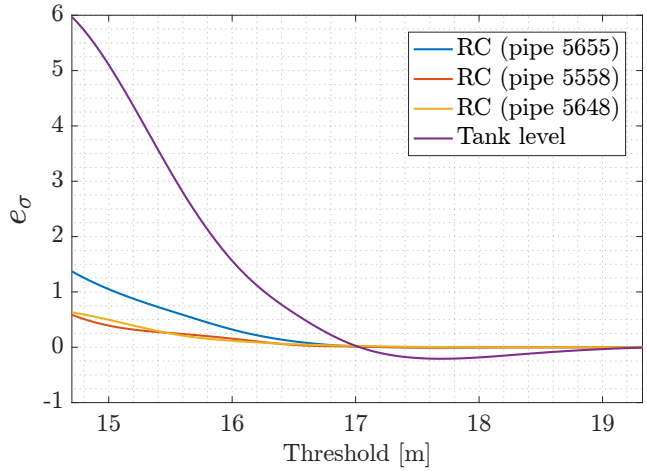


Figure 4.11: Elasticity coefficients versus threshold. Distribution parameter: standard deviation

4.3.3 Failure analysis

One important question regarding the hydraulic performance assessment of complex water distribution networks is how likely a given failure event is. This issue is addressed by

the hydraulic reliability analysis of the system, as discussed in Section 4.3.1. Nevertheless, it is pivotal to also understand what happens when a system fails, identifying the specific reasons that lead to failure and how does the system behave when failure occurs. This type of study is known as *failure analysis*, a crucial task for proper decision-making regarding the design, maintenance and operation of real-life water distribution networks. The following results demonstrate that the proposed framework provides valuable information to this end.

4.3.3.1 Conditional distributions of uncertain parameters

A single run of subset simulation generates samples conditional to the sequence of intermediate failure events (see Appendix A). As discussed in [25], these samples can be used directly to perform failure analysis. In particular, the conditional marginal histograms of uncertain parameters suggest the most probable causes of failure. If the conditional histogram of a certain parameter is significantly different from the unconditional probability density function, then that parameter is expected to have a high influence towards determining failure. Conversely, if the conditional histogram of that parameter remains similar to the unconditional probability density function, then it is not likely to be relevant regarding failure occurrence. In order to demonstrate this type of analysis, consider a failure event F^* defined by a minimum allowable pressure $\mu^* = 15$ [m], that is

$$F^* = \{\boldsymbol{\theta} \in \Omega_\theta : \mu(\boldsymbol{\theta}) < \mu^* = 15 \text{ [m]}\} \tag{4.3.1}$$

Observe that the corresponding conditional samples are obtained in the last stage of subset simulation (see Figure 4.8). Figure 4.12 shows the conditional histograms of three parameters: the prescribed head of the storage tank (h_{tank}), and the roughness coefficients of pipes 5558 and 5655 (identified in Figure 4.9). For reference purposes, the corresponding unconditional probability density functions are plotted with dashed lines. Note that the conditional histograms are different from the unconditional distributions. In fact, the prescribed head of the storage tank shows a significant shift of probability mass towards the low value region, while the histograms associated to the roughness coefficients show some shift towards the low value region. Hence, these parameters seem to be part of the dominant factors that govern failure. These results are consistent with Figures 4.10 and 4.11, where these three parameters were identified as the most important ones regarding system reliability. It is also important to remark that the shift of probability mass towards the low value region in all cases is consistent with the physical behavior of the system: lower tank heads diminish hydraulic heads in all nodes, whereas lower Hazen-Williams coefficients are

associated with higher head losses in pipes.

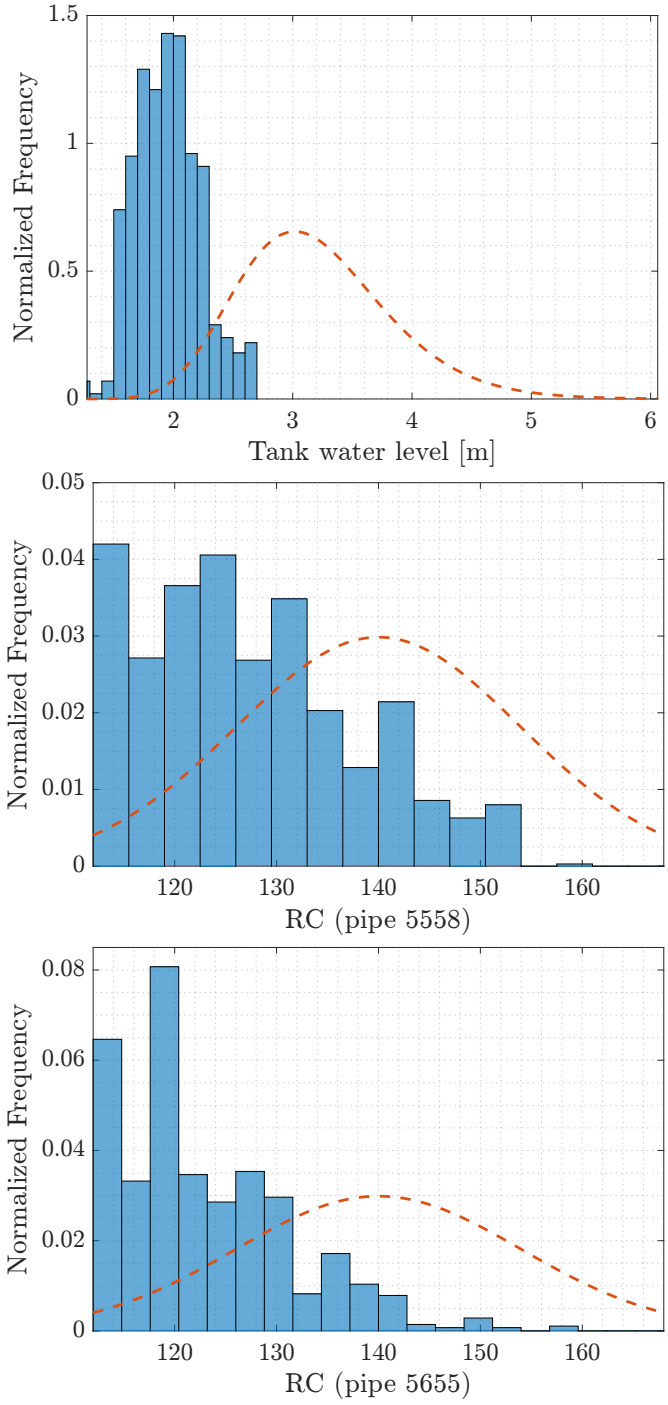


Figure 4.12: Conditional histograms of samples. Tank head and roughness coefficients

A similar analysis can be made with respect to nodal demands. However, in this case,

it was observed that the conditional distributions are not significantly different to the unconditional ones. Figure 4.13 shows the conditional histograms of two representative nodal demands, namely node numbers 938 and 3125. These nodes are identified in Figure 4.9. Again, the unconditional probability functions are plotted with dashed lines. Note that the conditional histograms are not significantly different from the unconditional distributions. Slight shifts of the probability mass are observed, which can be directed towards higher values (e.g. nodal demand 3125) or lower values (e.g. nodal demand 938). Similar results are observed for the remaining nodal demands. Therefore, it is expected that a single nodal demand is not relevant to determine failure, which is consistent with the results obtained in Section 4.3.2.

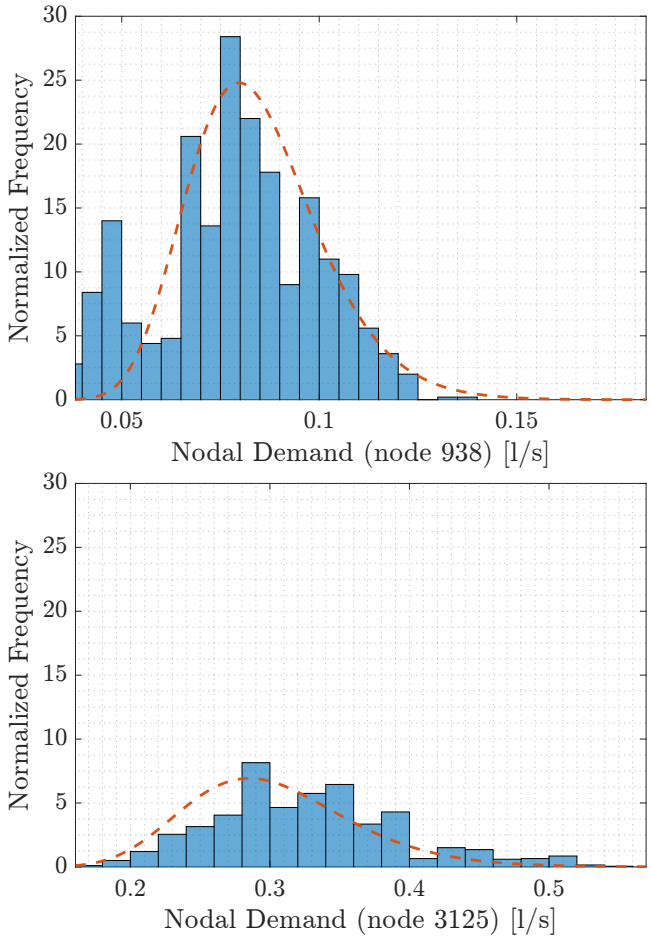


Figure 4.13: Conditional histogram of samples. Nodal demands

4.3.3.2 Failure modes

Real-life utility networks can be understood as a system composed of different sub-components or sub-systems. Those individual sub-components interact in different and complex ways to determine the performance of the entire network. Therefore, it is important to identify which sub-components are the ones that govern failure, in order to provide information for decision-making. Such issue is directly related to failure analysis.

According to the utility function defined in Equation (4.1.1), the corresponding water distribution network can be understood as a set of components connected in series, where each of them corresponds to a node whose response is the corresponding pressure head. In this way, the i-th failure mode F^i is defined as

$$F^i = \{\boldsymbol{\theta} \in \Omega_\theta : p_i(\boldsymbol{\theta}) < \mu^*\} \tag{4.3.2}$$

where $p_i(\boldsymbol{\theta})$ is the pressure head at node $i \in \Omega_{\text{nodes}}$. In other words, the i-th failure mode identifies the region in the space of uncertain parameters that gives pressure head at node $i \in \Omega_{\text{nodes}}$ below the critical threshold μ^*.

Samples obtained from subset simulation can be directly used to compute the conditional failure probability of the i-th failure mode, that is, $P(F^i|F)$. This quantity corresponds to the probability that a certain node fails, given that system failure has occurred. Considering the failure event F^* defined in Equation (4.3.1), the conditional failure probabilities of all failure modes are shown in Table 4.2. Recall that this information is obtained directly from the last stage of subset simulation. In this case, $P(F^{3130}|F^*) = 1$ and $P(F^i|F^*) = 0, i = 1, 3129$. Then, it is clear that node 3130 is the most critical one in terms of pressure levels. This analysis gives relevant information about the performance of different zones of the network, which can be used to improve the hydraulic reliability of the system. For instance, adding pipes to improve the connectivity of node 3130 or including control elements in the surrounding area (e.g. hydropneumatic tanks) seem to be reasonable options to improve the performance of the water distribution network.

Table 4.2: Conditional probabilities of failure modes

Failure mode (F^i) node number	$P(F^i\|F^*)$ %
1 - 3129	0
3130	100

4.4 Hydraulic performance in alternative scenarios

For illustration purposes, alternative operational scenarios of the water distribution system of Chiguayante are studied in this section. It is shown that the proposed stochastic framework provides valuable information that allows the comparison, in terms of the hydraulic performance, of different configurations of a water distribution system.

4.4.1 Robustness with respect to demand changes

Water distribution networks are subjected to constant changes in the spatial configuration of their demand, as a result of the dynamic nature of users. The proposed framework allows to evaluate the changes on the hydraulic reliability of the system when different scenarios of demand configuration are considered. Such analysis provides a valuable insight about the robustness of water distribution networks with respect to demand changes. This type of information can be useful for future planning considerations, such as new commercial, residential or industrial developments.

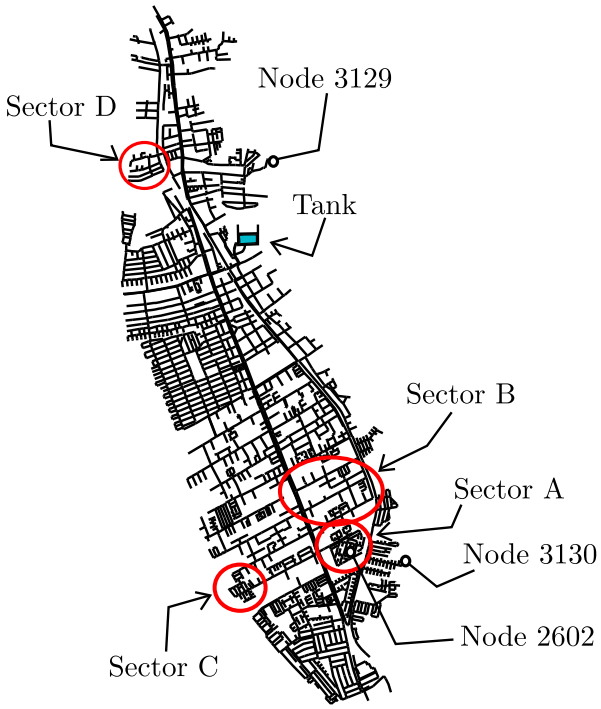


Figure 4.14: Identification of demand zones and some illustrative nodes

For illustration purposes, four different scenarios are considered in this work. In each scenario, the demand in a sector of the network is assumed to be three times the one

considered under regular operational conditions (Section 4.3). The rest of the system characteristics are kept unmodified. In other words, the mean values of the nodal demands in a specific sector are multiplied by a factor equal to three in each scenario. Figure 4.14 shows the sectors that are modified in each case. The water demand in sectors A, B, C and D corresponds to 1.9%, 2.9%, 1.4% and 1.1% of the total demand in the nominal network, respectively.

Figure 4.15 shows the probability of failure versus threshold level for all the demand scenarios. For illustration purposes, the curve corresponding to regular operational conditions (i.e., the one presented in Figure 4.8) is reproduced in the figure, identified as the *nominal case*. It is seen that, in all scenarios, the failure probability for any given threshold is increased with respect to the nominal case. This is a expected result from the physical point of view, since the hydraulic heads in the network are expected to decrease under the perturbations considered here (local demand increases).

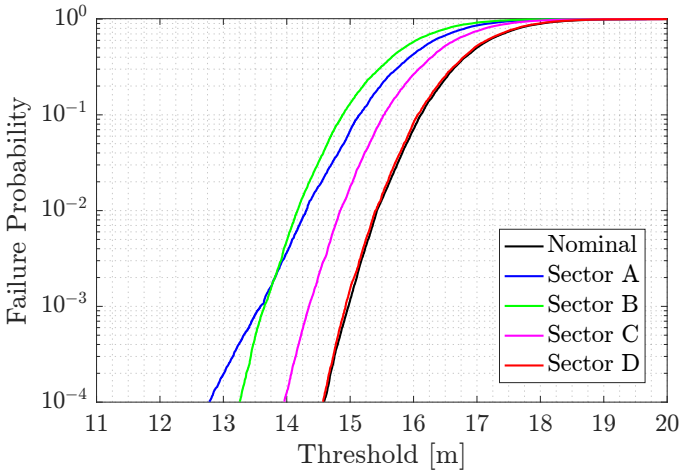


Figure 4.15: Failure probability in terms of critical threshold level. Sectional demand increments

Different demand configurations have different effects on the hydraulic performance of the system. For instance, the curve associated to sector D is very similar to the one of the *nominal* case, which means that a demand increase in sector D has almost no effect on the hydraulic reliability of the water distribution network. On the contrary, the curves associated to Sectors A and B show higher failure probabilities than the ones observed in the nominal case. In other words, demand increases in these sectors have a significant influence on the hydraulic performance of the water distribution network. Increasing the

demand in Sector C, on the other hand, has an intermediate effect on hydraulic reliability. To illustrate the effect of the different demand conditions, Table 4.3 shows the failure probabilities for each scenario associated to an allowable minimum pressure of 14.60 [m]. For example, note that the failure probability changes from 1×10^{-4} in the nominal case to 4×10^{-2} under a demand increase in Sector B. It is also interesting to note that the effect of Sector A for small threshold levels, which are associated to small failure probabilities, is more significant than sector B, even though the total demand in sector B (2.9% of the total demand in the network) is higher than the one in Sector A (1.9% of the total demand in the network).

Table 4.3: Failure probabilities for different demand scenarios, $\mu^* = 14.60$ [m]

Scenario	Failure probability
Nominal case	1×10^{-4}
Sector A	2×10^{-2}
Sector B	4×10^{-2}
Sector C	3×10^{-3}
Sector D	1×10^{-4}

Failure analysis can be used in this type of study to understand how the changes in the demand distribution modify the behavior of the system. For illustration purposes, Table 4.4 shows the conditional probabilities of the failure modes for each scenario, conditional to some failure event F^*, that is, $P(F^i|F^*)$. Here, F^i is the i-th failure mode defined in Equation (4.3.2) and F^* is a failure event such that $P[F^*] = 10^{-3}$. In other words, the results presented in Table 4.4 are obtained with the samples generated at the last stage of subset simulation. The corresponding nodes are presented in Figure 4.14. When the demand of Sectors B and C increases, the minimum pressure is always observed in node 3130, i.e. the failure mode remains unchanged with respect to the *nominal* case. In a similar manner, node 3129 is the one that fails under an increase of the demand in Sector D. On the other hand, two failure modes are observed when increasing the demand in Sector A: nodes 2602 and 3130, with conditional probabilities $P(F^{2602}|F^*) = 86\%$ and $P(F^{3130}|F^*) = 14\%$, respectively. This shows that node 2602 is more likely to fail than node 3130 in this scenario. It is seen that the most critical zone of the network, in terms of the pressure levels, can change when modifying the spatial distribution of the demand. The approach is able to identify these critical zones based on a single run of subset simulation for each scenario.

Table 4.4: Failure modes for different scenarios of demand changes

Scenario	Failure mode (F^i) node number	$P(F^i\|F^*)$ %
A	2602	86
	3130	14
B	3130	100
C	3130	100
D	3129	100

4.4.2 Effect of redundancy

One important concept in the design of water distribution networks is that of redundancy [7, 20, 70], which allows to ensure the critical function of the system when some components are not operating. This subsection studies the hydraulic performance of the network when some pipes are nonfunctional. In particular, five network configurations are studied. Each of them assumes that only one pipe is not working, due to repair operations or some other event. The corresponding link element is removed from the hydraulic model. In other words, each scenario assumes that no flow is developed in a given pipeline. The pipes under consideration (representative links of more than 90 [m] long) and some nodes of interest are presented in Figure 4.9.

The curves of failure probability in terms of the threshold level for the different configurations are shown in Figure 4.16, including the case with all pipes being functional (i.e., the curve presented in Figure 4.8). Note that, for all threshold levels, the probabilities of failure under regular operational conditions are lower than the ones observed in scenarios with nonfunctional pipes. This clearly shows that increasing the levels of pipe redundancy has a beneficial effect into the hydraulic reliability of this system. Another interesting observation is that different nonfunctional pipes have different effects on the hydraulic reliability. For instance, the results observed in the scenarios associated to pipe 5348 and pipe 5529 are very similar to the ones obtained when all pipes are functional. On the other hand, the reliability of the system decreases significantly by removing pipe 5646 or pipe 5650. In this manner, this type of analyses can identify the pipes that mostly affect the hydraulic reliability of the system when they are non-operational.

The proposed stochastic framework provides the information required to identify the failure modes in each scenario: failure analysis indicates that node 3130 is the one that fails for the configurations defined by removing pipe numbers 5348, 5529 and 5650, whereas

node number 3109 (see Figure 4.9) fails for the configuration defined by removing pipe number 5646. This information is relevant, for instance, to make informed-decisions regarding network management, such as scheduling maintenance operations.

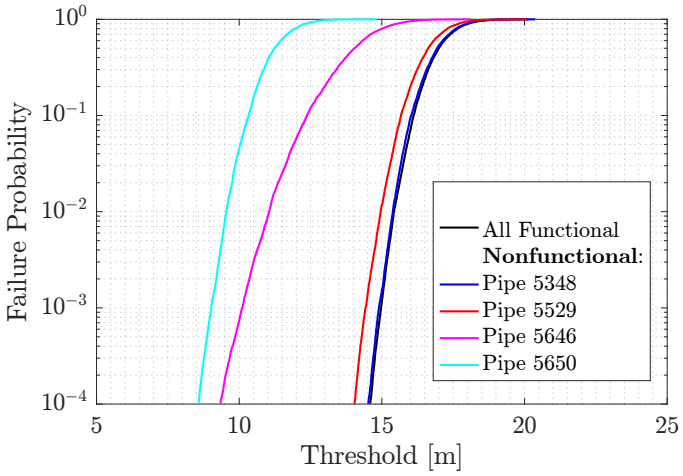


Figure 4.16: Failure probability in terms of critical threshold level for different redundancy scenarios

The analysis presented here can be extended to the case when the probability of the network being at a given configuration is known. In that case, the total failure probability can be estimated by multiplying the chance of each configuration occurring by the failure probability of this configuration, and then summing up all possible configurations [20]. This type of analysis can also be used to compare the performance, in terms of their redundancy, of non-fixed layout networks, which is particularly important during design stages. Finally, it is noted that, although the previous analysis was focused on the redundancy of pipes, it can be extended, in principle, to other components and devices as well.

4.5 Extension to quasi-dynamic conditions

The stochastic approach can be extended to study water distribution systems under quasi-dynamic conditions. This type of analysis allows to assess the hydraulic performance of water distribution networks during a period of interest (for example, a regular day of operation). In this section, a hydraulic reliability problem is formulated to study the water distribution network of Chiguayante under quasi-dynamic conditions.

4.5.1 Extended period simulation

The extended period simulation considered in EPANET [26] calculates the hydraulic response of the system at N_T instants of time $\{t_1, \ldots, t_{N_T}\}$. The nodal heads and pipe flows are computed for each discrete time step by considering steady-state conditions. The prescribed conditions at each time step (e.g. water volume or water quality in regulation elements) are obtained by considering the information from the previous instant and additional conditions (e.g. water inflow at each time step). The quasi-dynamic analysis considered here involves 24 instants of time, starting at 00:00 hours and ending at 23:00 hours. In other words, 24 steady-state simulations are needed to characterize the hydraulic response of the water distribution network in this scenario.

The regulation tank and pipelines of the hydraulic system were described in Section 4.1.1. On the other hand, it is assumed that all the demand nodes follow the same normalized demand pattern (see Figure 4.17). Note that the maximum demand takes place at 13:00 hours. For illustration purposes, an intermediate scenario in terms of the demand level is considered here. In this context, the maximum nodal demands are equal to the values presented in Figure 4.2, but multiplied by a factor equal to 75%.

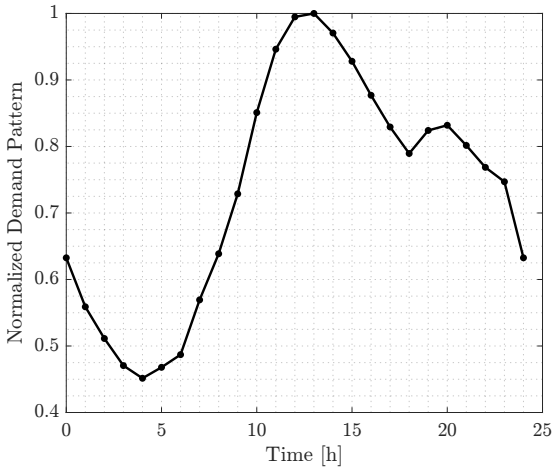


Figure 4.17: Normalized demand pattern for all nodes, quasi-dynamic conditions

A pumping system supplies water to the regulation tank. The nominal volume of pumped water over time is shown in Figure 4.18 (left). The water volume in the regulation tank corresponding to the deterministic network (i.e. the hydraulic model characterized by the nominal values of the network state parameters) is presented in Figure 4.18 (right).

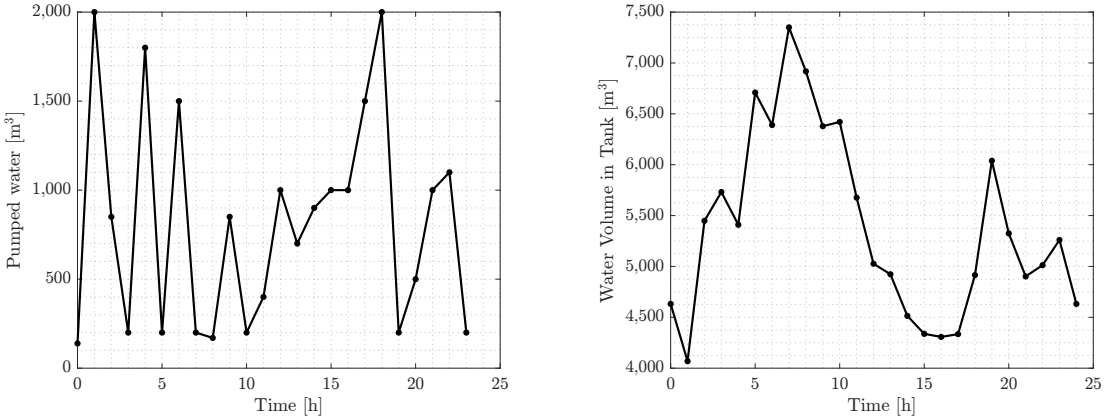


Figure 4.18: Left: network inflow over time, deterministic network. Right: water volume in tank, deterministic network

4.5.2 Uncertain parameters

Four different groups of uncertain parameters are considered here: the pipe roughness coefficients, the prescribed tank head at the beginning of the simulation, the demands at each node and time, and the volume of water pumped at each time step. The probabilistic characterization of these parameters is described in what follows.

- The roughness coefficients of the pipes (Hazen-Williams coefficients) are characterized in the same manner that in Section 4.1.3. This gives a total of 5655 random variables.

- The initial tank head is characterized as a uniform random variable, centered at the nominal value and with range equal to 2.0 [m]. This gives one random variable.

- The demand at each time instant and node is modeled as a log-normal random variable, with mean value equal to the corresponding nominal value and a coefficient of variation of 20%. All variables are independent. Since there are 3130 nodes and 24 time steps, there is a total of 75120 random variables associated to the demand conditions.

- The network inflow at each time step is characterized as a truncated normal random variable, with mean value equal to the nominal value, a coefficient of variation of 15%, and ranging between 75% and 125% of the mean value. This gives a total of 23 random variables.

This formulation gives a total of 80799 random variables. It is clear, then, that this problem is highly dimensional from the reliability point of view.

4.5.3 Utility function

The utility function $\mu(\boldsymbol{\theta})$ is defined as the absolute minimum pressure head in the network during the simulation period, that is

$$\mu(\boldsymbol{\theta}) = \min_{\substack{i \in \Omega_{\text{nodes}} \\ t \in \{t_1, \ldots, t_{N_T}\}}} p_{i,t}(\boldsymbol{\theta}) \tag{4.5.1}$$

where $p_{i,t}(\boldsymbol{\theta})$ is the pressure head at node $i \in \Omega_{\text{nodes}}$ and time $t \in \{t_1, \ldots, t_{N_T}\}$.

4.5.4 Uncertainty propagation analysis

An uncertainty propagation analysis is first carried out, considering a sample size of 5×10^4. Figure 4.19 shows the normalized histogram of the network minimum pressure observed at three representative time instants: 00:00, 13:00 and 20:00 hours. It is observed that the distribution of minimum pressure at 00:00 is almost uniform, whereas the distribution at 13:00 and 20:00 are more similar to a Gaussian distribution. Furthermore, the minimum pressure at 00:00 ranges between 19.2 [m] and 22.6 [m], while the minimum pressure at 13:00 ranges from 15.0 [m] to 22.0 [m]. In other words, the variability of the minimum pressure at 13:00 is larger than the one observed at 00:00. This analysis provides an insight into the effect of the uncertain conditions into the hydraulic performance of the network during a complete simulation period.

Figure 4.20 shows the normalized histogram of the absolute minimum pressure head in the network during the simulation period, that is, the histogram of the utility function values obtained for the sample under consideration. For comparison purposes, the absolute minimum pressure head of the network characterized in a deterministic manner is presented with a dashed line in the same figure, corresponding to 19.09 [m]. Note that this value is higher than the sample average. On the other hand, a variation of approximately 8 [m] in the system response is observed. In fact, the histogram is flatter than the one presented in Figure 4.7. Thus, the effect of uncertainty into the hydraulic performance of the system is considerable.

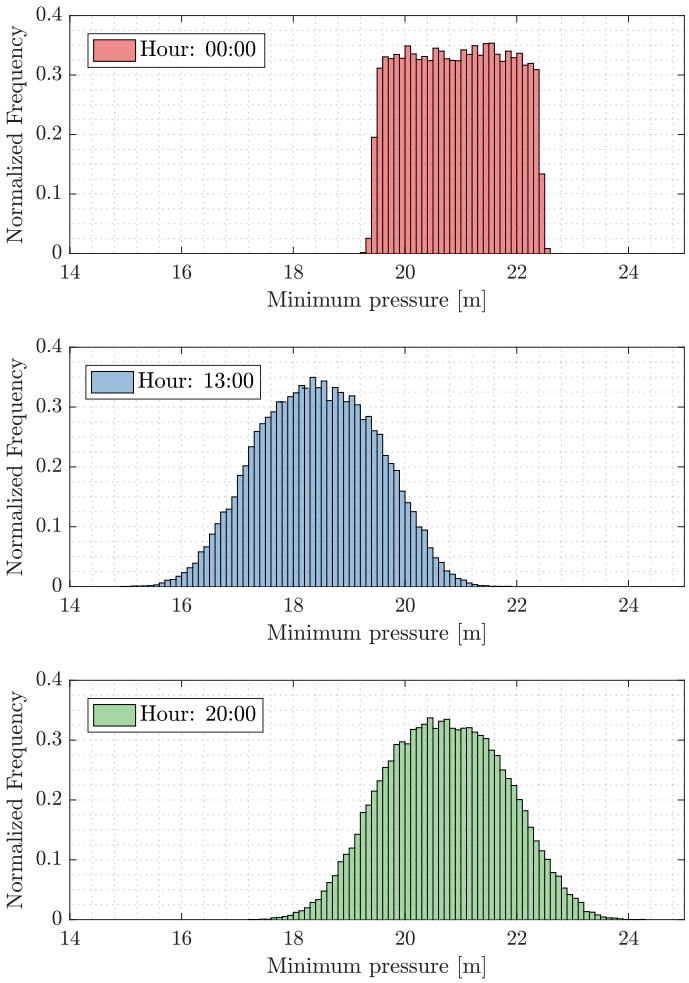


Figure 4.19: Normalized histogram of network minimum pressure head at different instants of time

4.5.5 Hydraulic reliability analysis

The results of the hydraulic reliability analysis considering the utility function presented in Section 4.5.3 are now discussed. Note that this problem is very challenging, since it is highly dimensional from the probabilistic point of view (80799 random variables). Figure 4.21 (left) shows the failure probability versus threshold levels obtained in three independent runs of subset simulation. For comparison purposes, the curve corresponding to the sample used in the uncertainty propagation analysis is also presented. Such curve, identified as *MCS*, reaches a failure probability equal to 3×10^{-4}. The coefficient of variation corresponding to the MCS estimate of this failure probability is equal to 26% (sample size of 5×10^4). It is observed that the differences between the curves associated to independent runs are rather small. The curves are also consistent with the results

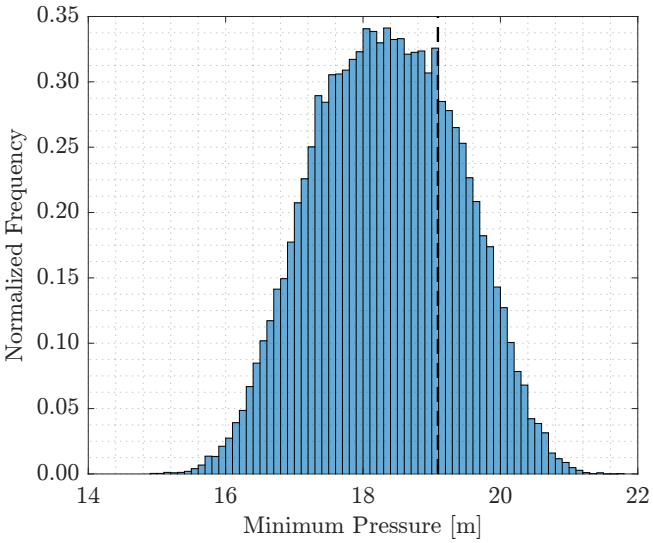


Figure 4.20: Normalized histogram of network minimum pressure head, quasi-dynamic conditions

obtained by direct simulation, notwithstanding the high dimensionality of the reliability problem (80799 random variables). The good performance of subset simulation for this case can be explained based on the fact that a relatively small number of uncertain parameters are predominant towards determining the hydraulic performance of this particular water network, similar to the behavior observed in sections 4.3.2 and 4.3.3. Nonetheless, different scenarios in terms of the probabilistic characterization of the system (e.g. the inclusion of spatio-temporal correlation of nodal demands) can lead to a greater number of random variables having an important effect on the hydraulic reliability of the network. The evaluation of the performance of the proposed approach in such scenarios is left for future research efforts.

Figure 4.21 (right) presents the failure probability trend versus thresholds obtained as an average of five independent simulation runs. As in Section 4.3.1, this information allows to understand the hydraulic performance of the network for different levels of minimum allowable pressure. For instance, if a minimum allowable pressure of 15.5 [m] is considered, then the probability of failure is equal to 10^{-3}.

4.5.6 Failure analysis

Following the ideas presented in Section 4.3.3, the failure mode F_i^t corresponds to the region in the space of network state parameters giving an unacceptable pressure head at

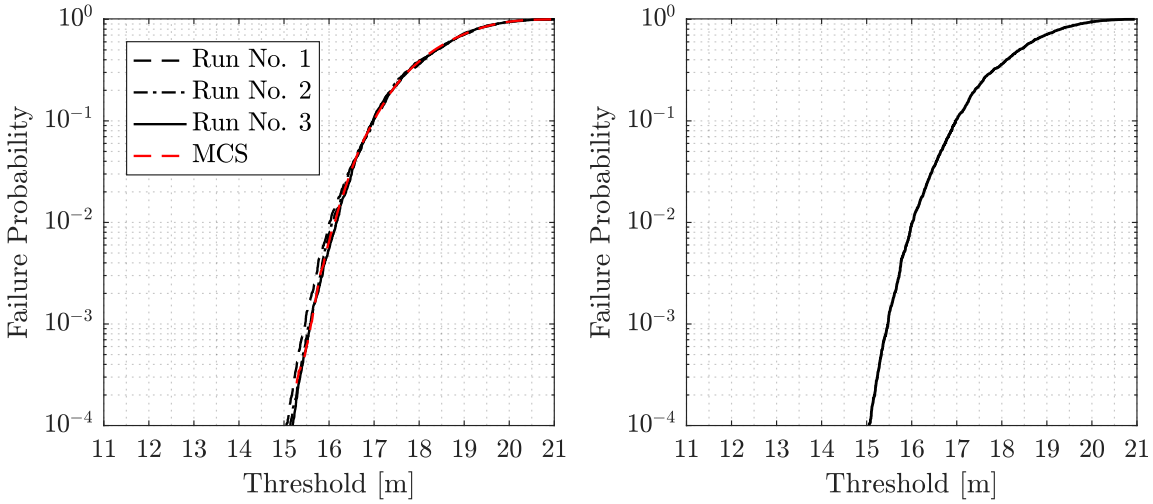


Figure 4.21: Failure probability versus threshold level in quasi-dynamic conditions. Left: Monte Carlo Simulation (up to 3×10^{-4}) and three independent runs of subset simulation (up to 10^{-4}). Right: average of five independent runs of subset simulation

node $i \in \Omega_{\text{nodes}}$ and time $t \in \{t_1, \ldots, t_{N_T}\}$. Formally,

$$F_i^t = \{\boldsymbol{\theta} \in \Omega_\theta : p_{i,t}(\boldsymbol{\theta}) < \mu^*\} \tag{4.5.2}$$

In the same way, it is also important to identify the nodes that are more likely to present the lowest pressure heads during the entire simulation period. In this context, the *spatial failure modes* F_i, $i \in \Omega_{\text{nodes}}$ are defined as

$$F_i = \bigcup_{t \in \{t_1, \ldots, t_{N_T}\}} F_i^t \tag{4.5.3}$$

$$F_i = \left\{ \boldsymbol{\theta} \in \Omega_\theta : \min_{t \in \{t_1, \ldots, t_{N_T}\}} p_{i,t}(\boldsymbol{\theta}) < \mu^* \right\} \tag{4.5.4}$$

Similarly, the *temporal failure modes* F^t, $t \in \{t_1, \ldots, t_{N_T}\}$, are only concerned with the specific instant of time when failure happens. These failure modes are defined as

$$F^t = \bigcup_{i \in \Omega_{\text{nodes}}} F_i^t \tag{4.5.5}$$

$$F^t = \left\{ \boldsymbol{\theta} \in \Omega_\theta : \min_{i \in \Omega_{\text{nodes}}} p_{i,t}(\boldsymbol{\theta}) < \mu^* \right\} \tag{4.5.6}$$

Then, it is possible to estimate, as a post-process of the reliability assessment, the chance that node i fails at a specific time t conditional to a certain failure event $F^\star$, that is

$P(F_i^t|F^\star)$. For illustration purposes, Table 4.5 shows the probabilities of each failure mode conditional to the occurrence of a failure event F^*. Such failure event verifies $P(F^\star) = 10^{-3}$, according to Figure 4.21 (right). This means that the conditional probabilities are estimated based on the samples at the last stage of subset simulation. Note that the minimum pressure is always observed at node 3130. Thus, increasing the network capacity in the adjacent zone seems to be likely to increase the overall hydraulic reliability of the water distribution network.

Table 4.5: Conditional probabilities for different failure modes, quasi-dynamic conditions

Node (i)	$P(F_i\|F^\star)$	$P(F_i^t\|F^\star)$
3130	100%	12:00 (03.6%)
		13:00 (66.3%)
		14:00 (29.9%)
		15:00 (00.2%)

4.6 Final remarks

This chapter presented an application example involving the hydraulic reliability analysis of a large-scale, real-life water distribution network in the south of Chile. The analysis was formulated in terms of minimum pressure heads, although other responses of interest can be considered as well. Different types of analyses were carried out, including uncertainty propagation, hydraulic reliability assessment, reliability sensitivity analysis, and failure analysis. Several scenarios were addressed, such as different demand configurations and different non-functional pipes. The extension of the analyses to quasi-dynamic conditions was also demonstrated, showing the capabilities of the proposed approach in the context of extended period simulation. In all cases, the proposed framework allows to gain a valuable insight about the hydraulic performance of the water distribution network under uncertain conditions. Such information can be useful for assisting water utility managers to make complex decisions regarding the design, maintenance and operation of complex water distribution networks.

The results demonstrate that the methodology effectively and efficiently handles high-dimensional reliability problems arising from hydraulic performance assessment of real-life water distribution systems. It was shown that the performance of the proposed approach remains unaffected by the complexity of the distribution system and the size of the relia-

bility problem, as opposed to previously reported works in this subject. In other words, the proposed methodology represents a change of paradigm in the uncertainty analysis of water distribution networks, since the complexity and dimensionality of the network is not a restriction anymore.

Chapter 5

BAYESIAN APPROACH FOR DETECTION-RELATED PROBLEMS

Detection-related problems are directly related to the hydraulic performance of water distribution networks. Examples of these problems are leakage detection, contaminant source characterization and connectivity detection. Clearly, solving this type of problems can help to improve the overall level of service of water utility networks. This chapter formulates a simulation-based Bayesian model updating approach to handle detection-related problems in complex water distribution networks. The proposed methodology provides explicit representations of unavoidable modeling errors and measurement noise arising in detection processes of real-life systems. The approach is also able to handle ill-conditioned problems usually found in real-life water distribution systems.

5.1 Background

Detection-related problems correspond to the identification of irregular operational conditions that affect the performance of a water distribution network. Available information to this end includes field measurements, prior engineering knowledge and physics-based modeling. As discussed in Section 2.2, some examples of this class of problems in the context of water distribution networks are contaminant source characterization, leakage detec-

tion and connectivity detection. Detection processes involve unavoidable uncertainties that deteriorate the quality of available information, such as measurement noise and modeling errors. Moreover, practical issues regarding monitoring processes limit the amount and quality of measured data. Thus, detection-related problems usually arise as ill-conditioned in real-life applications.

This work introduces a simulation-based Bayesian model updating approach to handle detection-related problems in water distribution networks. The approach provides a realistic representation of the uncertainties associated with hydraulic modeling, measured data and prior engineering information. Here, the Bayesian framework is adapted to the problems of interest and it is coupled with a commercial-level hydraulic simulator [26]. Although the proposed methodology can be used, in principle, for a variety of identification-type of problems in pressurized distribution networks, the application examples presented here address the problems of leakage detection (see Chapter 6) and connectivity detection (see Chapter 7) in real-life water distribution networks.

5.2 Bayesian model class selection

In this section, detection-related problems are formulated as *Bayesian model class selection* problems. Suppose that a given network is not operating with the expected level of performance, due to some unknown and irregular condition that needs to be identified. Awareness about such condition can be provided, for instance, by water utility managers expertise. Consider now measurement data D that is obtained from a water distribution network. For instance, data D can contain pressure or flow measurements at some network points, which may suggest the presence of leakage in the network.

Assume now that N_{class} feasible hypotheses to characterize the unknown condition have been formulated. Each hypothesis is described by means of a probabilistic model class $M^\ell, \ell = 1, \ldots, N_{\text{class}}$, which is defined as the set of all hydraulic models $M^\ell\left(\boldsymbol{\theta}^\ell\right)$ parametrized by the network state vector $\boldsymbol{\theta}^\ell \in \mathbb{R}^{n_p^\ell}$ [51]. This vector is described in a probabilistic manner by means of a multivariate distribution. For instance, if the location of a single leak needs to be identified, the model class M^ℓ considers all hydraulic models that present leakage at the ℓ-th location, defined through some parameters $\boldsymbol{\theta}^\ell$ (e.g. leakage

intensity). Additionally, the set of all probabilistic model class $\mathbf{M}$ is

$$\mathbf{M} = \left\{ M^{\ell} \,:\, \ell = 1, \ldots, N_{\text{class}} \right\} \tag{5.2.1}$$

where a *prior probability* $P\left[M^{\ell}|\mathbf{M}\right]$ is assigned to each probabilistic model class, such that $\sum_{\ell=1}^{N_{\text{class}}} P\left[M^{\ell}|\mathbf{M}\right] = 1$. It is important to remark that, in this setting, probability is interpreted on the idea of reasonable expectation. Formally, probability is a measure of the plausibility of a given hypothesis under available information [50]. Then, $P\left[M^{\ell}|\mathbf{M}\right]$ represents how plausible is each probabilistic model class in the absence of data D (prior state of knowledge). The usual situation involves uninformative prior knowledge about the plausibility of each probabilistic model class, which gives $P\left[M^{\ell}|\mathbf{M}\right] = 1/N_{\text{class}}$ according to the indifference principle.

The explicit incorporation of data D into the analysis allows to obtain revised information about each probabilistic model class. In particular, it is possible to obtain the *posterior* probability $P\left[M^{\ell}|\mathbf{M}, D\right]$ of each probabilistic model class, which is a measure of the plausibility of the model class when data D is taken into account. The computation of these posterior probabilities is known as the *model class selection problem*. In this setting, the model class with the highest posterior probability is interpreted as the solution of the detection-related problem, that is, the actual network condition. The posterior probabilities are given by

$$P\left[M^{\ell}|\mathbf{M}, D\right] = \frac{P\left[D|M^{\ell}\right] P\left[M^{\ell}|\mathbf{M}\right]}{\sum_{\iota=1}^{N_{\text{class}}} P\left[D|M^{\iota}\right] P\left[M^{\iota}|\mathbf{M}\right]}, \qquad \ell = 1, \ldots, N_{\text{class}} \tag{5.2.2}$$

where $P\left[M^{\ell}|\mathbf{M}\right]$ is already defined, and $P\left[D|M^{\ell}\right]$ is the evidence of model class M^{ℓ}. The last quantity is a probabilistic measure of how likely is obtaining data D when the ℓ-th model class is considered. The computation of the evidence is directly related to the Bayesian model updating problem.

5.3 Bayesian model updating

The Bayesian framework allows to obtain revised information about the system, based on available data D. In this setting, the updated joint probability density function $p\left(\boldsymbol{\theta}^{\ell}|M^{\ell}, D\right)$ can be obtained (posterior probability density function). This distribution quantifies the plausibility of the model parameters $\boldsymbol{\theta}^{\ell}$ when model class M^{ℓ} and data D

are considered. It is important to remark that the probabilistic description of the network state parameters $\boldsymbol{\theta}^\ell$ is used to represent the uncertainty associated to the modeling of the unknown network condition, that is, epistemic uncertainty. According to Bayes' Theorem, the posterior distribution is given by

$$p\left(\boldsymbol{\theta}^\ell | M^\ell, D\right) = \frac{p\left(D|M^\ell, \boldsymbol{\theta}^\ell\right) p\left(\boldsymbol{\theta}^\ell | M^\ell\right)}{P\left[D|M^\ell\right]} \tag{5.3.1}$$

where $p\left(D|M^\ell, \boldsymbol{\theta}^\ell\right)$ is the likelihood function based on the predictive distribution for the network response of model $M^\ell(\boldsymbol{\theta}^\ell)$, the quantity $P\left[D|M^\ell\right] = \int p\left(D|M^\ell, \boldsymbol{\theta}^\ell\right) p\left(\boldsymbol{\theta}^\ell|M^\ell\right) d\boldsymbol{\theta}^\ell$ is the evidence of model class M^ℓ, and $p\left(\boldsymbol{\theta}^\ell|M^\ell\right)$ is the prior probability density function selected for the probabilistic model class M^ℓ. The last information is used to quantify the initial plausibility of each predictive model defined by the value of the parameters $\boldsymbol{\theta}^\ell$, allowing in this manner prior information to be incorporated. In most cases a uniform distribution is considered, since only the feasible range of each parameter is usually known. On the other hand, the likelihood function considered in this work is described in the following section.

5.3.1 Likelihood function

The likelihood function $p\left(D|M^\ell, \boldsymbol{\theta}^\ell\right)$ is a measure of how plausible is to obtain measurement data D from hydraulic model $M^\ell(\boldsymbol{\theta}^\ell)$. In what follows, it is assumed that D contains N_F flow tests performed at N_L monitoring locations. Each flow test corresponds to a single realization of the measurements, including N_Q quantities of interest for each location. The corresponding flow data is contained in the vector $\mathbf{y}^* \in \mathbb{R}^{N_{\text{data}}}$, where $N_{\text{data}} = N_F \times N_L \times N_Q$. For example, the flow data set may consist of pressure heads, flow rates, or concentration rates obtained at the N_L monitoring locations. Let also $\mathbf{y}(\boldsymbol{\theta}^\ell) \in \mathbb{R}^{N_{\text{data}}}$ be a vector containing all the corresponding flow quantities at the monitoring locations, computed from model $M^\ell(\boldsymbol{\theta}^\ell)$. The vector $\mathbf{e}(\boldsymbol{\theta}^\ell) \in \mathbb{R}^{N_{\text{data}}}$ comprising all the prediction errors is defined as

$$\mathbf{e}(\boldsymbol{\theta}^\ell) = \mathbf{y}^* - \mathbf{y}(\boldsymbol{\theta}^\ell) \tag{5.3.2}$$

Note that the prediction errors are defined as differences between model predictions and sensor measurement data, which are due to both flow network modeling errors and measurement devices accuracy. Such issues are unavoidable in the hydraulic modeling and data collection processes of real-life water distribution networks.

This work considers a deterministic mapping between the network state parameters and the flow predictions, $\boldsymbol{\theta} \rightarrow \mathbf{y}(\boldsymbol{\theta}^\ell)$. In this work, the prediction errors are modeled as normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}$. The likelihood function $p\left(D|M^\ell, \boldsymbol{\theta}^\ell\right)$ is then written as

$$p\left(D|M^\ell, \boldsymbol{\theta}^\ell\right) = \det\left(2\pi\boldsymbol{\Sigma}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} J\left(\boldsymbol{\theta}^\ell, \mathbf{y}^*\right)\right] \tag{5.3.3}$$

where $J\left(\boldsymbol{\theta}^\ell, \mathbf{y}^*\right)$ is a weighted measure of fit between the model prediction and the flow data, given by

$$J\left(\boldsymbol{\theta}^\ell, \mathbf{y}^*\right) = \left[\mathbf{y}^* - \mathbf{y}(\boldsymbol{\theta}^\ell)\right]^T \boldsymbol{\Sigma}^{-1} \left[\mathbf{y}^* - \mathbf{y}(\boldsymbol{\theta}^\ell)\right] \tag{5.3.4}$$

In the actual implementation, the prediction errors are assumed independent and, therefore, the covariance matrix $\boldsymbol{\Sigma}$ is a diagonal matrix comprising the prediction error variances. Note that this assumption implies that prediction errors, from different locations and different flow tests, are statistically independent. Nevertheless, different prediction error model classes can be used as well, including models that consider correlation between prediction errors [71, 72]. The effect of such considerations into the effectiveness of the proposed method is left for future research efforts.

Finally, it is also possible to consider stochastic hydraulic models [73, 74] within the proposed approach, allowing to explicitly include irreducible uncertainties (e.g., stochastic demand of users) into the detection process. In this scenario, the mapping $\boldsymbol{\theta} \rightarrow \mathbf{y}(\boldsymbol{\theta}^\ell)$ for model $M^\ell(\boldsymbol{\theta}^\ell)$ is probabilistic and flow predictions $\mathbf{y}(\boldsymbol{\theta}^\ell)$ follow a multivariate probability density function $p(\mathbf{y}(\boldsymbol{\theta}^\ell))$. This means that the likelihood function $p\left(D|M^\ell, \boldsymbol{\theta}^\ell\right)$ can be obtained, in principle, by simulation. In fact, current developments in this area [75] can be integrated in the proposed framework to address detection-related problems involving stochastic hydraulic models. Such considerations are left for future research efforts.

5.3.2 Simulation-based approach

For globally identifiable cases, that is, when the set of most probable model parameters is a singleton, asymptotic approximations of the Bayesian predictive integrals have been used in a number of applications with sufficient accuracy [44, 51, 52]. In this case, the posterior distribution of the model parameters $\boldsymbol{\theta}^\ell$ is very peaked and is asymptotically approximated by a multi-dimensional Gaussian distribution centred at the most probable value of the model parameters. In this setting, the most probable value for the model parameters are the ones that maximize the posterior probability density function in Equa-

tion (5.3.1). The corresponding evidence can be estimated based on this value for the model parameters.

The cases previously described usually arise when large amount of data and a relatively accurate modeling of the system under study are available. However, usual situations in real-life utility networks lead to cases where the posterior probability density function is not very peaked, flat or multimodal. In those cases, the validity of using asymptotic approximations is doubtful. To avoid these difficulties and to treat potentially ill-posed detection-problems, a simulation-based Bayesian model updating method is adopted here as a tool for obtaining a posterior sample and estimating the corresponding evidence. Specifically, an advanced simulation technique called transitional Markov chain Monte Carlo is implemented [4, 54].

5.4 Advanced simulation technique

The transitional Markov chain Monte Carlo algorithm (TMCMC) is a multilevel simulation approach introduced in [4] to address several problems in the context of Bayesian model updating. A number of applications have shown the effectiveness and generality of this technique in Bayesian model updating, model class selection, model averaging, etc [4, 76–78]. A thorough description of the algorithm, including a pseudo-code with the actual implementation, is provided in Appendix C. For completeness, the main ideas and key features of the technique are described in the next subsections.

5.4.1 Basic ideas of transitional Markov chain Monte Carlo

The main idea of the algorithm is to sample from a sequence of intermediate non-normalized distributions $\{p_j(\boldsymbol{\theta}^\ell) : j = 0, \ldots, m\}$ that converges to the posterior distribution. Such sequence is defined as

$$p_j(\boldsymbol{\theta}^\ell) \propto p(\boldsymbol{\theta}^\ell|M^\ell)p(D|M^\ell, \boldsymbol{\theta}^\ell)^{\alpha_j}, \qquad j = 0, \ldots, m \tag{5.4.1}$$

where $0 = \alpha_0 < \cdots < \alpha_m = 1$, and $\propto$ means proportional. The method starts sampling from the prior distribution ($j = 0$) and ends sampling from the posterior distribution ($j = m$). The α_j-values are selected such that subsequent distributions present a similar

shape. As a result, the algorithm generates $m+1$ sets of samples

$$\{\boldsymbol{\theta}_{jk}^{\ell} \,:\, k=1,\ldots,N_j\}, \qquad j=0,\ldots,m \tag{5.4.2}$$

where $\boldsymbol{\theta}_{jk}^{\ell}$ is the k-th sample obtained at stage j, and N_j is the total number of samples generated at stage j. Markov chain Monte Carlo [66, 67] is applied at stages $j=1,\ldots,m$ to obtain the required sample.

5.4.2 Evidence estimate

One of the most relevant features of the algorithm, in the proposed framework, is the estimation of the evidence $P[D|M^{\ell}] \approx \widehat{S}_{\ell}$ as [4]

$$P\left[D|M^{\ell}\right] \approx \widehat{S}_{\ell} = \prod_{j=0}^{m-1}\left[\frac{1}{N_j}\sum_{k=1}^{N_j} p(D|M^{\ell},\boldsymbol{\theta}_{jk}^{\ell})^{\alpha_{j+1}-\alpha_j}\right] \tag{5.4.3}$$

where $\{p(D|M^{\ell},\boldsymbol{\theta}_{jk}^{\ell}) : k=1,\ldots,N_j\}$ are the likelihood values already obtained at stage j.

5.4.3 Posterior sampling

The method not only indicates the most plausible model class, but also provides revised information about the corresponding model parameters in terms of a sample from the posterior distribution. Besides, the simulation technique can handle globally identifiable as well as strictly unidentifiable systems. In this way, the approach is more general than traditional techniques that try to find one single value for the model parameters.

5.4.4 Implementation issues

High performance computing techniques at the computer hardware level can be considered to enhance the computational efficiency of the proposed approach. In fact, the transitional Markov chain Monte Carlo method contains a large number of chains which are perfectly parallel [4, 68]. Thus, a number of computer workers can handle the generation of samples corresponding to the different chains. The load balance in the computer workers can be based, in principle, on a static or dynamic job-scheduling scheme [68, 69]. In addition, parallelization can also be used at the model class level. Actually, the probabilistic model classes are independent from each other, which means that the estimation of the evidence of the different model classes can be performed concurrently. Therefore, these analyses can be carried out simultaneously taking advantage of available parallelization techniques.

Chapter 6

APPLICATION 2: LEAKAGE DETECTION

Leakage involves hydraulic resource escaping from a water distribution system in an uncontrolled manner. Such a phenomenon constitutes one of the most important problems in water utility networks. This chapter demonstrates the application of the proposed Bayesian approach to the study of a leakage detection problem, involving a real-life water distribution system. Two scenarios in terms of the leakage location are taken into account. The effects of sensors configuration, modeling errors and measurement noise into the quality of the detection process are investigated. The results show the effectiveness of the proposed approach, as well as its robustness to model and measurement errors. Moreover, it is shown that the approach is able to identify thresholds for these errors beyond which no reliable identification of the leakage location is possible.

6.1 Problem formulation

6.1.1 Description of the network

The hydraulic network analysed in this chapter corresponds to the water distribution system of Concepción, a city of 250000 inhabitants located in the south of Chile. The layout of the EPANET model is shown in Figure 6.1. The North-South dimension is about 12.2 [km], whereas the East-West dimension is about 8.7 [km]. The hydraulic model

consists of 10158 nodes, 9896 pipes and 20 storage tanks. In addition, the distribution system has 1506 valves and 8 impulse pumps. In the context of this application example, steady-state conditions are considered. Appendix D presents a more detailed description of the steady-state solution of water distribution networks in the context of EPANET.

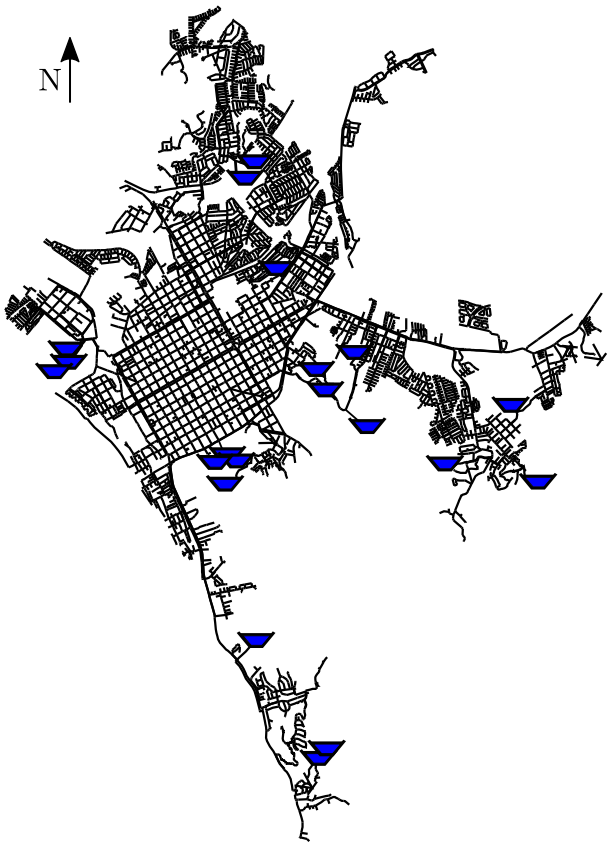


Figure 6.1: Water distribution system of Concepción

The total demand of the network is 1690 [l/s], which is distributed over 5209 nodes (i.e. 4949 nodes have no demand). The corresponding nominal demands of the nodes are shown in Figure 6.2. These conditions represent a standard scenario in terms of the users requirements. The minimum and maximum nodal demands are 1.0×10^{-3} [l/s] and 35.0 [l/s], respectively. The water distribution system comprises 504 [km] of pipelines, modeled with 9896 pipes. The lengths of the pipes, arranged in an increasing manner, are presented in Figure 6.3. Note that the minimum and maximum lengths are 0.3 [m] and 1437.5 [m], respectively. Pipes of five different materials are used in the network, namely, high density polyethylene (HDPE), polyvinyl chloride (PVC), asbestos cement (AC), steel, and cast iron. The corresponding nominal roughness coefficients (Hazen-Williams coefficients) of

the different materials are 140, 140, 130, 100 and 80, respectively. Information about the number and total length of pipes of the different materials is provided in Table 6.1.

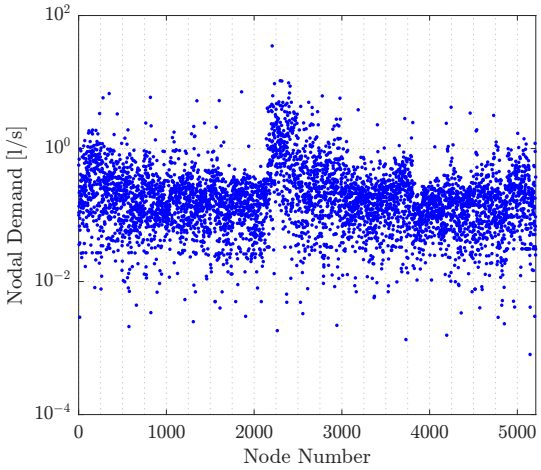


Figure 6.2: Nodal demands in hydraulic model of Concepción

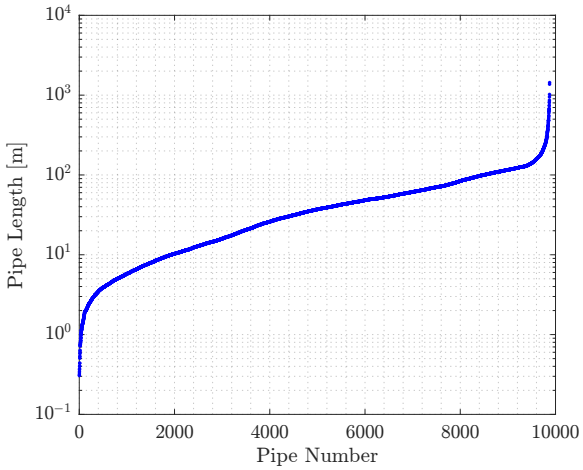


Figure 6.3: Lengths of pipes in hydraulic model of Concepción

Table 6.1: Number of pipes and total length of different materials. Application example 2

Material	Number of pipes	Total length [km]
High density polyethylene (HDPE)	237	11.8
Polyvinyl chloride (PVC)	5123	238.7
Asbestos cement (AC)	2809	148.4
Steel	178	20.7
Cast iron	1549	84.4

6.1.2 Definition of sectors with leakage

For illustration purposes, two scenarios in terms of the leakage location are considered. In the first case, leakage is assumed to occur in a peripheral sector of the network. The corresponding sector is identified in Figure 6.4 as Leakage Sector 1. In the second case, leakage is assumed to be located in a central sector of the network, identified in Figure 6.4 as Leakage Sector 2. Note that both sectors are quite different regarding their interaction with the rest of the network: the central position of Leakage Sector 2 leads to a more complex influence of the rest of the network, in comparison lo Leakage Sector 1.

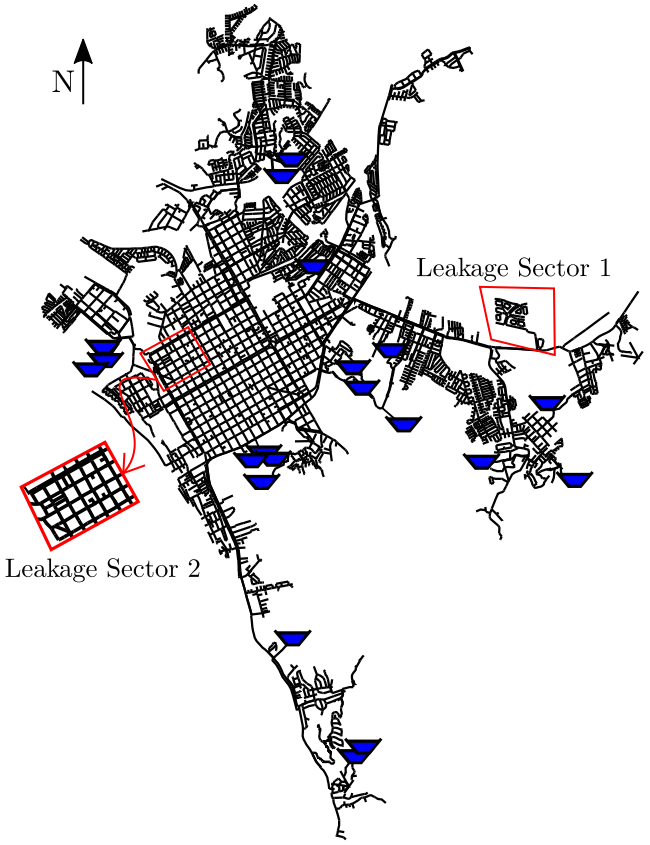


Figure 6.4: Leakage sectors in hydraulic network of Concepción

Figure 6.5 (left) shows the layout of Leakage Sector 1 in more detail, which contains 95 nodes and 100 pipes. On the other hand, Figure 6.5 (right) presents the layout of Leakage Sector 2, consisting of 45 nodes and 50 pipes. Both sectors are different in terms of their topology: Leakage Sector 1 shows a branched topology, whereas Leakage Sector 2 presents a looped topology. Note that other network sectors can be considered in this framework as well.

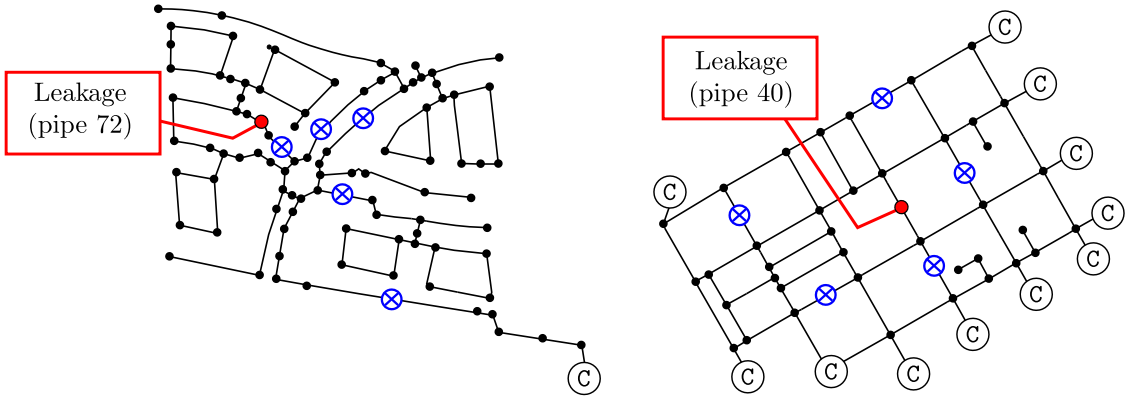


Figure 6.5: Layout of Leakage Sector 1 (left) and Leakage Sector 2 (right). C: connection nodes to the rest of the network. ⊗: location of flow meters

In terms of leakage modeling, this work considers a demand-driven approach. The leakage location is simulated by adding a node at the midpoint of the damaged pipe. The prescribed flow demand in that node corresponds to the amount of leakage (leakage intensity). Specifically, it is assumed that leakage in Sector 1 occurs in pipe number 72, while pipe number 40 is the one that leaks in Sector 2 (see Figure 6.5). The leakage intensity considered for each case is equal to 5% of the total demand of the sector. These quantities correspond to realistic scenarios in engineering practice. The corresponding numerical values are presented in Table 6.2. Note that the leakage intensities in Sector 1 and Sector 2 correspond to 0.020% and 0.065% of the total demand of the network, respectively. Validation calculations have shown that leakage location cannot be identified correctly for leakage severity smaller than 3% of the water volume supplied in each sector.

Table 6.2: Definition of leakage in each sector

	Total demand [l/s]	Leakage location	Leakage intensity (Q_{leak}) [l/s]
Leakage Sector 1	8.1	Pipe 72	0.405
Leakage Sector 2	22.0	Pipe 40	1.100

It is important to remark that the scenarios considered in this work, which assume that leakage is already known to be located in some specific sector of the network, are quite realistic. In fact, current methodologies allow to determine, based on available field data, if there are water losses in a certain sector of the network [79,80]. Such state of knowledge is usually referred to as *leakage awareness* [81]. Thus, the objective of this application

example is to determine the specific position, at the level of hydraulic model components, where leakage is located in a specific sector. In other words, the proposed approach is used in this chapter to locate leaks in a sector that has already been identified to contain leaks.

6.1.3 Simulated data

The data D considered in the detection process consists of simulated measurements. These measurements are obtained from a prescribed hydraulic model that is referred to as *actual network*. In order to consider more realistic conditions, this work explicitly takes into account both measurement errors and modeling errors. In this setting, measurement errors are related to differences between the flow measurements used for the detection process and flow predictions from the actual network, whereas modeling errors correspond to differences between the *actual network* and the class of hydraulic models used for the detection process, i.e. the hydraulic networks considered in the set of probabilistic model classes $\mathbf{M}$.

In the context of this chapter, only one quantity of interest is obtained at each location ($N_Q = 1$). Then, simulated data is defined as

$$\mathbf{y}^* = \left\langle \mathbf{y}_1^{*T}, \ldots, \mathbf{y}_{N_F}^{*T} \right\rangle^T \tag{6.1.1}$$

where $\mathbf{y}_i^* \in \mathbb{R}^{N_L}, i = 1, \ldots, N_F$ is a vector containing the measurements obtained at the N_L locations for the i-th flow test, defined by

$$\mathbf{y}_i^* = \left\langle y_{i,1}^*, \ldots, y_{i,N_L}^* \right\rangle^T \tag{6.1.2}$$

and $y_{i,j}^*, j = 1, \ldots, N_L$ is the measurement obtained at the j-th measurement during the i-th flow test.

Following some of the ideas presented in [44], the simulated measurements obtained for each flow test $\mathbf{y}_i^*$ are generated as

$$y_{ij}^* = y_{ij}^{\text{model}} + y_{ij}^{\text{noise}} \tag{6.1.3}$$

where y_{ij}^{model} is the flow prediction obtained from flow test i at location j in the actual network, and y_{ij}^{noise} accounts for the measurement error that comes from the corresponding sensor.

The characteristics of the hydraulic model considered to obtain y_{ij}^{model} deviate from the nominal characteristics defining the class of models used in the identification process. Specifically, the characteristics that are perturbed from their nominal values are the pipe roughness coefficients and the nodal demands. The variation from the nominal values of the previous characteristics at each pipe and node is assumed to follow a zero mean uniform distribution with bounds $(-\alpha, \alpha)$ and $(-\beta, \beta)$, respectively. These values represent the magnitude of the model errors expressed as a percentage of the nominal values of the network properties. The perturbed characteristics define the *actual network*, which is assumed to be representative of the actual behaviour of the water distribution system. Then, it is clear that the class of probabilistic model classes considered for the detection process is not capable to reproduce exactly the behaviour of the actual network.

The term y_{ij}^{noise} is assumed to be representative of the magnitude of measurement errors obtained in flow test i at location j. In the context of this work, y_{ij}^{noise} are assumed to be independent and uniformly distributed random variables with bound $(-\gamma, \gamma)$. The magnitude of γ represents the size of measurement error expressed as a percentage of the actual network predictions, such that $|y_{ij}^{\text{noise}}| \leq \gamma |y_{ij}^{\text{model}}|$.

As already discussed in the previous chapter, several quantities of interest can be considered to carry out the detection process. In the context of leakage detection, it is possible to consider pressure measurements or flow measurements. Nevertheless, preliminary results showed that pressure measurements were not very informative to carry out the detection process in this large-scale water distribution network. Therefore, the flow data set considered here consists of flow rates in the pipes. Specifically, five measurement locations in each leakage sector are considered, as presented in Figure 6.5. It can be seen that these monitoring devices (flow meters) are spread over important sections of each leakage sector and they are relatively far from the vicinity of the leaked pipe (pipe 72 in Leakage Sector 1 and pipe 40 in Leakage Sector 2).

6.1.4 Probabilistic model classes

In this application example, each model class M^ℓ considers a node at the mid-point of a pipe in the section under consideration, where the corresponding nodal demand $\boldsymbol{\theta}^\ell = \theta^\ell$ is the leakage intensity (demand-driven approach). In other words, the hydraulic model $M^\ell(\theta^\ell)$ considers a node at the midpoint of the ℓ-th pipe with prescribed demand θ^ℓ.

Then, each probabilistic model considers one uncertain parameter, interpreted as the leakage intensity at the ℓ-th location. In the context of this chapter, the model class $M^\ell, \ell = 1, \ldots, N_{\text{class}}$ is called the ℓ-th *leakage event*.

The set of probabilistic model classes $\mathbf{M}$ considers all pipes within the sector where leakage is assumed to be. Thus, the total number of probabilistic model classes coincides with the total number of pipes in the sector, since only one leakage location is assumed to happen. In a more general setting, if the sector under consideration consists of P pipe sections where leakage can occur and there are K leakage locations (pipe sections that simultaneously present leakage), the total number of leakage events is $N_{\text{class}} = P!/K!(P-K)!$. Besides, each model class is parametrized by a vector $\boldsymbol{\theta}^\ell \in \Theta^\ell \subset \mathbb{R}^K$ that specifies the amount of leakage outflow at the corresponding K leakage locations of leakage event ℓ. In this work, only one leakage location is considered, that is $K = 1$ and $N_{\text{class}} = P$. This is a realistic assumption, since the damage in pipes is expected to occur progressively. Then, Leakage Sector 1 gives $N_{\text{class}} = 100$ and Leakage Sector 2 gives $N_{\text{class}} = 50$. Model class M^{72} corresponds to the actual leakage location in Sector 1, whereas leakage in Sector 2 corresponds to M^{40}.

The likelihood function of each probabilistic model class was described in Section 5.3.1. On the other hand, the prior distribution for each probabilistic model class is uniform, ranging from zero to three times the actual leakage intensity (Q_{leak}). This upper limit is chosen to consider a rather wide support region for leakage intensity values. In fact, preliminary calculations show that any other value for the upper limit can be selected without affecting the performance of the detection procedure, as long as the support region remains sufficiently large to contain the actual amount of leakage.

$$p\left(\theta^\ell | M^\ell\right) = \begin{cases} \dfrac{1}{3Q_{\text{leak}}}, & 0 \leq \theta^\ell \leq 3Q_{\text{leak}} \\ 0, & \text{otherwise} \end{cases}, \quad \ell = 1, \ldots, N_{\text{class}} \tag{6.1.4}$$

6.1.5 Implementation details

The Bayesian model updating procedure was performed, for each probabilistic model class, by using the transitional Markov chain Monte Carlo method with 1000 samples per stage in both sectors. On the other hand, the covariance matrix in equations (5.3.3) and (5.3.4) is taken as $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and σ^2 is a positive number.

6.2 Idealized scenario

An idealized scenario with $\alpha = \beta = \gamma = 0$ and only one flow test is considered here. Then, the nominal characteristics of the hydraulic models used for the identification procedure coincide with the ones of the actual network. Besides, the flow predictions from the actual network are directly used for the identification process (there are no measurement errors). The corresponding results for both sectors are discussed in what follows.

6.2.1 Leakage Sector 1

As already pointed out, a total of 100 probabilistic model classes are considered for this sector. The normalized evidences of all model classes are presented in Figure 6.7. Since all probabilistic model classes have the same prior probability, the most probable model class is the one with the highest evidence. It is observed that M^{72} is the most probable leakage event, which coincides with the actual location of the leakage (see Figure 6.5). Moreover, the evidence values corresponding to undamaged pipes are negligible. Thus, the proposed approach is successful in detecting the actual location of leakage when no modeling or measurement errors are introduced. Recall that the leakage intensity corresponds to 0.02% of the total water volume supplied in the network (see Table 6.2).

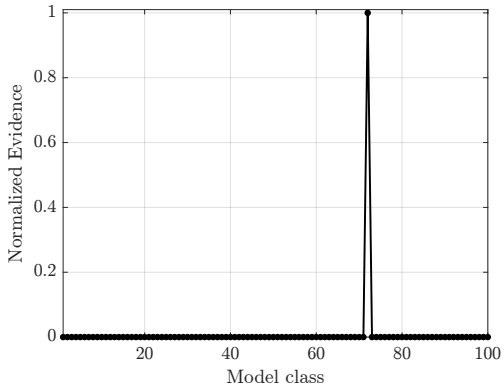


Figure 6.6: Normalized evidences of all model classes. Leakage Sector 1. Idealized scenario

The simulation-based approach formulated in this work allows to obtain a sample from the posterior distribution. This allows to get an insight into the identification process and the behavior of the system. Figure 6.7 shows the histograms of the samples of the model parameter θ^{72} (leakage intensity) during the different stages of the transitional Markov chain Monte Carlo method. Note that six simulation steps are required to obtain a

sample from the posterior distribution and the evidence estimate. The posterior histogram (stage 5) shows that the samples of this model parameter are distributed around the actual value $\theta^{72} = 0.405$. Moreover, the mean of the posterior sample is equal to $\bar{\theta}^{72} = 0.404$. In other words, this model class is able to identify the leakage intensity.

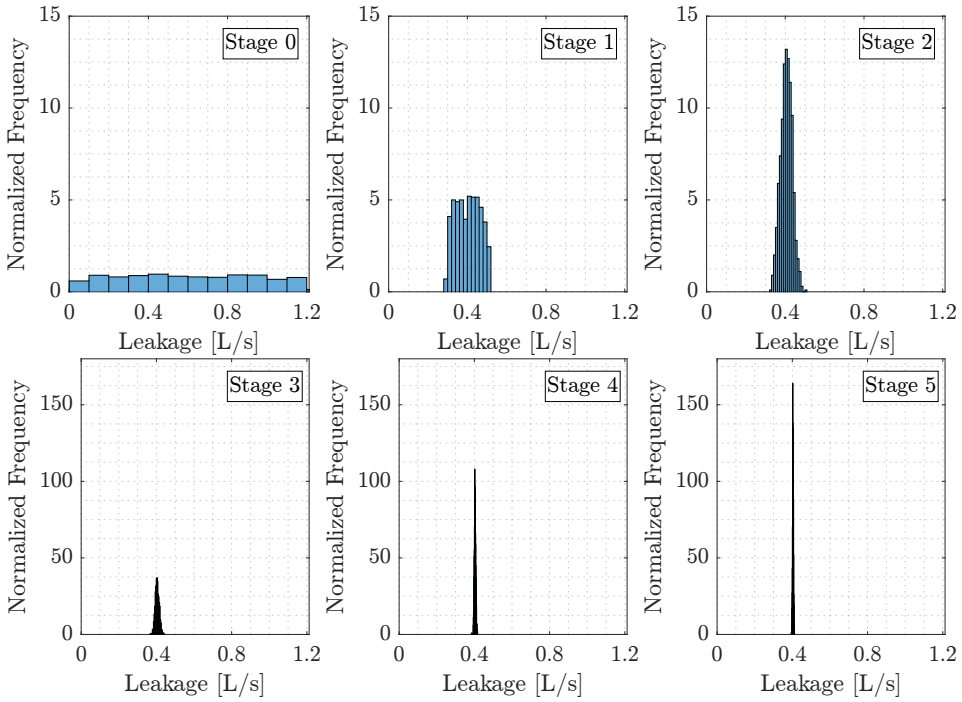


Figure 6.7: Histograms of samples obtained at each stage of TMCMC. Leakage sector 1

6.2.2 Leakage Sector 2

The detection procedure in Leakage Sector 2 considers 50 model classes. The normalized evidences of all model classes are presented in Figure 6.9. These results are analogous to the ones observed in Leakage Sector 1: the proposed method identifies correctly the leakage location (pipe number 40, see Figure 6.5). Recall that, in this case, the leakage intensity corresponds to 0.065% of the total water volume supplied in the network (see Table 6.2). Figure 6.9 shows the normalized histograms of model parameter θ^{40} (leakage intensity) obtained at different stages of the transitional Markov chain Monte Carlo method. The posterior histogram (stage 4) is distributed around the actual value $\theta^{40} = 1.10$. Then, the method is able to identify the actual leakage intensity for this central sector of the network.

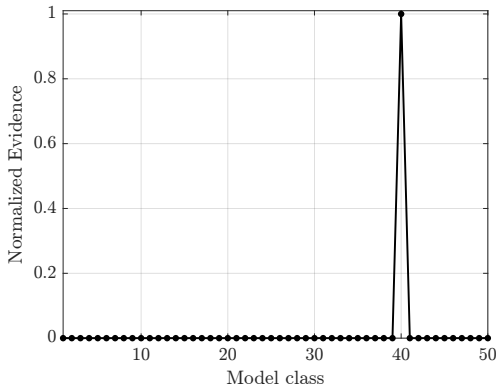


Figure 6.8: Normalized evidences of all model classes. Leakage Sector 2. Idealized scenario

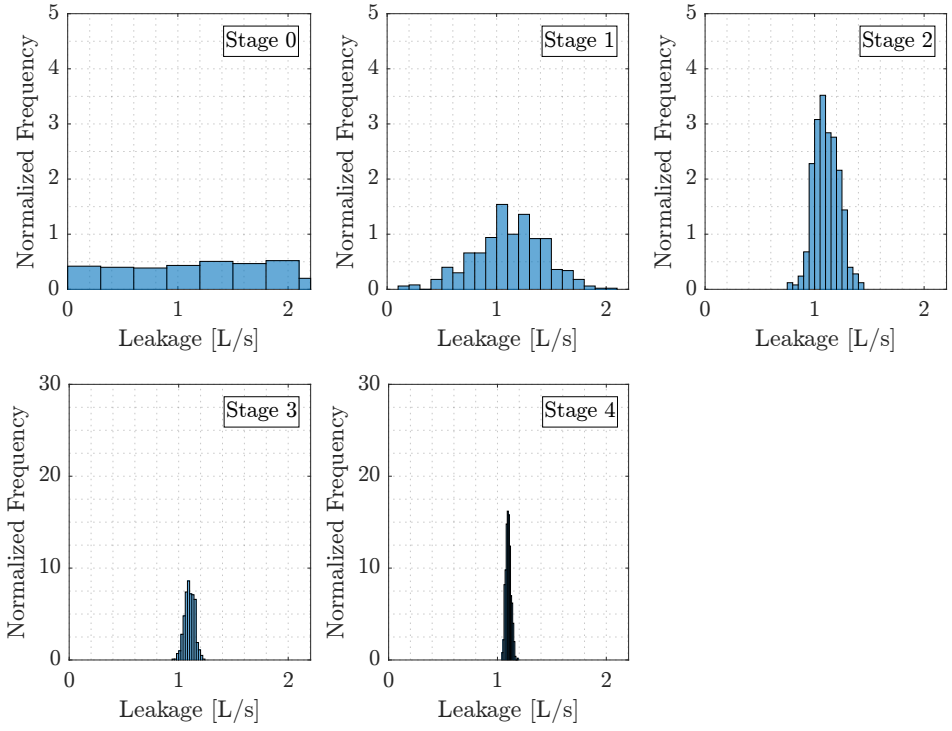


Figure 6.9: Histograms of samples obtained at each stage of TMCMC. Leakage sector 2

6.3 Effect of modeling errors: roughness coefficients

This section presents the results of the detection process when considering errors only in the roughness coefficients, that is, $\beta = \gamma = 0$ and $\alpha \neq 0$. Specifically, two levels of error intensity are considered: $\alpha = 5\%$ and $\alpha = 10\%$. As already pointed out, model errors in pipes are imposed by perturbing the values of all pipe roughness coefficients in

the hydraulic network considered to obtain the data for the identification procedure. Note that this setting gives a total of 9896 hydraulic model parameters that deviate from the nominal characteristics.

6.3.1 Leakage Sector 1

The results associated with the detection of the leakage location in Sector 1 under modeling errors in the pipe roughness coefficients are presented in Figure 6.10. A single flow test is considered for both intensity levels. Note that the maximum value of the evidence is obtained for probabilistic model class M^{72} in the cases $\alpha = 5\%$ and $\alpha = 10\%$. The evidences of the other probabilistic model classes are almost zero. This shows that the proposed approach allows to clearly identify the damaged pipe in Leakage Sector 1 under relatively large modeling errors introduced in the roughness coefficients of the actual network. In fact, validation calculations show that the detection process is successful for modeling errors as high as $\alpha = 20\%$, which are unfeasible for the type of pipelines considered in this application example.

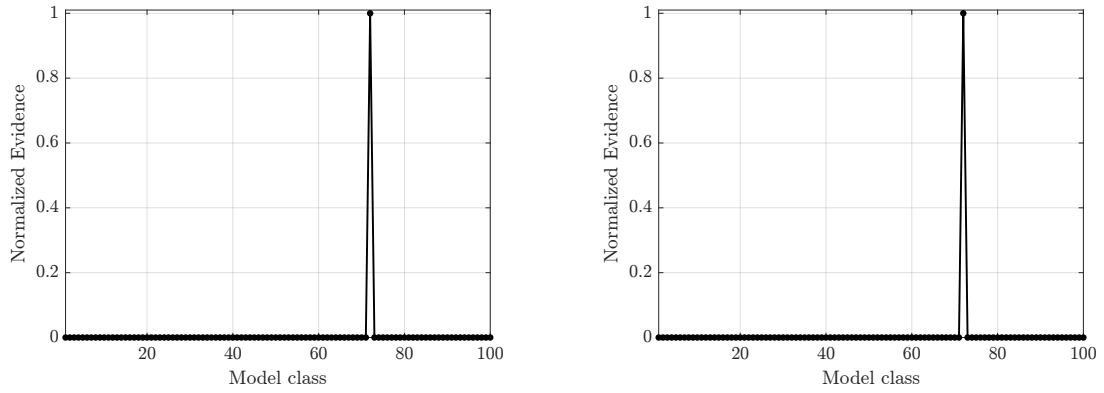


Figure 6.10: Normalized evidences. Leakage Sector 1. Modeling error of pipe roughness coefficients. Left: $\alpha = 5\%$. Right: $\alpha = 10\%$

The expected leakage intensity can be estimated as the average of the posterior samples obtained for the most probable model class, which gives 0.38 [l/s] and 0.40 [l/s] for $\alpha = 5\%$ and $\alpha = 10\%$, respectively. These values compare well to the actual amount of leakage. However, such estimates correspond to a single realization of the measurements. Clearly, the results will change if the identification process is carried out with different simulated data. In order to assess the actual bias of the estimate of the leakage intensity, independent runs of the detection process were implemented considering $\alpha = 10\%$. In other words, a

number of identification processes associated to different realizations of the measurements (i.e. data obtained from actual networks with different pipe roughness coefficients) were developed to obtain a number of posterior mean estimates. The corresponding sample average of the mean estimates in terms of the number of independent runs of the detection process is presented in Figure 6.11. It is observed that the sample average stabilizes very fast to the actual amount of leakage, $\theta^{72} = 0.404$ [l/s]. Thus, the estimate of the leak intensity is practically unbiased. Recall that only one flow test is considered to obtain these results.

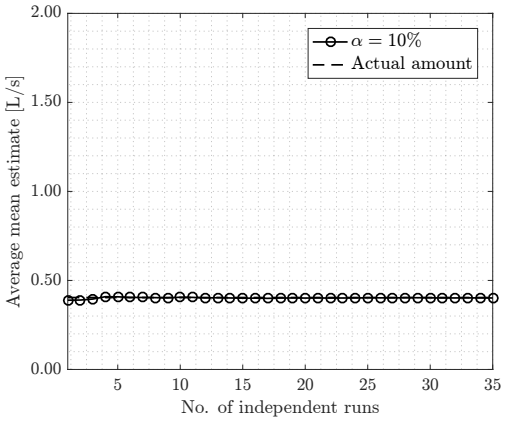


Figure 6.11: Sample average of the mean estimates of leakage intensity in terms of the number of independent detection processes. Modeling error of pipe roughness coefficients ($\alpha = 10\%$). Leakage Sector 1

6.3.2 Leakage Sector 2

Validation calculations have shown that the proposed approach is able to identify the actual leakage location in Sector 2 when considering an intermediate level of modeling errors in the pipe roughness coefficients ($\alpha = 5\%$). Nonetheless, the effectiveness of the proposed approach is decreased when a higher level of modeling errors in pipe roughness coefficients is introduced, namely $\alpha = 10\%$. As depicted in Figure 6.12 (left), the most probable model class corresponds to pipe number 28 when only one flow test is considered in the case of $\alpha = 10\%$. Thus, the proposed approach is not able to identify the correct location of the leakage event with only one flow test.

The performance of the method is improved when more information is available. Figure 6.12 (right) shows the normalized evidences of probabilistic model classes when ten flow tests ($N_F = 10$) are considered, that is, the flow rates at the specified locations correspond-

ing to ten different realizations of the pipe roughness coefficients. In this case, the proposed approach correctly identifies the actual leakage location (M^{40}). These results show that the effectiveness of the proposed approach in Leakage Sector 2 is more sensitive to the effect of modeling errors in pipe roughness coefficients than in Leakage Sector 1. This is due to the central location of Leakage Sector 2, which leads to a more complex interaction with the rest of the network. On the other hand, the sample average of the mean estimates of the amount of leakage in terms of the number of independent runs of the detection process for this case converges to a value very close to the actual amount of leakage, $\theta^{40} = 1.10$ [l/s]. Therefore, the identification process is quite robust to model errors in pipe roughness coefficients for the network and the levels of uncertainty under examination.

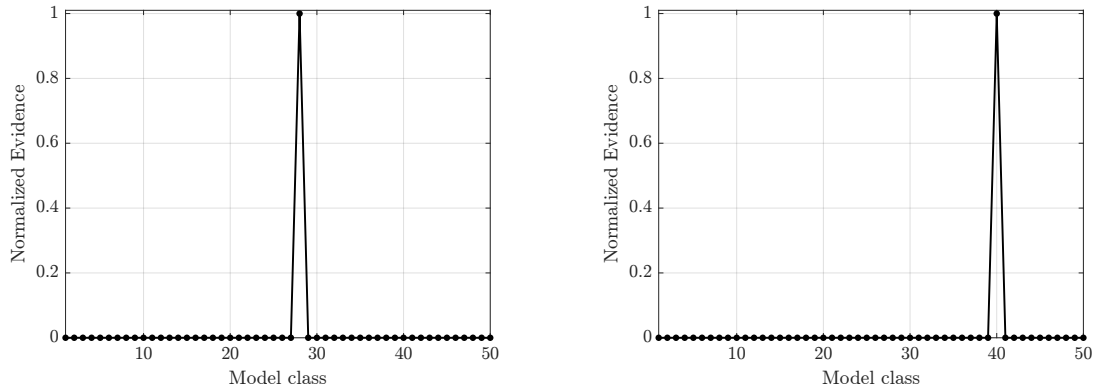


Figure 6.12: Normalized evidences. Leakage Sector 2. Modeling error of pipe roughness coefficients $\alpha = 10\%$. Left: one flow test. Right: ten flow tests

6.4 Effect of modeling errors: nodal demands

This section presents the results of the detection process when considering errors only in the nodal demands, that is, $\alpha = \gamma = 0$ and $\beta \neq 0$. Specifically, two levels of error intensity are considered: $\beta = 5\%$ and $\beta = 10\%$. In this setting, a total of 5209 hydraulic model parameters in the actual network deviate from the nominal characteristics.

6.4.1 Leakage Sector 1

When considering modeling error in the nodal demands, the results of the detection process in Leakage Sector 1 are similar to the ones observed in Section 6.3.1. Figure 6.13 shows the normalized evidences corresponding to two levels of modeling errors in the nodal

demands, namely $\beta = 5\%$ and $\beta = 10\%$. The leakage location is correctly identified in both cases. In order to study the bias of the mean estimate, Figure 6.14 shows the sample average of the mean estimates of the leakage intensity in terms of the number of independent runs of the detection process. Note that the agreement between the stabilized sample average and the actual amount of leakage is excellent. Only one flow test has been considered to obtain these results.

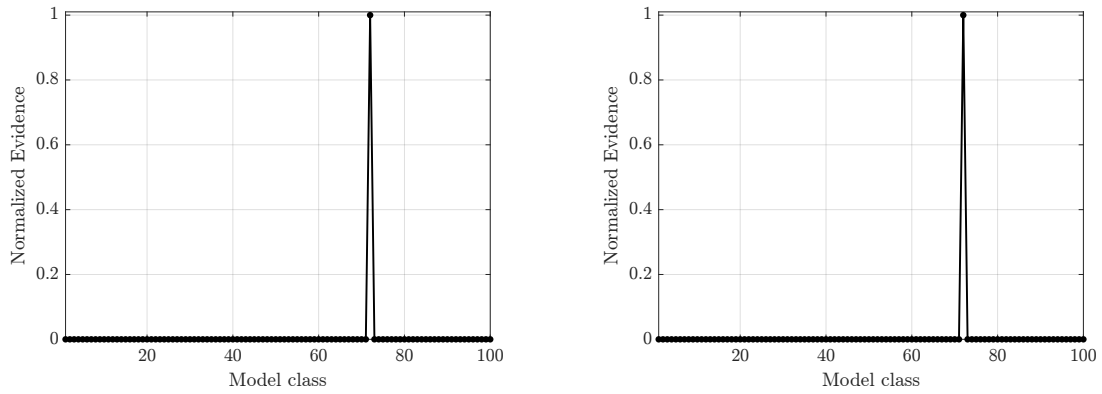


Figure 6.13: Normalized evidences. Leakage Sector 1. Modeling error of nodal demands. Left: $\beta = 5\%$. Right: $\beta = 10\%$

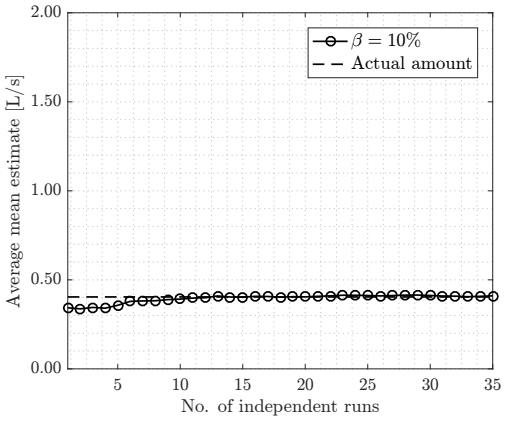


Figure 6.14: Sample average of the mean estimates of leakage intensity in terms of the number of independent detection processes. Modeling error of nodal demands ($\beta = 10\%$). Leakage Sector 1

6.4.2 Leakage Sector 2

The results obtained for Leakage Sector 2 are qualitatively similar to the ones obtained for Leakage Sector 1. The proposed approach is able to identify the location of leakage for

the levels of modeling errors in nodal demands considered in this work, namely $\beta = 5\%$ and $\beta = 10\%$. Moreover, the sample average of the mean estimates of the leakage intensity in terms the number of independent runs of the detection process stabilizes rapidly to the actual amount of leakage, $\theta^{40} = 1.10$ [l/s]. The previous results have been obtained with only one flow test. Thus, for the network under examination and within the levels of uncertainties considered, the identification process is quite robust to model errors in nodal demands.

6.5 Effect of measurement errors

This section presents the results of the detection process when considering only measurement errors, that is, $\alpha = \beta = 0$ and $\gamma \neq 0$. Recall that the measurement errors are simulated by adding a zero mean uniform noise in the data generated by the actual network. Specifically, two intensity levels are considered here: $\gamma = 2\%$ and $\gamma = 5\%$. These errors are within reasonable values considering current sensor technologies and common engineering practice. In this case, the characteristics of the actual network coincide with the nominal characteristics of the hydraulic models used for the identification procedure.

6.5.1 Leakage Sector 1

Figure 6.15 shows the normalized evidences for the cases $\gamma = 2\%$ (left) and $\gamma = 5\%$ (right) obtained when only one flow test is available ($N_F = 1$). It is observed that the method identifies correctly the leakage location for the case $\gamma = 2\%$. However, the method fails to identify the correct location with only one flow test under larger measurement errors ($\gamma = 5\%$). In fact, the maximum normalized evidence corresponds to leakage at pipe 10 in Figure 6.15 (right). The performance of the identification process improves when including more data about the behaviour of the system. Figure 6.16 shows the normalized evidence when considering 15 flow tests at each location. Note that the actual leakage location is identified as the most probable (model class M^{72}). The corresponding sample average of the mean estimates of the amount of leakage in terms of the number of independent runs of the detection process, is shown in Figure 6.17. As in the previous cases, the sample average stabilizes very fast to the actual amount of leakage, $\theta^{72} = 0.405$ [l/s], which suggest that the estimate of the amount of leakage is also unbiased for this case.

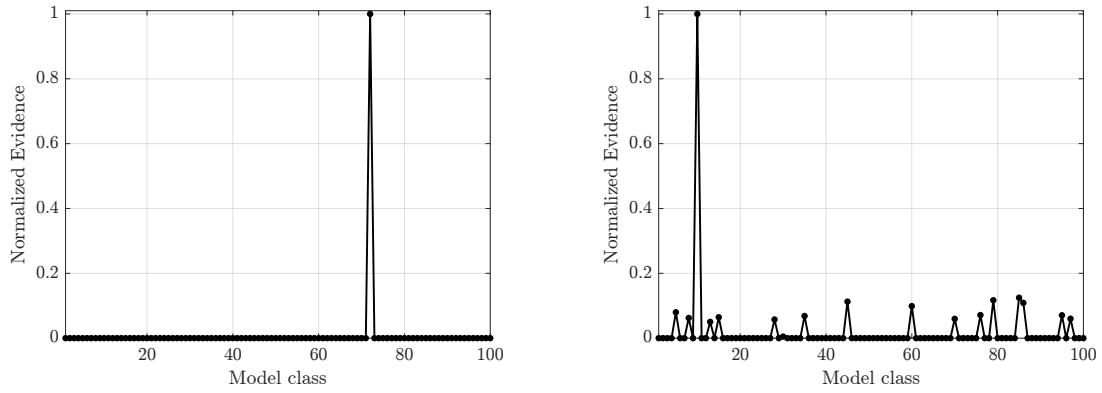


Figure 6.15: Normalized evidences. Leakage Sector 1. Measurement noise. Left: $\gamma = 2\%$. Right: $\gamma = 5\%$. One flow test at each monitoring location

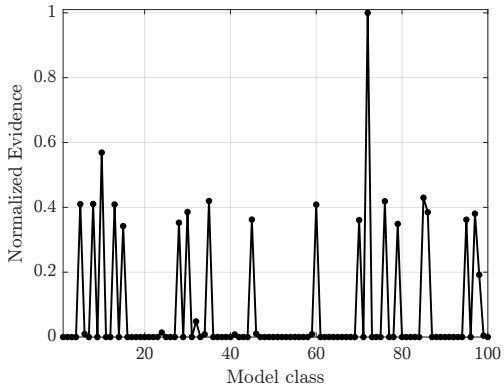


Figure 6.16: Normalized evidences. Leakage Sector 1. Measurement noise: $\gamma = 5\%$. 15 flow tests at each monitoring location

6.5.2 Leakage Sector 2

As in the previous cases, the results associated with Leakage Sector 2 are qualitatively similar to the ones obtained for Leakage Sector 1. When considering one flow test and moderate measurement errors ($\gamma = 2\%$), the normalized evidences of probabilistic model classes are practically identical to the ones observed in the idealized scenario (see Figure 6.8). On the other hand, the method requires more flow tests to detect the actual location under larger measurement errors, i.e. $\gamma = 5\%$. Figure 6.18 (left) shows the normalized evidences associated to the case $\gamma = 5\%$ and 20 flow tests at each location. In this case, pipe 47 is identified as the most probable one (incorrect leakage location). Validation calculations, which are presented in Figure 6.18 (right), show that only when 40 flow tests are included in the detection process, the most probable model class corresponds to the actual leakage event (M^{40}).

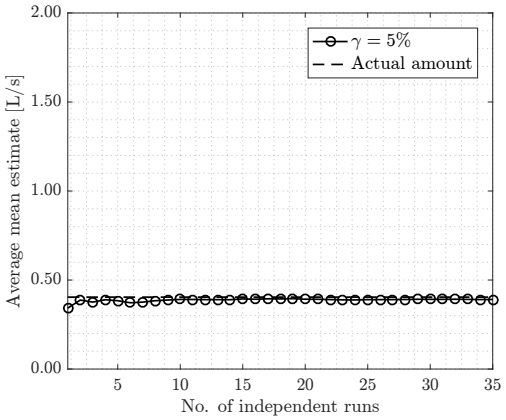


Figure 6.17: Sample average of the mean estimates of leakage intensity in terms of the number of independent detection processes. Measurement noise ($\gamma = 5\%$). 15 flow tests at each location. Leakage Sector 1

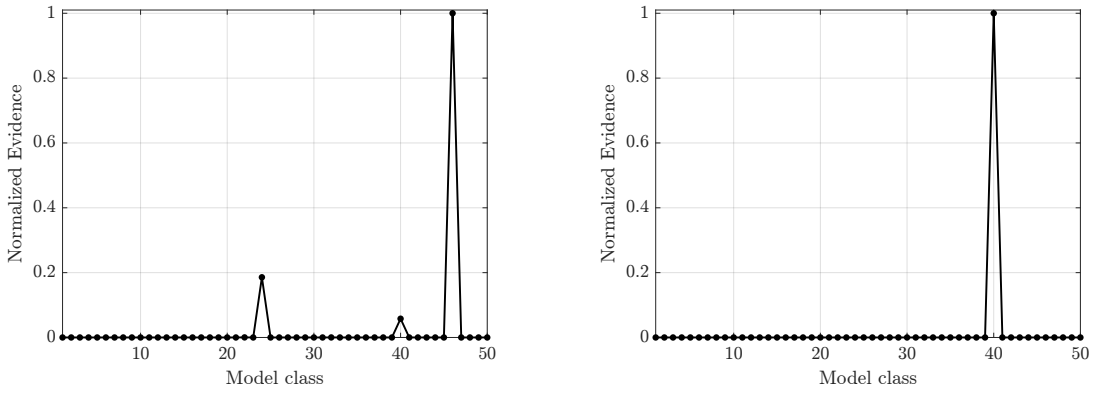


Figure 6.18: Normalized evidences. Leakage Sector 2. Measurement noise: $\gamma = 5\%$. Left: 20 flow tests at each monitoring location. Right: 40 flow tests at each monitoring location

The sample average of the mean estimates of the amount of leakage in terms of the number of independent runs of the detection process is presented in Figure 6.19. The results correspond to $\gamma = 5\%$ and 40 flow tests at each monitoring location. It is seen that the sample average converges very rapidly to the actual leakage intensity, $\theta^{40} = 1.10$ [l/s]. This suggests that, as in the previous cases, the mean estimate is unbiased.

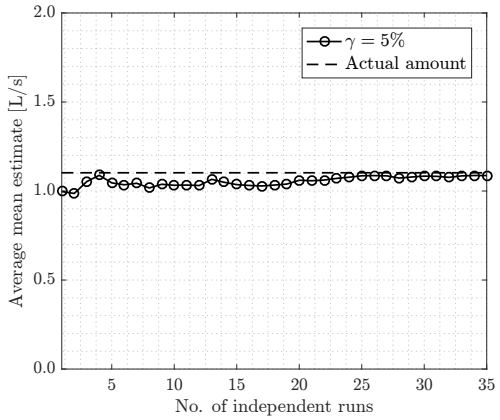


Figure 6.19: Sample average of the mean estimates of leakage intensity in terms of the number of independent detection processes. Measurement noise ($\gamma = 5\%$). 40 flow tests at each location. Leakage Sector 2

6.6 Effect of model and measurement errors

This section investigates the performance of the detection procedure when modeling and measurement errors are considered simultaneously. For illustration purposes, large measurement errors ($\gamma = 5\%$), and large modeling errors in roughness coefficients and nodal demands ($\alpha = \beta = 10\%$) are considered.

Figure 6.20 shows the normalized evidences obtained in Leakage Sector 1, corresponding to 45 flow tests. The method is able to identify the actual leakage location in this scenario, since model class M^{72} maximizes the posterior probability. Note that the posterior probability values of the other model classes are relatively larger than the ones observed when considering only measurement errors (see Figure 6.16). It is also observed that the leakage location cannot be correctly detected if less than 45 flow tests are considered. This number of flow tests represents the minimum amount of measurements beyond which no reliable identification is possible. On the other hand, the average of the mean estimates of the leakage intensity stabilizes very fast to the actual value.

Further calculations show that the identification associated to Leakage Sector 2 is qualitatively similar to the one obtained in Leakage Sector 1. That is, the minimum number of flow tests required for a correct identification is increased with respect to the previous sections.

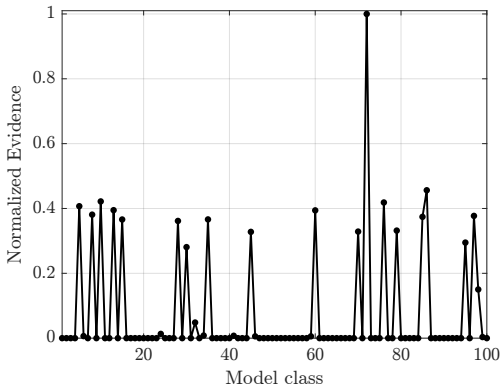


Figure 6.20: Normalized evidences. Leakage Sector 1. Error in roughness coefficients: $\alpha = 10\%$. Modeling error in nodal demands: $\beta = 10\%$. Measurement noise: $\gamma = 5\%$. 45 flow tests at each monitoring location

6.7 Effect of sensors configuration

This section investigates the effect of sensors into the effectiveness of the proposed approach. For illustration purposes, an alternative 5-sensor layout distributed closer to the leakage location in Sector 1 is considered. The effectiveness of the approach under different configurations in terms of the number of sensors is also demonstrated in this section. The results show the relevance of the design of effective monitoring strategies.

6.7.1 Alternative sensor layout

The location of sensors within the water distribution systems has a significant impact into the quality of the leakage detection procedure. In order to demonstrate this, an alternative configuration of flow meters in Leakage Sector 1 is considered. This alternative layout is presented in Figure 6.21. Note that the flow meters are now located closer to the actual location of the leakage (pipe number 72), in comparison to the configuration presented in Figure 6.5 (left).

The results obtained using the alternative layout show that the detection processes under modeling errors are similar to the ones observed with the previous sensor configuration. In other words, the new configuration is as informative as the original one when only modeling errors are considered. The location and severity of leakage are correctly identified with only one flow test in this case.

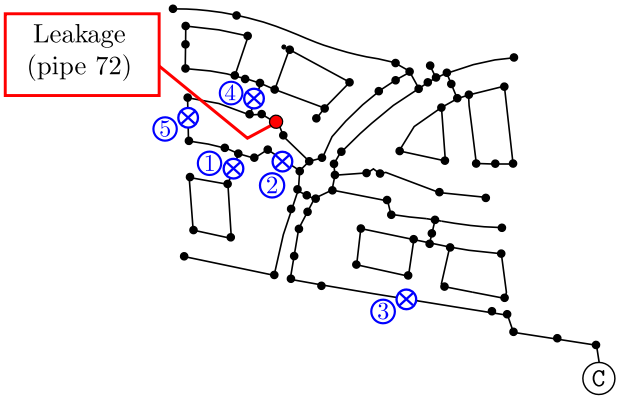


Figure 6.21: Alternative layout of flow meters. C: connection nodes to the rest of the network. ⊗: location of flowmeters

In the case of moderate measurement errors ($\gamma = 2\%$), the performance of the new layout is similar to the one observed for the original configuration (see Section 6.5.1). Under larger measurement errors ($\gamma = 5\%$), however, the alternative sensor arrangement seems to improve the detection procedure when compared to the original layout. In fact, the results presented in Figure 6.22 (left) show that model class M^{72} presents the highest evidence when only one flow test at each monitoring location is considered, although leakage event 57 also presents a relatively high evidence value. Nonetheless, the identification of the correct leakage location is significantly clearer when considering 10 flow tests at each sensor location, as depicted in Figure 6.22 (right). On the other hand, the corresponding sample average of the mean estimates in terms of the number of independent runs of the detection process is shown in Figure 6.23. Once again, the results suggest the unbiasedness of the mean estimate, since the sample average converges to the actual leakage intensity. As a final remark, the identification results obtained from the new sensor configuration are most robust in the sense that the correct leakage location can be predicted with a smaller number of flow tests at each monitoring location.

6.7.2 Number of sensors

The effect of the number of sensors on the detection performance of the proposed scheme is now addressed. For illustration purposes, five scenarios are considered in terms of the number of flow meters. The description of these scenarios is presented in Table 6.3, in connection with the layout and identification of sensors presented in Figure 6.21. The results of the detection process for each scenario are summarized in Table 6.4. These results correspond to the case of large measurement errors ($\gamma = 5\%$) and 10 flow tests at each monitoring location. Note that the leakage location cannot be correctly identified below a

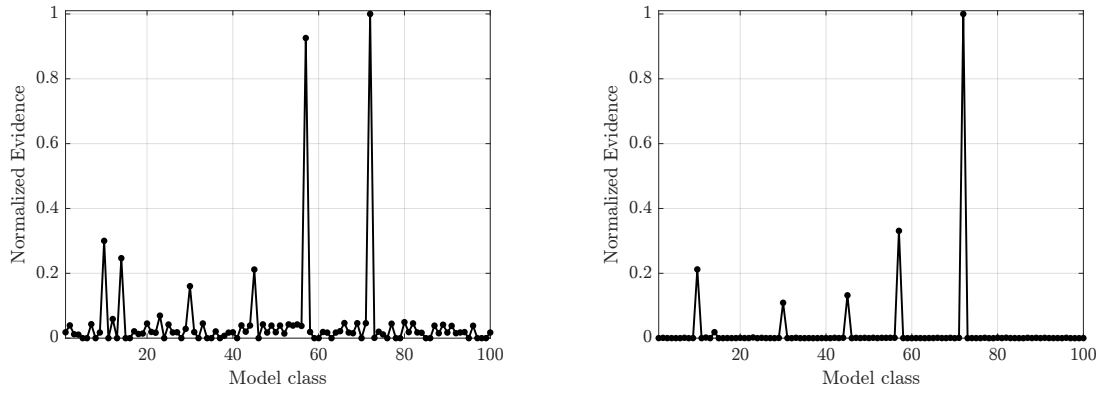


Figure 6.22: Normalized evidences. Alternative sensor layout. Measurement noise: $\gamma = 5\%$. Left: 1 flow test at each monitoring location. Right: 10 flow tests at each monitoring location

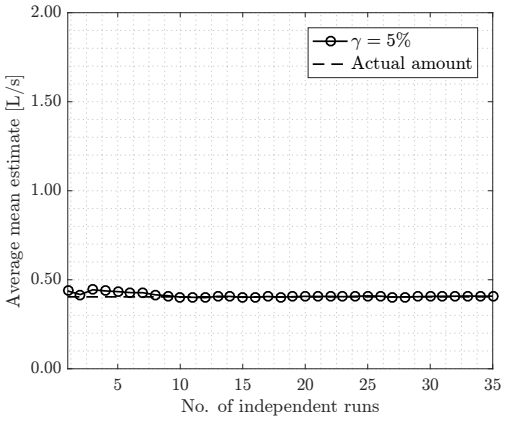


Figure 6.23: Sample average of the mean estimates of leakage intensity in terms of the number of independent detection processes. Measurement noise ($\gamma = 5\%$). 10 flow tests at each location. Alternative sensor location

minimum number of sensors. Even though the analysis considers an incomplete number of possible scenarios (in terms of the sensors utilized), it is clear that the number of sensors plays a significant role on the performance of the identification process.

The analyses conducted in this section show that adequate monitoring strategies can have a significant impact on detection processes. Moreover, sampling design methods for locating sensors can improve detection and diagnosis results. These issues are a pivotal task for monitoring critical infrastructure systems [53,81–83]. The fundamental goal is to obtain the maximum possible information from the water distribution network by selecting the optimal number and location of sensors. Moreover, the economic feasibility and impact of

Table 6.3: Description of scenarios in terms of the flow meters used during the identification process

Scenario	Sensors used
1	1
2	1–2
3	1–2–3
4	1–2–3–4
5	1–2–3–4–5

Table 6.4: Most probable model classes for different scenarios

Scenario	Most probable model classes
1	(several incorrect locations are predicted)
2	14 (incorrect leakage location)
3	72 (correct leakage location)
4	72 (correct leakage location)
5	72 (correct leakage location)

the monitoring strategies are important issues that need to be explicitly considered into the decision-making process. These are challenging problems and constitute a future research effort.

6.8 Computational cost

In this application example, each hydraulic simulation of the network (the solution of one set of non-linear equations associated to steady-state conditions) takes approximately 0.22 [s]. The total number of hydraulic simulations depends on the number of stages required by the transitional Markov chain Monte Carlo method and the sample size specified for each stage (1000 samples per stage in this case). This gives an average of 22 [min] required to estimate the evidence of each probabilistic model class. Considering that all model classes are completely independent, the entire model class selection process takes about 4.5 hours. The previous computational efforts are based on the implementation of the proposed approach in available four-core multi-threaded computer units.

6.9 Final remarks

The results presented in this chapter have shown that the proposed approach can be applied to leakage detection problems involving real-life water distribution networks. Moreover, the methodology allows to correctly predict the location and the severity of the prescribed leakage for the network under investigation. The effect of modeling errors and measurement noise was investigated by considering them to generate the synthetic measurements. Validation calculations showed that the performance of the detection process becomes more sensitive to these errors as the leakage intensity decreases. Moreover, it was observed that no reliable diagnosis can be obtained beyond certain thresholds for leakage intensity, modeling errors and measurement noise. If these thresholds are trespassed, available data does not allow to properly represent the behaviour of the actual system and, therefore, no reliable identification can be made. The proposed approach is able to identify these thresholds, providing valuable information about the water utility network. Finally, the location and number of sensors have a significant effect on the capabilities of the proposed methodology. The definition of optimal monitoring strategies to improve the effectiveness of detection processes is a highly relevant issue that is left for future research efforts.

Chapter 7

APPLICATION 3: CONNECTIVITY DETECTION

Current management of water utility networks involves a constant evolution of their topology, changes in their control elements and definition of new operational schedules, in order to address new users requirements and changing environmental conditions. Examples of these situations include the construction of pipelines, control valves and water supply facilities. Some of the issues arising in this rapidly changing context are related to the incorrect representation of water distribution systems at decision-making levels. One of the common problems faced by water utility managers in this regard is the imprecise topological characterization of hydraulic models. In particular, incorrect connectivity in terms of pipe elements is usually found in real-life hydraulic models. In this chapter, the proposed Bayesian framework is implemented to handle the connectivity detection problem. Here, the objective is to identify the unknown location of a certain pipeline based on available data. The application example involves the hydraulic model of a real-life water distribution network under quasi-dynamic conditions. The results show the applicability and effectiveness of the proposed approach in this class of problems, which naturally arise as strictly unidentifiable systems.

7.1 Problem formulation

7.1.1 Description of the network

The case of study considered in this chapter involves the water distribution system of an urban area in Viña del Mar, Chile. The system comprises two distribution sectors, namely Sector A and Sector B, that are supposedly independent. This setting is depicted in Figure 7.1, where Sector A is identified with black lines and Sector B with red lines.

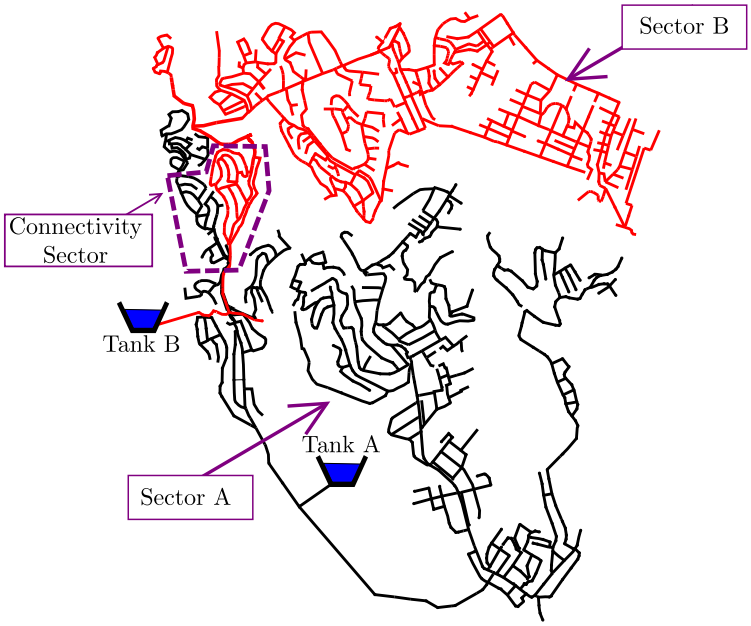


Figure 7.1: Water distribution network of an urban sector in Viña del Mar

The hydraulic model is constructed in the commercial-level software EPANET, consisting of 1441 nodes (555 nodes belong to sector A and 886 nodes belong to sector B), 1478 pipes, 13 pressure reduction valves, 62 throttle control valves and 2 storage tanks. Quasi-dynamic conditions are considered in the context of this application example. The corresponding extended period simulation represents a regular day from 08:00 to 19:00 hours (12 steady-state simulations). Appendix D presents more details regarding the steady-state simulation of pressurized pipe systems.

The network topology planned by the water utility managers, which is the one represented in the hydraulic model, considers that there is no interaction between both sectors (that is, no network component connects them). In other words, each sector is supplied by

only one tank: tank A supplies Sector A, while Sector B is supplied by tank B. The volumes of water in tanks A and B during the simulation period are presented in Figure 7.2. The physical characteristics of these regulation elements are summarized in Table 7.1, where minimum and maximum heads refer to meters above sea level. One of the important physical characteristics of this system is that tank A is located about 120 [m] higher than tank B. As a consequence, hydraulic heads of sector A are higher than the ones of sector B in areas with similar elevation.

Table 7.1: Description of regulation elements - Application example 3

Tank	Volume [m^3]	Minimum head [m]	Maximum head [m]	Users (houses)	Maximum demand [l/s]
Tank A	2000	309.21	314.31	2200	98.4
Tank B	500	184.00	188.44	1400	63.6

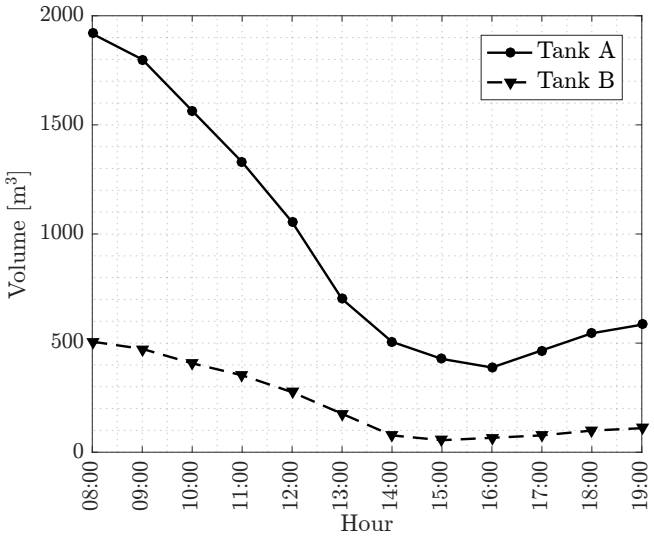


Figure 7.2: Volume of water in each tank during the simulation period

The hydraulic network comprises 67.56 [km] of pipelines, modeled with 1478 pipe elements. The lengths of the pipes arranged in an increasing manner are presented in Figure 7.3, ranging from 2.81 [m] to 258.7 [m]. Pipes of four different materials are considered: high density polyethylene (HDPE), polyvinyl chloride (PVC), asbestos cement (AC) and ductile iron. The corresponding Hazen-Williams coefficients are 150, 140, 140 and 120, respectively. Table 7.2 presents the number and total length of pipes of the different materials.

Table 7.2: Number of pipes and total length of different materials. Application example 3

Material	Number of pipes	Total length [km]
High density polyethylene (HDPE)	7	0.52
Polyvinyl chloride (PVC)	36	1.71
Asbestos cement (AC)	1177	56.54
Ductile iron	258	8.79

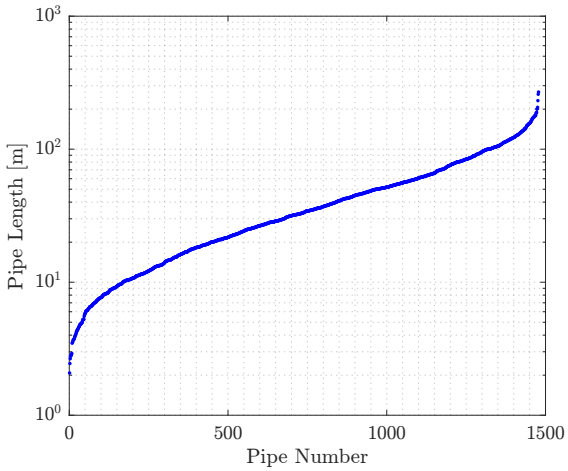


Figure 7.3: Pipe lengths of the water distribution system - Application example 3

As already pointed out, this application example considers an extended period simulation of 12 time steps, between 08:00 AM and 19:00 PM. The normalized demand pattern of a representative node of the network is illustrated in Figure 7.4. Note that the pattern follows the typical consumption behaviour of a residential zone in Chile. All demand nodes present the same normalized pattern. It is also seen that the peak demand is expected around 12:00 PM (noon), corresponding to a total demand of 162 [l/s]. The minimum and maximum nodal demands of the system at this time of the day are 1.54×10^{-2} [l/s] and 35.43 [l/s], respectively.

7.1.2 Topological characterization error

The operational scenario considered by water utility managers to model the hydraulic distribution system assumes that flow paths between Sectors A and B do not exist (hydraulic independence between sectors). Nonetheless, it is believed that the actual water distribution system does have a pipeline that connect both sectors. Indeed, available field data strongly suggests that connectivity between both sectors does exist. Hence, there is a

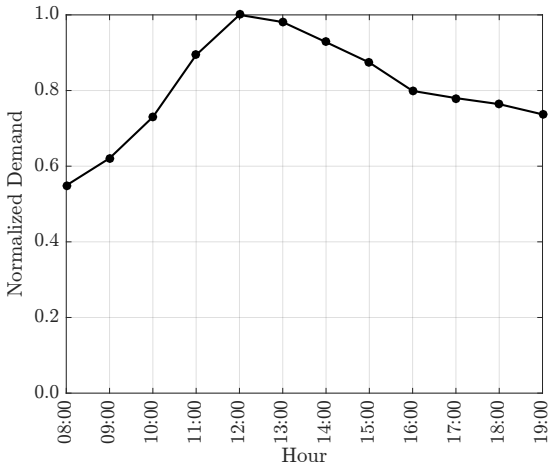


Figure 7.4: Normalized demand pattern of a representative node of the network

topological characterization error in the hydraulic model, which must be detected in order to properly update the engineering model. In this manner, the hydraulic performance of the system could be evaluated in its current and real condition, so better decision-making processes regarding the repair, maintenance and enhancement of the water distribution network can be developed.

The actual operational condition of the system, namely the actual connectivity between sectors A and B, presents several practical issues that directly affect the hydraulic performance of the water distribution network. First, the existence of connectivity between both sectors and the physical characteristics of the system lead to flow from Sector A to Sector B. In other words, a great part of the water volume pumped to Tank A is returned to sector B, which is located at a lower level (see Table 7.1). This represents a clear energy waste that makes the system more inefficient. Second, the increased flow in a portion of Sector B increases the hydraulic energy losses in such zone, reducing the pressure heads downstream. Then, the overall hydraulic performance of the system is worsened. Finally, it is more complicated (or virtually impossible) to isolate specific sectors of the system using valves, since the actual topology is unknown. This leads to higher operational costs and poor decision making about the water distribution system. For these reasons, the importance of the connectivity detection problem in the context of hydraulic performance assessment becomes evident.

The topological characterization error may occur due to different reasons. For example, it is possible to have problems in Geographic Information System (GIS) data. Current

modeling techniques rely on GIS, which provides functions to handle most of the information required by water utility managers: network layout and connectivity, topographic survey, pipe characteristics, distribution and classification of users, detection of possible interaction and interferences with other utility networks, etc. However, one of the common errors in the GIS data are missing pipelines in the network layout, which leads to incorrect connectivity modeling [27]. Additionally, connectivity errors might be caused by errors in data processing, poor information management, and missing pipes in official water system drawings.

The actual pipe that is missing from the network model is known to be located within a reduced area of the network, identified in Figure 7.1 as *Connectivity Sector*. That network zone is shown in more detail in Figure 7.5 (left), where the location of the connecting pipeline is also presented. This link element has the following hydraulic properties: diameter of 100 [mm], Hazen-Williams coefficient equal to 140 (asbestos cement), and length of 48.9 [m].

7.1.3 Definition of connectivity events

In this application example, it is necessary to identify the nodes that are actually connected in the connectivity sector and the hydraulic properties that define the corresponding link element. Thus, each probabilistic model class addresses a potential connection within the prescribed sector, by introducing a pipe element from a node in Sector A to a node in Sector B. Specifically, fifteen feasible connections are considered, which are shown in Figure 7.5 (right). Note that model class M^4 coincides with the actual connectivity event. Here, the model class $M^\ell, \ell = 1, \ldots, N_{\text{class}}$ is called the ℓ-th connectivity event, where $N_{\text{class}} = 15$. Besides, it is assumed that no prior information is available regarding the plausibility of each connectivity event, and therefore $P\left[M^\ell|\mathbf{M}\right] = 1/15$.

To define the probabilistic model class $M^\ell, \ell = 1, \ldots, N_{\text{class}}$, a pipe parametrized by $\boldsymbol{\theta}^\ell$ is introduced in the hydraulic model. Recall that the hydraulic properties of a pipe in EPANET are defined by the length, diameter and roughness coefficient [26]. The length of the connecting pipe is already fixed by the coordinates of the connected nodes. Then, each probabilistic model class considers two uncertain parameters $\boldsymbol{\theta}^\ell = \langle \theta_1^\ell, \theta_2^\ell \rangle^T$, where θ_1^ℓ is the diameter of the pipe and θ_2^ℓ is the corresponding roughness coefficient (Hazen-Williams coefficient). The likelihood function of each probabilistic model class is described in Section 5.3.1. On the other hand, the prior distribution for each probabilistic model

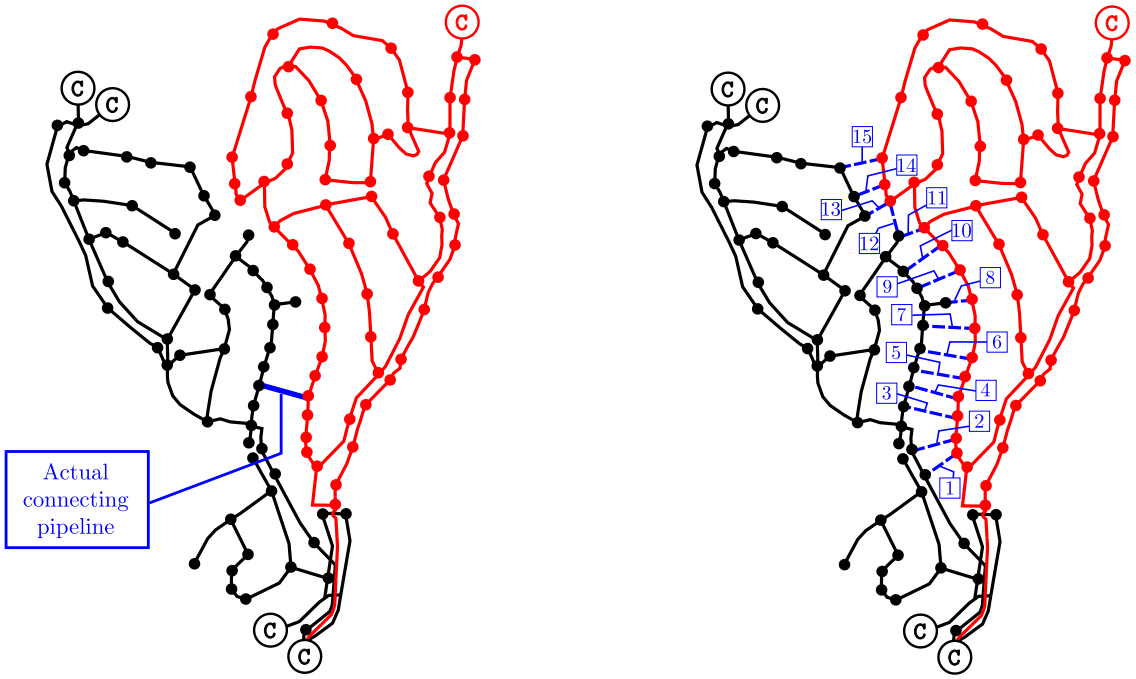


Figure 7.5: Actual connecting pipe in connectivity sector (left) and definition of connectivity events (right). C: connection nodes to the rest of the network

class is uniform. The diameter (θ_1^ℓ) ranges from 50 [mm] to 250 [mm], while the roughness coefficient (θ_2^ℓ) ranges from 80 to 150. This distribution is chosen according to the the feasibility of pipe diameters and construction materials used in the particular sector under consideration. Nevertheless, preliminary calculations showed that the definition of the range of these parameters does not seem to have an impact into the effectiveness of the detection process.

7.1.4 Simulated data

In connection with Section 6.1.3, data D consists of simulated measurements from a prescribed hydraulic model known as the *actual network*. The data generation process takes explicitly into account both measurement noise and modeling errors. Preliminary calculations showed that pressure measurements were not very informative in this case. Then, only flow measurements are considered. Specifically, eleven monitoring devices distributed over the connectivity sector are considered (see Figure 7.6). The flow rates obtained during all the simulation period (from 08:00 AM to 19:00 PM) are included in the data used for the detection process. This gives $N_Q = 12$ quantities of interest (flows) at each monitoring

location. Then, simulated data $\mathbf{y}^* \in \mathbb{R}^{N_{\text{data}}}$ is defined as

$$\mathbf{y}^* = \left\langle \mathbf{y}_1^{*T}, \ldots, \mathbf{y}_{N_F}^{*T} \right\rangle^T \tag{7.1.1}$$

where $\mathbf{y}_i^* \in \mathbb{R}^{N_L \times N_Q}, i = 1, \ldots, N_F$ is the vector containing all measurements obtained at the N_L locations during the i-th flow test. This vector is defined as

$$\mathbf{y}_i^* = \left\langle \mathbf{y}_{i,1}^{*T}, \ldots, \mathbf{y}_{i,N_L}^{*T} \right\rangle^T \tag{7.1.2}$$

where $\mathbf{y}_{ij}^* \in \mathbb{R}^{12}$ is the vector containing the flow rates obtained at location j during the i-th flow test, defined by

$$\mathbf{y}_{ij}^* = \langle y_{i,j,1}, \ldots, y_{i,j,12} \rangle^T \tag{7.1.3}$$

and $y_{i,j,k}^*$ is the flow rate obtained at time step k and location j during flow test i. According to Section 6.1.3, each datum used for the detection process is generated as

$$y_{ijk}^* = y_{ijk}^{\text{model}} + y_{ijk}^{\text{noise}} \tag{7.1.4}$$

where y_{ijk}^{model} is obtained from the actual network and y_{ijk}^{noise} is assumed to be representative of the measurement errors. Modeling errors are included in pipe roughness coefficients (intensity α) and nodal demands (intensity β), while measurement errors are assumed to be a uniform noise with intensity γ. The reader is referred to Section 6.1.3 for more details regarding the inclusion of errors in the data simulation process.

7.1.5 Implementation details

The Bayesian model updating procedure was performed by using the transitional Markov chain Monte Carlo method with 500 samples per stage. The covariance matrix in Equations (5.3.3) and (5.3.4) is taken as $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and $\sigma^2 \in \mathbb{R}^+$.

7.2 Detection process in idealized scenario

The idealized scenario assumes no modeling or measurement errors in the detection process, that is, $\alpha = \beta = \gamma = 0$. The normalized evidences for each probabilistic model class are presented in Figure 7.7, considering only one flow test at each monitoring location ($N_F = 1$). It is seen that the actual connectivity event is correctly identified in this case, since model class M^4 presents the highest evidence. In fact, the posterior probability of

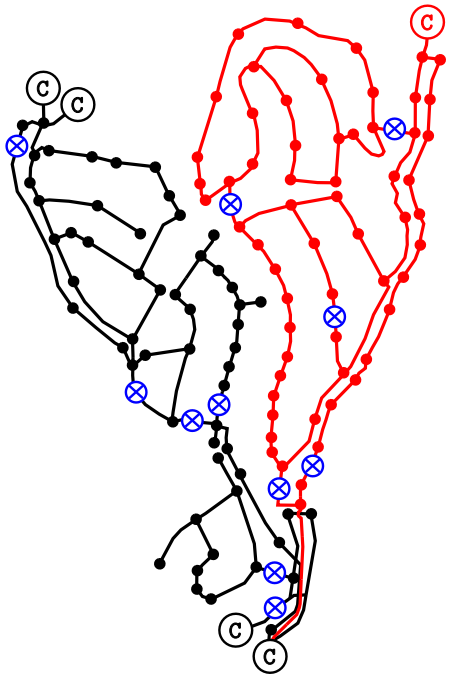


Figure 7.6: Layout of flowmeters

model class M^4 is almost one, while the other model classes present a posterior probability almost equal to zero.

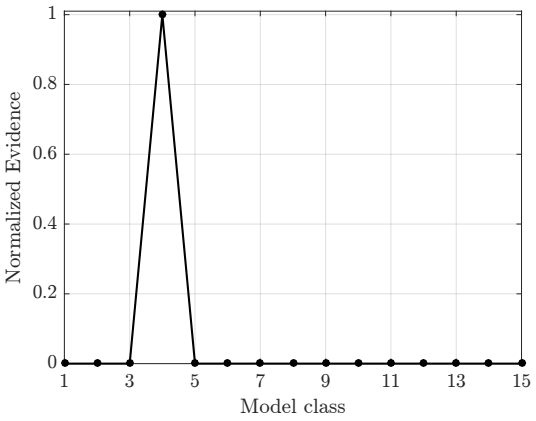


Figure 7.7: Normalized evidences of all model classes. Connectivity detection. Idealized scenario

Figure 7.8 shows the samples obtained at each stage of the transitional Markov chain Monte Carlo algorithm for probabilistic model class M^4. A total of six stages are required to obtain a sample from the posterior distribution. The evolution of the scatter plots in the space $\langle \theta_1^4, \theta_2^4 \rangle$ shows how the samples converge when the actual connectivity event is

considered. The samples are clearly correlated in a certain direction in the parameter space, in such a way that an increase in the diameter (θ_1^4) is compensated by a decrease in the roughness coefficient (θ_2^4). This result is consistent with the physics [84] of the hydraulic system: an increase of the diameter (enhancement of the hydraulic capacity) is compensated by decreasing the Hazen-Williams coefficient (reduction of the hydraulic capacity). In fact, since the hydraulic head lost by water flowing in a pipe due to friction with the walls is modeled with the Hazen-Williams formula (see Equation (D.1.5)), any pair of values θ_1 and θ_2 such that ${\theta_1}^{4.870} \times {\theta_2}^{1.852} = \left(\theta_1^{\text{real}}\right)^{4.870} \times \left(\theta_2^{\text{real}}\right)^{1.852}$ will define a pipe with the same friction losses, where $\theta_1^{\text{real}} = 100$ [mm] and $\theta_2^{\text{real}} = 140$. In other words, all networks defined by points in that *manifold* are equivalent from the hydraulic point of view and, therefore, those networks have the same posterior probability density function values. This means that the system is strictly unidentifiable, since a complete region in the space of parameters maximize the posterior probability density function. It is remarked the proposed approach is able to identify this strong interaction between model parameters, which gives a valuable insight into the system behavior.

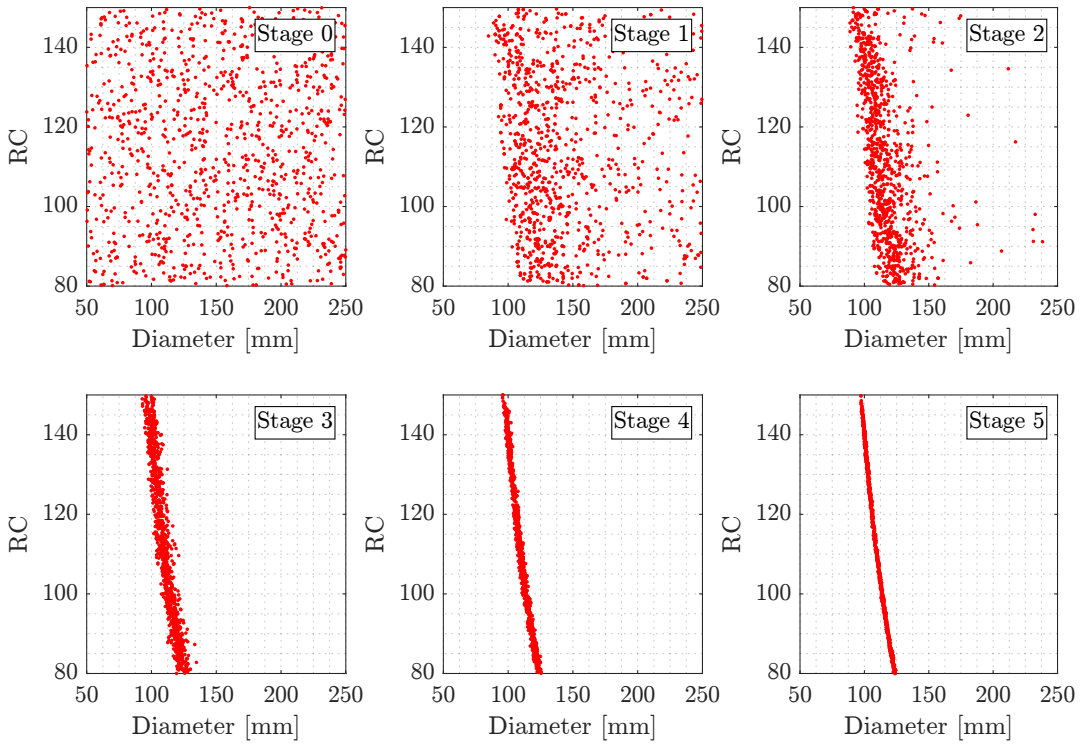


Figure 7.8: Samples at each stage of TMCMC. Connectivity detection. Idealized scenario

Available information allows to reduce the uncertainties in the model parameters. In

this context, Figure 7.8 demonstrates a significant reduction of the uncertainty in the diameter. The posterior sample shows that the parameter associated to the diameter ranges from 97.4 [mm] to 123.6 [mm], which includes the actual diameter of the connective pipe, 100 [mm]. Conversely, the uncertainty associated to the roughness coefficient remains practically unchanged: the support of the unidentifiable domain coincides with the support of the prior distribution. This observation agrees with the physical behavior of the system, since the hydraulic response of the network is more sensitive to the diameter of the connective pipe than to its Hazen-Williams coefficient. The previous results demonstrate some of the advantages of simulation-based Bayesian updating procedures over traditional or particular techniques that try to identify one best model when there is limited data about the behaviour of the system.

7.3 Effect of model and measurement errors

In order to investigate the effect of model and measurement errors, the simulated data is now generated as discussed in Section 6.1.3. Specifically, the following scenarios are studied here:

a) Modeling errors in pipe roughness coefficients only: $\beta = \gamma = 0, \alpha = 10\%$.

b) Modeling errors in nodal demands only: $\alpha = \gamma = 0, \beta = 10\%$.

c) Measurement errors only: $\alpha = \beta = 0, \gamma = 5\%$.

d) Combined modeling and measurement errors: $\alpha = \beta = 10\%, \gamma = 5\%$.

In all these cases, the proposed approach was able to correctly identify the actual connectivity event, favouring model class M^4 as the most probable one. In fact, the posterior probability of the rest of model classes remains very close to zero. Thus, the illustration of the normalized evidences is analogous to Figure 7.7. All these results were obtained by considering only one flow test at each monitoring location ($N_F = 1$). Therefore, the identification process is quite robust to model and measurement errors for this application example.

The identification of the parameters that best describe the connectivity event is affected by the quality and amount of available data. In this regard, it is interesting to observe how increasing the amount of measurements used in the detection process can help to reduce

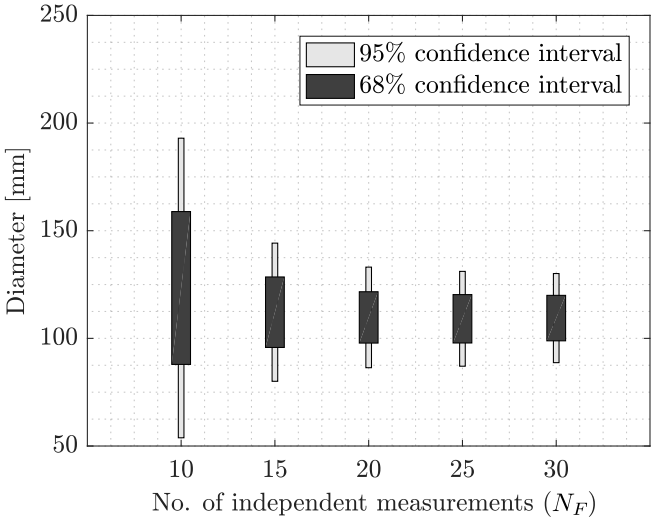


Figure 7.9: Confidence intervals of mean diameter estimate for different numbers of flow tests. Measurement errors: $\gamma = 5\%$

the uncertainty in the model parameters. Figure 7.9 shows the effect of the number of flow tests at each monitoring location (N_F) into the statistical information of the parameter related to the diameter of the actual connective pipe, θ_1. The statistics are based on 100 independent detection processes; in other words, the expected diameters associated to the posterior samples obtained from 100 independent sets of measurements. Specifically, the figure presents the 68% ($[\mu_{\theta_1} - \sigma_{\theta_1}, \mu_{\theta_1} + \sigma_{\theta_1}]$) and the 95% ($[\mu_{\theta_1} - 1.96\sigma_{\theta_1}, \mu_{\theta_1} + 1.96\sigma_{\theta_1}]$) credible intervals in the case of large measurement errors ($\gamma = 5\%$). In the previous notation, μ_{θ_1} and σ_{θ_1} correspond to the mean value and standard deviation of the parameter θ_1, respectively. The credible intervals represent the uncertainty in the expected value of the diameter under large measurement errors and different numbers of flow tests. Note that the variability of the posterior mean diameter decreases as the number of flow tests increases, which is reasonable since more information about the system is available. It is also observed that the actual pipe diameter, 100 [mm], is inside the credible intervals in all cases. The size of the credible intervals stabilizes after 20 flow tests. In other words, to obtain more than 20 flow tests at each monitoring location does not seem to improve the quality of the detection process in terms of the pipe diameter. On the other hand, all the detection processes showed that the support region of the roughness coefficient remains unchanged with respect to the prior distribution, as observed in the idealized scenario (see Figure 7.8).

7.4 Computational costs

In the context of this application example, each extended period simulation of the network takes approximately 0.12 [s]. A sample size of 500 is specified for each stage of the transitional Markov chain Monte Carlo method, which gives an average of 12 [min] required to estimate the evidence of each probabilistic model class. Then, the entire model class selection process takes about 2.5 hours. As in Chapter 6, the previous computational efforts are based on the implementation of the proposed approach in available four-core multi-threaded computer units.

7.5 Final remarks

This chapter has shown that the simulation-based Bayesian model updating approach can be successfully applied to connectivity detection problems in large-scale water distribution networks. These problems naturally arise as ill-conditioned, due to the strong interaction between the parameters that need to be updated. Nevertheless, the method was able to successfully identify the detection of the actual connectivity event with only one flow test at each monitoring location. The approach also proved to be quite robust to the modeling and measurement errors under consideration. On the other hand, the capability of drawing a posterior sample of the model parameters gives an important insight into the system behavior. This feature allowed to reduce the uncertainties in the diameter of the connective pipe. In this context, the identification scheme can identify the amount of information beyond which no further improvement into estimated model parameters can be achieved. Finally, the application example involved the integration of the Bayesian model updating technique with a quasi-dynamic analysis of the hydraulic system, demonstrating the generality and applicability of the proposed approach. This shows that the method can be extended, in principle, to other detection-related problems in complex utility networks.

Chapter 8

CONCLUSIONS

This work has presented a general framework based on probability theory to handle two types of problems related to the hydraulic performance of complex water distribution systems: hydraulic reliability assessment, and detection-related problems. The proposed approaches have been tested in several examples involving real-life hydraulic networks.

A stochastic approach has been proposed to address the hydraulic reliability assessment of water distribution networks. The approach was applied to the hydraulic reliability assessment of a real-life water distribution system, in terms of the minimum pressure head over the network. A number of analyses were conducted, including uncertainty propagation, reliability analysis, reliability sensitivity analysis, and failure analysis. Also, several network settings were studied, such as different demand configurations and redundancy scenarios. Results show the effectiveness and efficiency of the proposed framework in these high-dimensional reliability problems, providing valuable information for decision-making. The effectiveness of the approach under quasi-dynamic conditions was also demonstrated, which involved tens of thousands of uncertain parameters. In this regard, the approach represents a change of paradigm in the hydraulic reliability assessment of water distribution networks, since the size and complexity of the network are not a limitation anymore.

Future research efforts in the context of hydraulic reliability assessment involve the consideration of other types of failure modes, such as water quality, nodal flows, and flow velocities. The implementation of pressure-driven analysis in the proposed framework is also an additional research topic. Another subject for future work is the integration of hy-

draulic reliability with mechanical or structural reliability to obtain an overall performance measure of complex water distribution networks under uncertain conditions. Finally, another research topic is the extension of the approach to other types of complex utility networks operating under uncertain conditions such as general pressurized pipe networks, including oil pipelines and natural gas networks.

A Bayesian model updating approach has been proposed to handle detection-related problems in water distribution networks. The approach was tested on two problems involving real-life hydraulic models: leakage detection, and connectivity detection. According to the results obtained, the following conclusions can be made. First, the diagnosis of the system fails when the levels of modeling or measurement errors are above certain thresholds. Such thresholds depend on a number of factors, such as characteristics of the network, quality of monitoring devices, etc. The proposed methodology can identify these thresholds beyond which no reliable identification is possible. Second, the configuration of sensors significantly affects the performance of the detection process. Thus, optimal monitoring strategies can improve the capabilities of the approach, which should consider economical, technical and environmental aspects. Third, the proposed methodology is highly parallelizable, being suitable for a High Performance Computing environment. Finally, with the proper handling of the above aspects, the proposed approach is potentially a functional tool for solving a class of detection-related problems in complex water distribution networks.

Future research efforts in the context of the proposed Bayesian approach involve the consideration of actual measurements for the detection process. The integration of pressure-driven analysis is also a subject for future work. The consideration of stochastic demand models can also be explored in the future. Another research topic is the implementation of optimal sensor location strategies to improve the capabilities of the methodology. Finally, the proposed approach can be used to handle other detection-related problems in water utility networks. Work in this direction is currently under consideration.

To conclude, the probabilistic framework developed in this work provides general tools to handle hydraulic reliability assessment and detection-related problems in large-scale water distribution systems. The proposed methodologies allow to obtain valuable information about the hydraulic performance of this class of critical infrastructure networks, which can be useful to assist water utility managers in complex decision-making under uncertain conditions.

Appendix A

SUBSET SIMULATION

This appendix describes subset simulation, the stochastic simulation method considered in this work to perform the hydraulic reliability assessment of complex water distribution networks. The basic idea of the algorithm is discussed, its main characteristics are described, and a pseudo-code with the actual implementation is provided.

A.1 Main idea

Subset simulation [3, 25] is based on the decomposition of the failure domain F as a sequence of m nested *intermediate failure events* F_κ, $\kappa = 1, \ldots, m$ such that

$$F = F_m \subset F_{m-1} \subset \cdots \subset F_2 \subset F_1 = \bigcap_{\kappa=1}^{m} F_\kappa \tag{A.1.1}$$

which allows to compute the failure probability P_F as

$$P_F = P(F) = P(F_1) \times \prod_{\kappa=2}^{m} P(F_\kappa | F_{\kappa-1}) \tag{A.1.2}$$

Note that the failure probability P_F is now computed in terms of the unconditional failure probability $P(F_1)$ and the conditional failure probabilities $P(F_\kappa|F_{\kappa-1})$, $\kappa = 2, \ldots, m$. An appropriate definition of the sequence of intermediate failure events allows to compute these quantities very efficiently with sampling techniques. In this context, two main practical issues must be addressed to implement this simulation technique: (1) the definition

of the intermediate failure events and (2) the generation of samples conditional to such intermediate failure events. These topics are discussed in the next sections.

In this appendix, it is assumed that the uncertain parameters are independent standard Gaussian random variables, $\boldsymbol{\theta} \in \Omega_\theta \subset \mathbb{R}^{n_\theta}$. This assumption does not represent any loss of generality, since available transformation techniques allow the mapping from the standard Gaussian space to the space of physical parameters [85, 86].

A.2 Definition of intermediate failure events

The intermediate failure event F_κ is defined in terms of the intermediate threshold level μ_κ^* as

$$F_\kappa = \{\boldsymbol{\theta} \in \Omega_\theta : \mu(\boldsymbol{\theta}) < \mu_\kappa^*\}, \qquad \kappa = 1, \ldots, m \tag{A.2.1}$$

The selection of the intermediate failure events, which is actually the definition of the intermediate threshold levels, has a significant impact on the performance of the method. In the standard implementation [3, 25, 87], which is the one considered here, the intermediate threshold levels are adaptively defined so that the conditional failure probabilities are equal to a prestablished value. That is, the intermediate failure events verify $P(F_1) = P(F_\kappa|F_{\kappa-1}) = p_0, \kappa = 2, \ldots, m$, where the parameter p_0 is called the *conditional failure probability*. Validation calculations show that choosing any value of p_0 between 0.1 and 0.3 will lead to similar efficiency as long as subset simulation is implemented properly [3, 87].

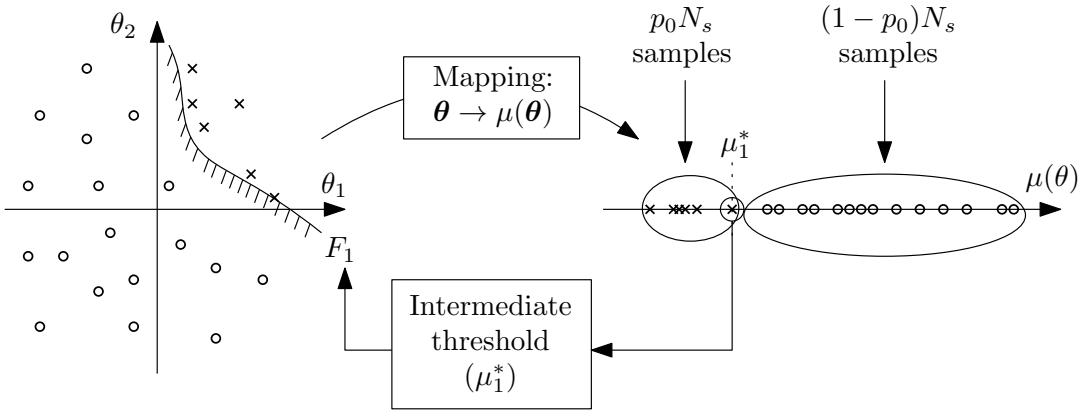


Figure A.1: Definition of the first intermediate failure event (F_1). The subsequent intermediate failure events follow the same idea

In this work, a constant sample size N_s is implemented, that is $N_\kappa = N_s, \kappa = 1, \ldots, m$. Now, consider the set of samples $\{\boldsymbol{\theta}_{\kappa-1,i} : i = 1, \ldots, N_s\}$ generated during the κ-th stage of subset simulation. The corresponding values of the utility function are $\{\mu(\boldsymbol{\theta}_{\kappa-1,i}) : i = 1, \ldots, N_s\}$. Then, the threshold level μ_κ^* associated to the intermediate failure event F_κ is defined as the $(p_0 N_s)$-th smallest value of the latter set. In this way, there are $p_0 N_s$ samples lying in the intermediate failure domain F_κ and $P(F_\kappa|P_\kappa - 1) = p_0$ holds. This is schematically represented in Figure A.1.

A.3 Conditional sampling: Markov chain Monte Carlo

Subset simulation requires the generation of samples distributed according to the conditional probability density function $p(\boldsymbol{\theta}|F_\kappa)$, that is

$$p(\boldsymbol{\theta}|F_\kappa) = p(\boldsymbol{\theta}) I_{F_\kappa}(\boldsymbol{\theta}) \tag{A.3.1}$$

Markov chain Monte Carlo methods arise as natural tools to generate conditional samples. In these methods, the samples are generated as a sequence of a *Markov chain*, that is, a stochastic process in which the distribution of each state depends only on the previous state. The Metropolis-Hastings algorithm [66, 67] is one of the most applied techniques in this context, although it is quite inefficient in high-dimensional reliability problems. Thus, the *modified Metropolis-Hastings* algorithm is considered [3]. A basic pseudo-code of this algorithm applied to generate a new sample $\boldsymbol{\theta}_{k+1}$, conditional to some failure event $F^\star$ and starting from the current sample $\boldsymbol{\theta}_k$, is provided here.

- **Step 1**: Generation of a candidate sample $\boldsymbol{\theta}^c$.
  - For $\iota = 1, \ldots, n_\theta$:
    1. Generate θ'_ι from the one-dimensional proposal distribution $p^*_\iota(\theta'_\iota; \theta_{k,\iota})$.
    2. Calculate the following quotient, based on the one-dimensional normal distribution $p_\iota(\cdot)$:
    $$r_\iota = p_\iota(\theta'_\iota)/p_\iota(\theta_{k,\iota})$$
    3. Generate a random number u uniformly distributed on [0,1]. If $u < r_\iota$, set $\theta^c_\iota = \theta'_\iota$ (accept component). Otherwise, set $\theta^c_\iota = \theta_{k,\iota}$ (reject component).
- **Step 2:** Verify that the candidate sample lies in $F^\star$.

  (a) If $\boldsymbol{\theta}^c = \langle \theta^c_1, \ldots, \theta^c_{n_\theta} \rangle^T \in F^\star$, set $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}^c$ (accept candidate sample).

(b) Otherwise, set $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$ (reject candidate sample).

The *proposal distribution* at the entry level $p_\iota^*(\cdot;\cdot)$ is defined as a uniform distribution centred at the current sample, that is

$$p_\iota^*(\theta'_\iota;\theta_{k,\iota}) = \frac{1}{\xi} I\left[\left(1-\frac{\xi}{2}\right)\theta_{k,\iota} < \theta'_\iota < \left(1+\frac{\xi}{2}\right)\theta_{k,\iota}\right] \tag{A.3.2}$$

where $\xi > 0$ is the spread parameter, that is, the width of the proposal distribution.

In brief, the modified Metropolis-Hastings algorithm generates each component of the candidate sample $\boldsymbol{\theta}^c$ based on the current sample $\boldsymbol{\theta}_k$ in an independent manner. Then, the next state of the Markov chain $\boldsymbol{\theta}_{k+1}$ is selected as $\boldsymbol{\theta}^c$ or $\boldsymbol{\theta}_k$ according to whether $\boldsymbol{\theta}^c$ lies in the failure domain or not.

In the previous setting, a total of $p_0 N_s$ samples lying in the intermediate failure domain F_κ are obtained at the $(\kappa-1)$-th simulation level. Thus, a total of $p_0 N_s$ independent Markov chains are generated using the modified Metropolis-Hastings at level κ, where each chain comprises $1/p_0$ conditional samples. This is schematically represented in Figure A.2, which shows a bi-dimensional case involving two simulation stages.

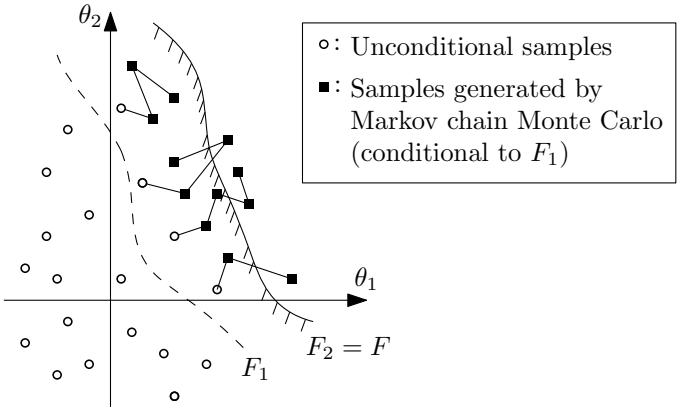


Figure A.2: Example of Markov chains generation during subset simulation - two levels

A.4 Updating of the spread parameter

One important aspect of subset simulation is the definition of the proposal distribution. Specifically, the spread parameter ξ of the uniform distribution has a great effect on the performance of the simulation procedure. This parameter affects the size of the

region covered by the Markov chain samples, and therefore it controls the efficiency of the method. If ξ is too small, the samples will lie very close to each other which leads to highly correlated Markov chains, worsening the quality of the failure probability estimate. On the other hand, a decrease in the acceptance rate is likely to occur if the parameter ξ is chosen too large, then increasing the number of repeated Markov chain samples and slowing down convergence. Therefore, the choice of the spread parameter ξ is a trade-off between acceptance rate and correlation due to proximity.

Practical experience indicates that the optimal rejection rate ranges roughly between 50% and 70% [87]. In this context, the present implementation involves an adaptive definition of the spread parameter. That is, the value of ξ is decreased or increased during the simulation according to whether the rejection rate R lies in the interval $[R_L, R_U]$, where R_L and R_R are lower and upper limits for the rejection rate. Specifically, after the generation of $N_{\text{mc}} < p_0 N_s$ Markov chains have been generated, the observed rejection rate R^* associated to these chains is computed and, then, the spread parameter ξ can be increased (if $R^* < R_L$) or decreased (if $R^* > R_L$). The following procedure is included in the κ-th stage ($\kappa = 2, \ldots, m$) of subset simulation to update the spread parameter.

1. Set $\xi = 1$.
2. Generate N_{mc} Markov chains of length $1/p_0$ according to Section A.3. Compute the total number of rejections of candidate samples obtained in these chains, N_R, and the average rejection rate as $R^* = N_R \,/(N_{\text{mc}}/p_0)$.
3. If $R^* > R_U$, set $\xi \leftarrow \gamma_1 \times \xi$. If $R^* < R_L$, set $\xi \leftarrow \gamma_2 \times \xi$.
4. If the total number of samples required for the current stage (N_s) has been generated, continue to the next stage. Otherwise, go back to Step 2.

In the above procedure, γ_1 and γ_2 are the updating factors. This setting leads to a total of $N_g = N_s p_0 / N_{\text{mc}}$ updating steps of the spread parameter.

A.5 Pseudo-code

1. Generate N_s samples $\{\boldsymbol{\theta}_{0,i} : i = 1, \ldots, N_s\}$ by direct Monte Carlo according to the probability density function $p(\boldsymbol{\theta})$. Set $\kappa = 1$.
2. Evaluate the utility function to obtain $\{\mu(\boldsymbol{\theta}_{\kappa-1,i}) : i = 1, \ldots, N_s\}$. Arrange these values in a decreasing manner.

3. Set μ_{aux} as the $[(1-p_0)N_s+1]$-th value of the sorted set $\{\mu(\boldsymbol{\theta}_{\kappa-1,i}) : i = 1,\ldots,N_s\}$.

   - If $\mu_{\text{aux}} > \mu^*$, set $\mu^*_\kappa = \mu_{\text{aux}}$.
   - Otherwise, go to step 8.

4. Define the κ-th intermediate failure domain as

$$F_\kappa = \{\boldsymbol{\theta} \in \Omega_\theta \,:\, \mu(\boldsymbol{\theta}) < \mu^*_\kappa\}$$

5. The sampling estimate for $P(F_\kappa)$ (if $k=1$) or $P(F_\kappa|F_{\kappa-1})$ (if $k>1$) is equal to p_0 by construction. Recall that p_0 and N_s are chosen such that $p_0 N_s$ is an integer number.

6. There are $p_0 N_s$ samples among $\{\boldsymbol{\theta}_{\kappa-1,i} : i = 1,\ldots,N_s\}$ whose utility function value is equal or smaller than μ^*_κ. Identify those samples for generating new samples with the modified Metropolis-Hastings algorithm.

7. Starting from each of these conditional samples, Markov chain Monte Carlo simulation (Section A.3) is used to generate an additional of $(1-p_0)N_s$ conditional samples that lie in F_κ. Update the spread parameter ξ every N_{mc} Markov chains (Section A.4). In this way, a total of N_s conditional samples $\{\boldsymbol{\theta}_{\kappa,i} : i = 1,\ldots,N_s\}$ at level κ are obtained. Set $\kappa \leftarrow \kappa + 1$ and go back to step 2.

8. The failure probability is estimated as

$$P_F \approx p_0^{m-1} \frac{1}{N_s} \sum_{i=1}^{N_s} I_F\left(\boldsymbol{\theta}_{m-1,i}\right)$$

   where $\{\boldsymbol{\theta}_{m,i} : i = 1,\ldots,N_s\}$ is the set of samples generated at the last stage of subset simulation (conditional level $m-1$).

Note 1: The generation of samples at Step 1 and the computation of their utility function values are completely parallelizable. The generation of each group of N_{mc} Markov chains is also parallelizable.

Note 2: If the algorithm is used to estimate the threshold values up to a certain failure probability, the procedure must complete a prescribed number of simulation stages.

Note 3: The following numerical values have been considered in this work: $N_s = 1000$, $p_0 = 0.1$, $N_{\text{mc}} = 10$, $R_U = 0.50$, $\gamma_1 = 0.80$, $R_L = 0.70$, $\gamma_2 = 1.20$. These parameters are problem dependent and, therefore, can be modified for other applications.

Appendix B

RELIABILITY SENSITIVITY ESTIMATES

B.1 Failure probability in terms of distribution parameters

Note that, according to Equation (3.2.3), the failure probability can be written in terms of the distribution parameters $\boldsymbol{\tau}$ as

$$\begin{aligned} P_F(\boldsymbol{\tau}) &= p_0^{m-1} \int_{\boldsymbol{\theta}\in\Omega_\theta} I_F(\boldsymbol{\theta}) p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}) d\boldsymbol{\theta} \\ &= p_0^{m-1} \int_{\boldsymbol{\theta}\in\Omega_\theta} I_F(\boldsymbol{\theta}) \frac{p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau})}{p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}^0)} p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}^0) d\boldsymbol{\theta} \end{aligned} \tag{B.1.1}$$

where $p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau})$ and $p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}^0)$ are the distributions of $\boldsymbol{\theta}$ conditional to F_{m-1} and distribution parameters $\boldsymbol{\tau}$ and $\boldsymbol{\tau}^0$, respectively. By definition, these conditional distributions are equal to

$$p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}) = \frac{I_{F_{m-1}}(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\tau})}{P_{F_{m-1}}(\boldsymbol{\tau})} \ , \quad p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}^0) = \frac{I_{F_{m-1}}(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\tau}^0)}{P_{F_{m-1}}(\boldsymbol{\tau}^0)} \tag{B.1.2}$$

where $P_{F_{m-1}}(\boldsymbol{\tau})$ and $P_{F_{m-1}}(\boldsymbol{\tau}^0)$ are the probabilities of the failure event F_{m-1} under distribution parameter vectors $\boldsymbol{\tau}$ and $\boldsymbol{\tau}^0$, respectively. Substituting $p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau})$ and $p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}^0)$ into Equation (B.1.1) and noting that $P_{F_{m-1}}(\boldsymbol{\tau}) = P_{F_{m-1}}(\boldsymbol{\tau}^0) = p_0^{m-1}$ by

construction, the probability of failure can be rewritten as [64]

$$P_F(\boldsymbol{\tau}) = p_0^{m-1} \int_{\boldsymbol{\theta}\in\Omega_\theta} I_F(\boldsymbol{\theta}) \frac{p(\boldsymbol{\theta};\boldsymbol{\tau})}{p(\boldsymbol{\theta};\boldsymbol{\tau}^0)} p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}^0) d\boldsymbol{\theta} \tag{B.1.3}$$

B.2 Reliability sensitivity estimation

The only term depending on the distribution parameters in Equation (B.1.3) is $p(\boldsymbol{\theta};\boldsymbol{\tau})$. According to the approach introduced in [64], such characterization of the failure probability in the subset simulation framework allows to compute the partial derivative defined in Equation (3.3.1) as

$$\left.\frac{\partial P_F(\boldsymbol{\tau})}{\partial \tau_j}\right|_{\boldsymbol{\tau}^0} = p_0^{m-1} \int_{\boldsymbol{\theta}\in\Omega_\theta} I_F(\boldsymbol{\theta}) \frac{\frac{\partial p}{\partial \tau_j}(\boldsymbol{\theta};\boldsymbol{\tau}^0)}{p(\boldsymbol{\theta};\boldsymbol{\tau}^0)} p(\boldsymbol{\theta}|F_{m-1};\boldsymbol{\tau}^0) d\boldsymbol{\theta} \tag{B.2.1}$$

under the assumption that τ_j does not affect the integration domain. The above expression can be estimated using the simulation results as

$$\left.\frac{\partial P_F(\boldsymbol{\tau})}{\partial \tau_j}\right|_{\boldsymbol{\tau}^0} \approx p_0^{m-1} \frac{1}{N_m} \sum_{i=1}^{N_m} I_F\left(\boldsymbol{\theta}_{m-1,i}^0\right) \frac{\frac{\partial p}{\partial \tau_j}(\boldsymbol{\theta}_{m-1,i}^0;\boldsymbol{\tau}^0)}{p(\boldsymbol{\theta}_{m-1,i}^0;\boldsymbol{\tau}^0)} \tag{B.2.2}$$

where $\{\boldsymbol{\theta}_{m-1,i}^0, i = 1,\ldots,N_m\}$ is the set of samples generated at the last stage of subset simulation under distribution parameter vector $\boldsymbol{\tau}^0$ of the probability density function $p(\boldsymbol{\tau})$. This means that a single run of the method is required to estimate the reliability sensitivity with respect to the distribution parameters. In other words, the reliability sensitivity estimation is a post-process of the simulation procedure: it does not require any other solution of the water distribution network. This formulation can be explicitly written for different probability density functions $p(\boldsymbol{\theta};\boldsymbol{\tau})$ and different distribution parameters τ_j. Besides, since the formulation presented in equations (B.2.1) and (B.2.2) is based on subset simulation, a single run of the algorithm provides estimates for the reliability sensitivity corresponding to different thresholds. In this way, the approach provides a full characterization of the reliability sensitivity trend with respect to different threshold levels [64, 65].

B.3 Reliability sensitivity estimator for particular cases

The partial derivative of the failure probability with respect to the mean value μ_{θ_j} and standard deviation σ_{θ_j} of the system parameter θ_j evaluated at $\mu^0_{\theta_j}$ and $\sigma^0_{\theta_j}$ for the case of a normal random variable are estimated as

$$\left.\frac{\partial P_F}{\partial \mu_{\theta_j}}\right|_{\boldsymbol{\tau}^0} \approx p_0^{m-1}\frac{1}{N_m}\sum_{i=1}^{N_m} I_F\left(\boldsymbol{\theta}^0_{m-1,i}\right) \times \left\{\frac{\left(\theta^0_{m-1,ij}-\mu^0_{\theta_j}\right)}{{\sigma^0_{\theta_j}}^2}\right\} \tag{B.3.1}$$

and

$$\left.\frac{\partial P_F}{\partial \sigma_{\theta_j}}\right|_{\boldsymbol{\tau}^0} \approx p_0^{m-1}\frac{1}{N_m}\sum_{i=1}^{N_m} I_F\left(\boldsymbol{\theta}^0_{m-1,i}\right) \times \left\{\frac{\left(\theta^0_{m-1,ij}-\mu^0_{\theta_j}\right)^2}{{\sigma^0_{\theta_j}}^2}-1\right\}\frac{1}{\sigma^0_{\theta_j}} \tag{B.3.2}$$

where $\{\boldsymbol{\theta}^0_{m-1,i}, i=1,\ldots,N_m\}$ is the set of samples generated at the last stage of subset simulation (conditional level m) under the distribution $p(\boldsymbol{\theta}|\boldsymbol{\tau^0})$, and $\theta^0_{m-1,ij}$ is the j-th component of the sample vector $\boldsymbol{\theta}^0_{m-1,i}$.

For the case of a log-normal random variable, the estimators are written as

$$\begin{aligned}\left.\frac{\partial P_F}{\partial \mu_{\theta_j}}\right|_{\boldsymbol{\tau}^0} \approx p_0^{m-1}\frac{1}{N_m}\sum_{i=1}^{N_m} I_F\left(\boldsymbol{\theta}^0_{m-1,i}\right) \times \Bigg\{&\left[\frac{\ln(\theta^0_{m-1,ij})-\mu_j}{\sigma_j^2}\right]\alpha_j \\ &+\left[\frac{\left(\ln(\theta^0_{m-1,ij})-\mu_j\right)^2}{\sigma_j^2}-1\right]\frac{1}{\sigma_j^2}\beta_j\Bigg\}\end{aligned} \tag{B.3.3}$$

and

$$\begin{aligned}\left.\frac{\partial P_F}{\partial \sigma_{\theta_j}}\right|_{\boldsymbol{\tau}^0} \approx p_0^{m-1}\frac{1}{N_m}\sum_{i=1}^{N_m} I_F\left(\boldsymbol{\theta}^0_{m-1,i}\right) \times \Bigg\{&\left[\frac{\ln(\theta^0_{m-1,ij})-\mu_j}{\sigma_j^2}\right]\lambda_j \\ &-\left[\frac{\left(\ln(\theta^0_{m-1,ij})-\mu_j\right)^2}{\sigma_j^2}-1\right]\frac{1}{\sigma_j^2}\lambda_j\Bigg\}\end{aligned} \tag{B.3.4}$$

where

$$\alpha_j = \left[2 - \exp\left(-\sigma_j^2\right)\right] \exp\left(-\left(\mu_j + \sigma_j^2/2\right)\right) \tag{B.3.5}$$

$$\beta_j = \left[\exp\left(-\sigma_j^2\right) - 1\right] \exp\left(-\left(\mu_j + \sigma_j^2/2\right)\right) \tag{B.3.6}$$

$$\lambda_j = -\left[\exp\left(\sigma_j^2\right) - 1\right]^{1/2} \exp\left(-\left(\mu_j + 3\sigma_j^2/2\right)\right) \tag{B.3.7}$$

with

$$\mu_j = \ln\left(\left(\mu_{\theta_j}^0\right)^2 \Big/ \sqrt{\left(\mu_{\theta_j}^0\right)^2 + \left(\sigma_{\theta_j}^0\right)^2}\right), \quad \sigma_j = \sqrt{\ln\left(1 + \left(\sigma_{\theta_j}^0/\mu_{\theta_j}^0\right)^2\right)} \tag{B.3.8}$$

and all other terms have been previously defined.

Appendix C

TRANSITIONAL MARKOV CHAIN MONTE CARLO

The transitional Markov chain Monte Carlo algorithm [4] is an advanced simulation technique implemented in this work to address detection-related problem in the context of a probabilistic framework. The stochastic simulation method, initially developed in the context of structural mechanics, allows to handle several problems arising in the Bayesian model updating framework. In what follows, the main idea and key features of the algorithm are discussed, and a pseudo-code with the actual implementation is provided.

C.1 Main idea

Consider a probabilistic model class M parametrized by some parameters $\boldsymbol{\theta} \in \mathbb{R}^{n_p}$, such that $M(\boldsymbol{\theta})$ represents a particular model defined by the specific value of $\boldsymbol{\theta}$. These parameters are characterized in a probabilistic manner [51]. Also, assume that some data D about the behavior of the real system is available. For instance, measurement data obtained by monitoring sensors. The goal of simulation-based Bayesian model updating is to obtain a sample from the posterior distribution

$$p\left(\boldsymbol{\theta}|M,D\right) = \frac{p\left(D|M,\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}|M\right)}{P[D|M]} \tag{C.1.1}$$

where $p(\boldsymbol{\theta}|M)$ is the prior distribution, $P[D|M]$ is the evidence of M, and $p(D|M,\boldsymbol{\theta})$ is the likelihood function, which is a probabilistic description of how plausible is to obtain data D from model $M(\boldsymbol{\theta})$. This task is very challenging, since in practical applications the posterior distribution $p(\boldsymbol{\theta}|M,D)$ is not a standard probability density function and it is only known up to the constant $P[D|M]$. An usual approach is the use of Markov chain Monte Carlo methods [66,67]. However, the previous methods cannot provide an estimate for the evidence and may fail when the posterior distribution is bimodal.

As already pointed out in Section 5.4, the main idea of the transitional Markov chain Monte Carlo (TMCMC) method is to obtain a posterior sample by means of a sequence of non-normalized intermediate distributions $\{p_j(\boldsymbol{\theta}) : j = 0,\ldots,m\}$ that converge to the posterior distribution [4]. This sequence is such that

$$p_j(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|M)p(D|M,\boldsymbol{\theta})^{\alpha_j}, \qquad j = 0,\ldots,m \tag{C.1.2}$$

where $0 = \alpha_0 < \cdots < \alpha_m = 1$ and $\propto$ means proportional. Note that $p_0(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|M)$ and $p_m(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|M,D)$. In other words, the method starts sampling from the prior distribution ($j = 0$), which is usually uniform, and ends sampling from the posterior distribution (up to a scaling constant). In this context, the parameter α_j can be interpreted as a measure of how much of the available information is used during the j-th stage. The construction of the sequence $\{p_j\}$ follows the approach introduced in [4] and is addressed in Section C.2 for completeness.

As a result of the simulation process, the algorithm produces a total of m sets of samples

$$\{\boldsymbol{\theta}_{jk} : k = 1,\ldots,N_j\} \qquad j = 0,\ldots,m \tag{C.1.3}$$

where $\boldsymbol{\theta}_{jk}$ is the k-th sample obtained at stage j, and N_j is the total number of samples generated at stage j. In this work, a constant sample size N_m is specified for all stages, that is $N_j = N_m$. The samples at stage $j = 1,\ldots,m$ are obtained by means of the Metropolis-Hastings algorithm [4,66,67], according to Section C.3.

The following features of this simulation algorithm are remarked: (1) the method starts sampling from the prior distribution, thus initially populating all the space of model parameters; (2) the simulation technique has been widely proved and tested in a number of applications, including globally and locally identifiable as well as strictly nonidentifiable

systems; (3) little calibration of the algorithm parameters is required; and (4) the construction of the algorithm allows to estimate the evidence of the model class as a by-product of the simulation, which is pivotal to solve the model class selection problem.

C.2 Selection of the intermediate distributions

In order to sample efficiently by using Markov chain Monte Carlo, the intermediate distributions must be chosen properly. Specifically, the change of the shape between two consecutive intermediate distributions should be small. To this end, the values for $\alpha_{j+1}, j = 0, \ldots, m-1$ are selected based on the samples obtained at stage j, in order to ensure a smooth transition between subsequent intermediate distributions. Consider the *plausibility weight* of the k-th sample obtained at stage j, w_{jk}, defined as

$$w_{jk} = \frac{p_{j+1}(\boldsymbol{\theta}_{jk})}{p_j(\boldsymbol{\theta}_{jk})} = \frac{p(\boldsymbol{\theta}_{jk}|M)p(D|M,\boldsymbol{\theta}_{jk})^{\alpha_{j+1}}}{p(\boldsymbol{\theta}_{jk}|M)p(D|M,\boldsymbol{\theta}_{jk})^{\alpha_j}} = p(D|M,\boldsymbol{\theta}_{jk})^{\alpha_{j+1}-\alpha_j} \tag{C.2.1}$$

The degree of uniformity of the plausibility weights is a good indicator of how similar are the shapes of $p_{j+1}(\boldsymbol{\theta})$ and $p_j(\boldsymbol{\theta})$. Then, the value of α_{j+1} is chosen so that the coefficient of variation of the plausibility weights is equal to a prescribed threshold [4]. In the actual implementation, the value of α_{j+1} is selected such that

$$\frac{\sigma_{wj}}{\mu_{wj}} = 1.00 \tag{C.2.2}$$

where

$$\mu_{wj} = \frac{1}{N_m}\sum_{k=1}^{N_m} w_{jk} \tag{C.2.3}$$

$$\sigma_{wj} = \sqrt{\frac{1}{N_m - 1}\sum_{k=1}^{N_m}\left(w_{jk} - \mu_{wj}\right)^2} \tag{C.2.4}$$

Note that if Equation (C.2.2) holds for $\alpha_{j+1} > 1$, then it is imposed that $m = j+1$ and $\alpha_{j+1} = 1$ (end of the simulation procedure). It is also remarked that the equation can be solved by using any suitable numerical technique. In particular, the bisection method is considered in this work.

C.3 Generation of intermediate samples

Once the intermediate distribution has been determined, i.e. the value of $\alpha_{j+1}, j = 0, \ldots, m-1$ has been defined, the Metropolis-Hastings algorithm [66,67] is applied to generate the samples $\{\boldsymbol{\theta}_{j+1,k}, k = 1, \ldots, N_m\}$ based on the samples from stage j. The *seeds* of the Markov chains are selected from samples generated at stage j (previous stage) according to a probability equal to their *normalized plausibility weight*, which is a measure of the plausibility that the sample $\boldsymbol{\theta}_{jk}$ be distributed according to $p_{j+1}(\boldsymbol{\theta})$. These normalized weights are explicitly defined as

$$\overline{w}_{jk} = \frac{w_{j,k}}{\sum_{\iota=1}^{N_m} w_{j,\iota}} \tag{C.3.1}$$

Each selected seed is the initial state of a Markov chain. If some sample from stage j is selected more than once during the simulation procedure, the last state of the corresponding Markov chain must be considered. The proposal probability density function is a Gaussian distribution centred at the current state of the chain, with covariance matrix $\boldsymbol{\Sigma}_j$ equal to a scaled version of the sample covariance at stage j, that is

$$\boldsymbol{\Sigma}_j = \beta^2 \sum_{k=1}^{N_m} \overline{w}_{jk} \left(\boldsymbol{\theta}_{jk} - \overline{\boldsymbol{\theta}}_j\right) \left(\boldsymbol{\theta}_{jk} - \overline{\boldsymbol{\theta}}_j\right)^T \tag{C.3.2}$$

$$\overline{\boldsymbol{\theta}}_j = \sum_{k=1}^{N_m} \overline{w}_{jk} \boldsymbol{\theta}_{jk} \tag{C.3.3}$$

where β^2 is a scaling parameter that determines the spread of the proposal distribution [4]. In this work, $\beta^2 = 0.04$. Note that this setting produces a set of Markov chains which are perfectly parallel, providing computational advantages in this regard.

C.4 Evidence estimate

It can be proven [4] that the expected value of w_{jk} is

$$E\left[w_{jk}\right] = \frac{\int f(\boldsymbol{\theta}|M) f(D|M,\boldsymbol{\theta})^{\alpha_{j+1}} d\boldsymbol{\theta}}{\int f(\boldsymbol{\theta}|M) f(D|M,\boldsymbol{\theta})^{\alpha_j} d\boldsymbol{\theta}} \tag{C.4.1}$$

Based on the above, μ_{wj} is an asymptotically unbiased estimator for the quotient

$\int f(\boldsymbol{\theta}|M)f(D|M,\boldsymbol{\theta})^{\alpha_{j+1}}d\boldsymbol{\theta}\big/\int f(\boldsymbol{\theta}|M)f(D|M,\boldsymbol{\theta})^{\alpha_j}d\boldsymbol{\theta}$ and therefore

$$S=\prod_{j=0}^{m-1}\mu_{wj}=\prod_{j=0}^{m-1}\left[\frac{1}{N_m}\sum_{k=1}^{N_m}w_{jk}\right] \tag{C.4.2}$$

is an asymptotically unbiased estimator for $P[D|M]$, that is $P[D|M]\approx S$. Thus, the algorithm provides an estimate of the evidence based on the plausibility weights obtained during the simulation.

C.5 Basic pseudo-code

1. Define β^2. Set $j=0$ and $\alpha_j=0$. Obtain a sample $\{\boldsymbol{\theta}_{0,k}:k=1,\dots N_m\}$ distributed according to the prior distribution $p(\boldsymbol{\theta}|M)$. Compute the likelihood values $\{f_{j,k}=p(D|M,\boldsymbol{\theta}_{j,k}):k=1,\dots N_m\}$. This is equivalent to direct Monte Carlo simulation.

2. Compute α^* such that

$$\frac{\sigma_w}{\mu_w}=1$$

where

$$\mu_w=\frac{1}{N_m}\sum_{k=1}^{N_m}f_{j,k}^{\alpha^*-\alpha_j},\qquad\qquad \sigma_w=\sqrt{\frac{1}{N_m-1}\sum_{k=1}^{N_m}\left(f_{j,k}^{\alpha^*-\alpha_j}-\mu_w\right)^2}$$

3. Set $\alpha_{j+1}=\min(1,\alpha^*)$. Compute

$$w_{j,k}=f_{j,k}^{\alpha_{j+1}-\alpha_j},\qquad\qquad S_j=\frac{1}{N_m}\sum_{k=1}^{N_m}w_{j,k},\qquad\qquad \bar{w}_{j,k}=\frac{w_{j,k}}{N_mS_j}$$

and

$$\bar{\boldsymbol{\theta}}_j=\sum_{k=1}^{N_m}\bar{w}_{j,k}\boldsymbol{\theta}_{j,k},\qquad\qquad [\Sigma]_j=\beta^2\sum_{k=1}^{N_m}\bar{w}_{j,k}\left(\boldsymbol{\theta}_{j,k}-\bar{\boldsymbol{\theta}}_j\right)\left(\boldsymbol{\theta}_{j,k}-\bar{\boldsymbol{\theta}}_j\right)^T$$

4. Set $\{\boldsymbol{\theta}_{j,k}^{\text{loc}}=\boldsymbol{\theta}_{j,k}:k=1,\dots,N_m\}$ and $\{f_{j,k}^{\text{loc}}=f_{j,k}:k=1,\dots,N_m\}$. These sets are used to keep track of the evolution of each Markov chain.

5. Apply the Metropolis-Hastings algorithm to generate N_m samples distributed according to the $(j+1)$-th distribution $p_{j+1}(\boldsymbol{\theta})\propto p(\boldsymbol{\theta}|M)p(D|M,\boldsymbol{\theta})^{\alpha_{j+1}}$.

- For $k = 1$ to N_m:

  (a) Select the ν-th Markov chain from the set $\{1, 2, \ldots, N_m\}$ according to their probability mass equal to the *normalized weights* $\{\bar{w}_{j,k} : k = 1, \ldots, N_m\}$. Select a *lead sample* $\boldsymbol{\theta}^{\text{lead}} = \boldsymbol{\theta}_{j,\nu}^{\text{loc}}$ and the corresponding likelihood value as $f^{\text{lead}} = f_{j,\nu}^{\text{loc}}$.

  (b) Generate a *candidate sample* $\boldsymbol{\theta}^{\text{cand}}$ from a multivariate normal distribution with covariance matrix $[\Sigma]_j$ and centred at $\boldsymbol{\theta}^{\text{lead}}$. Compute $f^{\text{cand}} = p(D|M, \boldsymbol{\theta}^{\text{lead}})$.

  (c) Compute the quotient

  $$\Upsilon = \frac{(f^{\text{cand}})^{\alpha_{j+1}}\, p(\boldsymbol{\theta}^{\,\text{cand}}|M)}{(f^{\text{lead}})^{\alpha_{j+1}}\, p(\boldsymbol{\theta}^{\,\text{lead}}|M)}$$

  and generate a random number $\xi \in [0, 1]$ from a uniform distribution.

  (d) If $\xi \leq \min\{\Upsilon, 1\}$, set $\boldsymbol{\theta}_{j+1,k} = \boldsymbol{\theta}^{\,\text{cand}}$, $f_{j+1,k} = f^{\,\text{cand}}$ and update the last element of the current Markov chain as $\boldsymbol{\theta}_{j,\nu}^{\text{loc}} = \boldsymbol{\theta}^{\,\text{cand}}$ and $f_{j,\nu}^{\text{loc}} = f^{\,\text{cand}}$. Otherwise, set $\boldsymbol{\theta}_{j+1,k} = \boldsymbol{\theta}^{\,\text{lead}}$ and $f_{j+1,k} = f^{\,\text{lead}}$.

6. If $\alpha_{j+1} < 1$, set $j \leftarrow j + 1$ and go back to step 2. Otherwise, set $m = j + 1$ and compute the *evidence estimator* as

$$S = \prod_{j=0}^{m-1} S_j$$

C.6 Actual Implementation

For some cases, the direct implementation of the modified TMCMC algorithm according to the previous pseudo-code presents some practical issues. In particular, extremely large likelihood values may affect (numerically) the computation of the normalized weights and, therefore, the generation of candidate samples. In order to avoid this type of problems, an alternative computation of the weights is considered in this work. This setting does not affect any essential feature of the original algorithm [4]. Assume that a routine $\mathcal{L}$ to compute the logarithm of the likelihood function (log-likelihood) is available, that is

$$\mathcal{L}(D|M, \boldsymbol{\theta}) = \ln\left(p(D|M, \boldsymbol{\theta})\right) \tag{C.6.1}$$

Now, considering the samples $\{\boldsymbol{\theta}_{j,k} : k = 1, \ldots, N_m\}$ generated at the k-th stage, the

maximum of the corresponding log-likelihood values is

$$\mathcal{L}^* = \max_{k=1,\ldots,N_m} \mathcal{L}_{j,k} \tag{C.6.2}$$

where $\mathcal{L}_{j,k} = \mathcal{L}(D|M, \boldsymbol{\theta}_{j,k})$. Then, the *scaled plausibility weights* $w^*_{j,k}$ are defined as

$$w^*_{j,k} = \frac{w_{j,k}}{\max\limits_{\iota=1,\ldots,N_m} w_{j,\iota}} = \exp\left[(\alpha_{j+1} - \alpha_j)(\mathcal{L}_{j,k} - \mathcal{L}^*)\right] \tag{C.6.3}$$

In this work, the scaled plausibility weights $w^*_{j,k}$ are computed in each stage of the simulation procedure, rather than the plausibility weights $w_{j,k}$. In this manner, the normalized weights are computed as $\bar{w}_{j,k} = w^*_{j,k} \Big/ \sum_{\iota=1}^{N_m} w^*_{j,\iota}$. Besides, the logarithm of the mean value of the plausibility weights is given by

$$\ln(S_j) = \ln\left(\frac{1}{N_m}\sum_{k=1}^{N_m} w^*_{j,k}\right) + (\alpha_{j+1} - \alpha_j)\mathcal{L}^* \tag{C.6.4}$$

and the logarithm of the evidence estimate is

$$\ln(S) = \sum_{j=0}^{m-1} \ln(S_j) \tag{C.6.5}$$

In what follows, the actual pseudo-code implemented in this work is provided.

Pseudo-code:

1. Define β^2. Set $j = 0$ and $\alpha_j = 0$. Obtain a sample $\{\boldsymbol{\theta}_{0,k} : k = 1, \ldots N_m\}$ distributed according to the prior distribution $p(\boldsymbol{\theta}|M)$. Compute the log-likelihood values $\{\mathcal{L}_{j,k} = \mathcal{L}(D|M, \boldsymbol{\theta}_{j,k}) : k = 1, \ldots N_m\}$. This is equivalent to direct Monte Carlo simulation.

2. Define $\mathcal{L}^* = \max\limits_{k=1,\ldots,N_m} \mathcal{L}_{j,k}$. Compute α^* such that

$$\frac{\sigma_w}{\mu_w} = 1$$

where

$$\mu_w = \frac{1}{N_m}\sum_{k=1}^{N_m} \exp\left\{(\alpha^* - \alpha_j)(\mathcal{L}_{j,k} - \mathcal{L}^*)\right\}$$

$$\sigma_w = \sqrt{\frac{1}{N_m - 1}\sum_{k=1}^{N_m}\left(\exp\left\{(\alpha^* - \alpha_j)(\mathcal{L}_{j,k} - \mathcal{L}^*)\right\} - \mu_w\right)^2}$$

3. Set $\alpha_{j+1} = \min(1, \alpha^*)$. Compute:

$$w^*_{j,k} = \exp\left\{(\alpha^* - \alpha_j)(\mathcal{L}_{j,k} - \mathcal{L}^*)\right\}, \qquad \bar{w}_{j,k} = \frac{w^*_{j,k}}{\sum_{\iota=1}^{N_m} w^*_{j,\iota}}$$

$$\ln(S_j) = \ln\left(\frac{1}{N_m}\sum_{k=1}^{N_m} w^*_{j,k}\right) + (\alpha_{j+1} - \alpha_j)\mathcal{L}^*$$

and

$$\bar{\boldsymbol{\theta}}_j = \sum_{k=1}^{N_m} \bar{w}_{j,k}\boldsymbol{\theta}_{j,k}, \qquad [\Sigma]_j = \beta^2 \sum_{k=1}^{N_m} \bar{w}_{j,k}\left(\boldsymbol{\theta}_{j,k} - \bar{\boldsymbol{\theta}}_j\right)\left(\boldsymbol{\theta}_{j,k} - \bar{\boldsymbol{\theta}}_j\right)^T$$

4. Set $\{\boldsymbol{\theta}^{\text{loc}}_{j,k} = \boldsymbol{\theta}_{j,k} : k = 1, \ldots, N_m\}$ and $\{\mathcal{L}^{\text{loc}}_{j,k} = \mathcal{L}_{j,k} : k = 1, \ldots, N_m\}$. These sets are used to keep track of the evolution of each Markov chain.

5. Apply the Metropolis-Hastings algorithm to generate N_m samples distributed according to the $(j+1)$-th distribution $p_{j+1}(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|M)p(D|M, \boldsymbol{\theta})^{\alpha_{j+1}}$.

   • For $k = 1$ to N_m:

   (a) Select the ν-th Markov chain from the set $\{1, 2, \ldots, N_m\}$ according to their probability mass equal to the *normalized weights* $\{\bar{w}_{j,k} : k = 1, \ldots, N_m\}$. Select a *lead sample* $\boldsymbol{\theta}^{\text{lead}} = \boldsymbol{\theta}_{j,\nu}$ and $\mathcal{L}^{\text{lead}} = \mathcal{L}^{\text{loc}}_{j,\nu}$.

   (b) Generate a *candidate sample* $\boldsymbol{\theta}^{\text{cand}}$ from a multivariate normal distribution with covariance matrix $[\Sigma]_j$ and centred at $\boldsymbol{\theta}^{\text{lead}}$. If $p(\boldsymbol{\theta}^{\text{cand}}|M) = 0$, set $\Upsilon = 1$, $\ln(\xi) = -1$ and go to Step 5-(d). Otherwise, compute $\mathcal{L}^{\text{cand}} = \mathcal{L}(D|M, \boldsymbol{\theta}^{\text{lead}})$.

   (c) Compute the quantity

$$\ln(\Upsilon) = \alpha_{j+1}\left(\mathcal{L}^{\text{cand}} - \mathcal{L}^{\text{lead}}\right) + \ln\left(p(\boldsymbol{\theta}^{\text{cand}}|M)\right) - \ln\left(p(\boldsymbol{\theta}^{\text{lead}}|M)\right)$$

   and generate a random number $\xi \in [0, 1]$ from a uniform distribution.

   (d) If $\ln(\xi) \leq \min\{\ln(\Upsilon), 0\}$, set $\boldsymbol{\theta}_{j+1,k} = \boldsymbol{\theta}^{\,\text{cand}}$, $\mathcal{L}_{j+1,k} = \mathcal{L}^{\,\text{cand}}$ and update the last element of the current Markov chain as $\boldsymbol{\theta}^{\text{loc}}_{j,\nu} = \boldsymbol{\theta}^{\,\text{cand}}$ and $\mathcal{L}^{\text{loc}}_{j,\nu} = \mathcal{L}^{\,\text{cand}}$. Otherwise, set $\boldsymbol{\theta}_{j+1,k} = \boldsymbol{\theta}^{\,\text{lead}}$ and $\mathcal{L}_{j+1,k} = \mathcal{L}^{\,\text{lead}}$.

6. If $\alpha_{j+1} < 1$, set $j \leftarrow j + 1$ and go back to step 2. Otherwise, set $m = j + 1$ and compute the logarithm of the *evidence estimator* as

$$\ln(S) = \sum_{j=0}^{m-1} \ln(S_j)$$

Appendix D

HYDRAULIC MODELING

Water distribution networks are spatially distributed systems that consist of several physical components (pipes, reservoirs, tanks, pumps, valves) developing complex, non-trivial interactions in order to fulfill the requirements of users. This appendix describes the fundamental physical principles considered to model the hydraulic system, as well as the formulation of the numerical method included in the hydraulic simulator implemented in this work (i.e., EPANET).

D.1 Physical principles

The engineering analysis of water distribution networks under steady-state conditions is based on two fundamental principles: mass conservation equations, and energy conservation equations. These relationships are described in what follows.

D.1.1 Mass conservation equations

The mass conservation principle implies, in the context of pressurized pipe networks, that the flow of water entering at some control point of a hydraulic system is equal to the total water outflow at the same point. Specifically, the sum of all water flows Q_{ij} into the j-th node of the pressurized distribution network must be equal to the nodal demand q_j

at that point, that is

$$\sum_j Q_{ij} - q_j = 0, \quad \forall j = 1, \ldots, N_{\text{node}} \tag{D.1.1}$$

where, by convention, Q_{ij} positive corresponds to flow from node i to node j.

D.1.2 Energy conservation equations

The principle of energy conservation is applied to each of the N_L link elements (pipes, vales, and pumps) included in the network. Specifically, if the p-th link of the network is defined from node I_p to node J_p, the head difference between such interconnected nodes must be equal to the headloss (negative of the head gain) in the corresponding link element, that is

$$H_{I_p} - H_{J_p} = \Delta h_p^s + \Delta h_p^f, \quad \forall p = 1, \ldots, N_L \tag{D.1.2}$$

where H_{I_p} is the piezometric head at node I_p, Δh_p^s are the minor headlosses in the p-th link element, and Δh_p^f are the corresponding friction headlosses. It is remarked that relationships (D.1.2) are sufficiently general to deal with all the link types included in EPANET, such as pipes, valves, and pumps [26]. In the specific case of pipe elements, the minor headlosses in the p-th pipeline are defined as

$$\Delta h_p^s = m_p^s |Q_p| Q_p \tag{D.1.3}$$

where m_p^s is the minor loss coefficient corresponding to pipe p, and $Q_p = Q_{I_p J_p}$. On the other hand, the friction headlosses are defined by relationships of the form

$$\Delta h_p^f = r_p^f |Q_p|^{n_p^f - 1} Q_p \tag{D.1.4}$$

where r_p^f is the resistance coefficient of pipe p, and n_p^f is the corresponding flow exponent. Several pipe headloss relationships can be implemented, such as the Darcy-Weisbach or the Chezy-Manning formulae. However, this work considers the use of the Hazen-Williams formula, which is given by

$$\Delta h_p^f = \frac{10.667 L_p}{D_p^{4.871} C_p^{1.852}} |Q_p|^{0.852} Q_p \tag{D.1.5}$$

where L_p, D_p, and C_p are the length, diameter, and Hazen-Williams coefficient, respectively, of the p-th pipe defined from node I_p to node J_p. The previous equation assumes

that all quantities are expressed in units from the International System of Units. Note that the hydraulic losses in pipe p decrease monotonically with the coefficient C_p, that is, increasing the value of the Hazen-Williams coefficient increases the hydraulic capacity of the corresponding pipe. Additionally, observe that any pair of values (D_p, C_p) giving the same denominator value in Equation (D.1.5) represents the same hydraulic behavior in terms of friction headlosses.

D.2 Numerical solution: Gradient method

Equations (D.1.1) and (D.1.2) represent a set of nonlinear equations where the unknowns are the pipe flows and the nodal heads. The Gradient method [84, 88] is implemented in the EPANET framework to obtain the network response at a given time (solution of the system of nonlinear equations). Such a iterative algorithm is chosen because of its global convergence property and its abiliity to handle large-scale systems very efficiently. Specifically, the approach introduced in [88] is implemented in the software. For completeness, the main ideas of the method are reproduced here. The reader is referred to [26] for further implementation details.

Consider a water network with N_J junction nodes, N_0 fixed-head nodes (tanks and reservoirs), and N_L links (pipes, valves, pumps). Note that the total number of nodes in the network is $N_{\text{node}} = N_J + N_0$. Then, equations (D.1.1) and (D.1.2) can be written in matrix form as

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{Q} \\ \mathbf{H} \end{Bmatrix} = \begin{Bmatrix} -\mathbf{B}_0\mathbf{H}_0 \\ \mathbf{q} \end{Bmatrix} \tag{D.2.1}$$

where $\mathbf{Q} \in \mathbb{R}^{N_L}$ is the vector containing the flowrates in each link, $\mathbf{H} \in \mathbb{R}^{N_J}$ is the vector of unknown nodal heads, $\mathbf{A} = \mathbf{A}(\mathbf{Q}) \in \mathbb{R}^{N_L \times N_L}$ is a diagonal matrix such that $\mathbf{AQ}$ is the vector containing the headlosses in each link, $\mathbf{B} \in \mathbb{R}^{N_L \times N_J}$ is the unknown-head nodes incidence matrix, $\mathbf{B}_0 \in \mathbb{R}^{N_L \times N_0}$ is the fixed-head nodes incidence matrix, and $\mathbf{H}_0 \in \mathbb{R}^{N_0}$ is the vector of fixed nodal heads. The unknown-head nodes incidence matrix is defined by

$$B_{ij} = \begin{cases} 1, & \text{if flow of pipe } i \text{ enters node } j \\ 0, & \text{if pipe } i \text{ and node } j \text{ are not connected} \\ -1, & \text{if flow of pipe } i \text{ leaves node } j \end{cases} \tag{D.2.2}$$

and the fixed-head nodes incidence matrix, or $\mathbf{A}_0$, is analogously defined. The numerical

solution of Equation (D.2.1) is based on a Newton-Raphson methodology [88]. The iterative scheme operates –in a simplified manner– as follows:

1. Define the tolerance ε and the initial estimate for the flow values $\mathbf{Q}^0$. Set $k = 0$.

2. Obtain the estimate for the head values $\mathbf{H}^{k+1}$ as the solution of the linear system

$$\left(\mathbf{B}^T\mathbf{D}^k\mathbf{B}\right)\mathbf{H}^{k+1} = \mathbf{F}^{k+1} \tag{D.2.3}$$

   where $\mathbf{D}^k \in \mathbb{R}^{N_L \times N_L}$ is a diagonal matrix whose diagonal terms are defined by

$$D_{pp}^k = \frac{1}{\left.\dfrac{\partial \Delta h_p}{\partial Q_p}\right|_{Q_p = Q_p^k}}, \qquad p = 1, \ldots, N_L \tag{D.2.4}$$

   and the right-hand side vector $\mathbf{F}^{k+1} \in \mathbb{R}^{N_L}$ is given by

$$\mathbf{F}^{k+1} = \left(\mathbf{B}^T\mathbf{Q}^k - \mathbf{q}\right) - \mathbf{B}^T\mathbf{D}^k\left(\mathbf{A}^k\mathbf{Q}^k + \mathbf{B}_0\mathbf{H}_0\right) \tag{D.2.5}$$

   with $\mathbf{A}^k = \mathbf{A}(\mathbf{Q}^k)$.

3. The estimate for the flow values is updated as

$$\mathbf{Q}^{k+1} = \mathbf{Q}^k - \mathbf{D}^k\mathbf{A}^k\mathbf{Q}^k - \mathbf{D}^k\left(\mathbf{B}_0\mathbf{H}_0 + \mathbf{B}\mathbf{H}^{k+1}\right) \tag{D.2.6}$$

4. If $\|\mathbf{Q}^{k+1} - \mathbf{Q}^k\|_1 > \varepsilon$, where $\|\cdot\|_1$ is the 1-norm operator, set $k \leftarrow k + 1$ and go to step 2. Otherwise, the solution of Equation (D.2.1) is given by $\mathbf{Q}^{k+1}$ and $\mathbf{H}^{k+1}$.

Note: The previous pseudo-code represents a basic version of the Gradient method. The actual implementation of this numerical technique in the EPANET framework, including the treatment of special network components such as pumps and valves, can be found in [26].

REFERENCES

[1] J. Birchmeier. Systematic assessment of the degree of criticality of infrastructures. In *Proceedings of ESREL 2007*, volume 1, pages 859–864, June 2007.

[2] E. Zio. Reliability Analysis of Complex Network Systems: Research and Practice in Need. *IEEE Reliability Society 2007 Annual Technology Report*, 2007.

[3] S.-K. Au and J.L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, 2001.

[4] J. Ching and Y.-C. Chen. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics*, 133(7):816–832, 2007.

[5] J. Ching and W.-C. Hsu. An efficient method for evaluating origin-destination connectivity reliability of real-world lifeline networks. *Computer-Aided Civil and Infrastructure Engineering*, 22(8):584–596, 2007.

[6] C. Xu and I.C. Goulter. Reliability-based optimal design of water distribution networks. *Journal of Water Resources Planning and Management*, 125(6):352–362, 1999.

[7] D.G. Yoo, D. Jung, D. Kang, and J.H. Kim. Seismic-Reliability-Based optimal layout of a water distribution network. *Water*, 8(2), 2016.

[8] K.M. Zuev, S. Wu, and J.L. Beck. General network reliability problem and its efficient solution by Subset Simulation. *Probabilistic Engineering Mechanics*, 40:25 – 35, 2015.

[9] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.

[10] M. Lemaire, A. Chateaumef, and J.-C. Mitteau. *Structural Reliability*. John Wiley and Sons, 2010.

[11] Y. Bao and L.W. Mays. Model for water distribution system reliability. *Journal of Hydraulic Engineering*, 116(9):1119–1137, 1990.

[12] L. Dueñas-Osorio and S.M. Vemuru. Cascading failures in complex infrastructure systems. *Structural Safety*, 31(2):157 – 167, 2009.

[13] N. Duan and L.W. Mays. Reliability analysis of pumping systems. *Journal of Hydraulic Engineering*, 116(2):230–248, 1990.

[14] J. Ghosh, K. Rokneddin, J.E. Padgett, and L. Dueñas-Osorio. Seismic reliability assessment of aging highway bridge networks with field instrumentation data and correlated failures, I: Methodology. *Earthquake Spectra*, 30(2):795–817, 2014.

[15] E. Kyriakides and M. Polycarpou. *Intelligent Monitoring, Control and Security of Critical Infrastructure Systems*. Springer-Verlag Berlin Heidelberg, 2015.

[16] E. Zio and G. Sansavini. Component criticality in failure cascade processes of network systems. *Risk Analysis*, 31(8):1196–1210, 2011.

[17] D.S. Kang, M.F.K. Pasha, and K. Lansey. Approximate methods for uncertainty analysis of water distribution systems. *Urban Water Journal*, 6(3):233–249, 2009.

[18] J.M. Torres, K. Brumbelow, and S.D. Guikema. Risk classification and uncertainty propagation for virtual water distribution systems. *Reliability Engineering & System Safety*, 94(8):1259 – 1273, 2009.

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

[20] A.J. Torii and R.H. Lopez. Reliability analysis of water distribution networks using the adaptive response surface approach. *Journal of Hydraulic Engineering*, 138(3):227–236, 2012.

[21] C. Xu, I. Goulter, and K. Tickle. Assessing the capacity reliability of ageing water distribution systems. *Civil Engineering and Environmental Systems*, 20(2):119–133, 2003.

[22] M.S. Islam, R. Sadiq, M.J. Rodriguez, H. Najjaran, and M. Hoorfar. Reliability assessment for water supply systems under uncertainties. *Journal of Water Resources Planning and Management*, 140(4):468–479, 2014.

[23] L.W. Mays. *Water distribution systems handbook*. McGraw-Hill, 2000.

[24] A. Ostfeld. Reliability analysis of water distribution systems. *Journal of Hydroinformatics*, 6(4):281, 2004.

[25] Y. Wang S.-K. Au. *Engineering risk assessment with Subset Simulation*. John Wiley & Sons Inc., 2014.

[26] L.A. Rossman. *EPANET 2 user manual and programmer's toolkits*. U.S. Environmental Protection Agency, Risk reduction engineering laboratory, Cincinati, USA, 2000.

[27] L.E. Johnson. *Geographic information systems in water resources engineering*. CRC Press, Boca Raton, FL, USA, 2009.

[28] D. Savic, G. Walters, P.G. Ashcroft, and A. Arscott. Hydroinformatics technology and maintenance of uk water networks. *Journal of Quality in Maintenance Engineering*, 3(4):289–301, 1997.

[29] J. Guan, M.M. Aral, M.L. Maslia, and W.M. Grayman. Identification of contaminant sources in water distribution systems using simulation-optimization method: case study. *Journal of Water Resources Planning and Management*, 132(4):252–262, 2006.

[30] A. Lambert and W. Hirner. *Losses from water supply systems: Standard terminology and recommended performance measures*. IWA, London, 2000.

[31] Informe de gestión del sector sanitario 2016. Technical report, Superintendencia de Servicios Sanitarios, Gobierno de Chile, 2016.

[32] R. Liemberger and M. Farley. Developing a non-revenue water reduction strategy, Part 1: Investigating and assessing water losses. In *IWA WWC 2004 Conference*, Marrakech, Morocco, 2004.

[33] H. Ramos, C. Reis, C. Ferreira, C. Falcao, and D. Covas. Leakage control policy within operating management tools. In *Proceedings of Water Software Systems: theory and applications*, Leicester, UK, 2001.

[34] Y. Gao, M.J. Brennan, and P.F. Joseph. On the effects of reflections on time delay estimation for leak detection in buried plastic water pipes. *Journal of Sound and Vibration*, 325(3):649–663, 2009.

[35] J.G. McNulty. An acoustic-based system for detecting, locating and sizing leaks in water pipelines. In *Proceedings of the 4th International Conference on Water Pipeline Systems: Managing Pipeline Assets in an Evolving Market*, York, UK, 2001.

[36] D.B. Field and B. Ratcliffe. Location of leaks in pressurised pipelines using sulphur hexafluoride as a tracer. Technical Report 80, Water Research Centre, 1978.

[37] R. Pilcher, S. Hamilton, H. Chapman, B. Ristovski, and S. Strapely. Leak location and repair guidance notes. In *International Water Association. Water Loss Task Forces: Specialist Group Efficient Operation and Management*, Bucharest, Romania, 2007.

[38] J.M. Muggleton, M.J. Brennan, R.J. Pinnington, and Y. Gao. A novel sensor for measuring the acoustic pressure in buried plastic water pipes. *Journal of Sound and Vibration*, 295(3-5):1085–1098, 2006.

[39] E. O'Brien, T. Murray, and A. McDonald. Detecting leaks from water pipes at a test facility using ground-penetrating radar. In *Proceedings of PEDS 2003 (Pumps, Electromechanical Devices and Systems Applied to Urban Management)*, Valencia, Spain, 2003.

[40] G. Moser, S.G. Paal, and I.F.C. Smith. Leak detection of water supply networks using error-domain model falsification. *Journal of Computing in Civil Engineering*, 32(2):04017077, 2018.

[41] M. Romano, Z. Kapelan, and D.A. Savić. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. *Journal of Hydroinformatics*, 15(3):634–651, 2013.

[42] S.R. Mounce, R.B. Mounce, and J.B. Boxall. Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of Hydroinformatics*, 13(4):672–686, 2011.

[43] J.-A. Goulet, S. Coutu, and I.F.C. Smith. Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. *Advanced Engineering Informatics*, 27(2):261–269, 2013.

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

[45] R. Puust, Z. Kapelan, D. Savic, and T. Koppel. Probabilistic leak detection in pipe networks using the SCEM-UA algorithm. In *Proceedings of the Water Distribution Systems Analysis Symposium 2006*, pages 1–2, Reston, VA, 2006. ASCE.

[46] M. Romano, Z. Kapelan, and D.A. Savić. Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management*, 140(4):457–467, 2014.

[47] M.A. Barandouzi, G. Mahinthakumar, R. Ranjithan, and E.D. Brill. Probabilistic mapping of water leakage characterizations using Bayesian approach. In *Proceedings of World Environmental and Water Resources Congress 2012*, pages 3248–3256, Reston, VA, 2012. ASCE.

[48] H. Wang and K.W. Harrison. Bayesian update method for contaminant source characterization in water distribution systems. *Journal of Water Resources Planning and Management*, 139(1):13–22, 2013.

[49] D.L. Boccelli and J.G. Uber. Incorporating spatial correlation in a Markov chain Monte Carlo approach for network model calibration. In *Proceedings of the World Water and Environmental Resources Congress 2005*, Reston, VA, 2005. ASCE.

[50] E.T. Jaynes. *Probability theory: The logic of science.* Cambridge University Press, Cambridge, 2003.

[51] J.L. Beck and L.S. Katafygiotis. Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics*, 124(4):455–461, 1998.

[52] L.S. Katafygiotis and J.L. Beck. Updating models and their uncertainties. II: Model identifiability. *Journal of Engineering Mechanics*, 124(4):463–467, 1998.

[53] K.V. Yuen. *Bayesian methods for structural dynamics and civil engineering.* John Wiley & Sons, 2010.

[54] W. Betz, I. Papaioannou, and D. Straub. Transitional markov chain monte carlo: Observations and improvements. *Journal of Engineering Mechanics*, 142(5):04016016, 2016.

[55] D. Straub and I. Papaioannou. Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, 141(3):04014134, 2015.

[56] P.S. Koutsourelakis, H.J. Pradlwarter, and G.I. Schuëller. Reliability of structures in high dimensions, part I: algorithms and applications. *Probabilistic Engineering Mechanics*, 19(4):409–417, 2004.

[57] M.A. Valdebenito, H.A. Jensen, H.B. Hernández, and L. Mehrez. Sensitivity estimation of failure probability applying line sampling. *Reliability Engineering & System Safety*, 171:99–111, 2018.

[58] S.-K. Au and J.L. Beck. Important sampling in high dimensions. *Structural Safety*, 25(2):139 – 163, 2003.

[59] L. Katafygiotis, T. Moan, and S.H. Cheungt. Auxiliary domain method for solving multi-objective dynamic reliability problems for nonlinear structures. *Structural Engineering and Mechanics*, 25(3):347–363, 2007.

[60] K.M. Zuev and L.S. Katafygiotis. The Horseracing Simulation algorithm for evaluation of small failure probabilities. *Probabilistic Engineering Mechanics*, 26(2):157 – 164, 2011.

[61] S.-K. Au and E. Patelli. Rare event simulation in finite-infinite dimensional space. *Reliability Engineering & System Safety*, 148(Supplement C):67 – 77, 2016.

[62] G.I. Schuëller and H.J. Pradlwarter. Benchmark study on reliability estimation in higher dimensions of structural systems – An overview. *Structural Safety*, 29(3):167 – 182, 2007.

[63] H.A. Jensen and M.A. Catalan. On the effects of non-linear elements in the reliability-based optimal design of stochastic dynamical systems. *International Journal of Non-Linear Mechanics*, 42(5):802 – 816, 2007.

[64] H.A. Jensen, F. Mayorga, and M.A. Valdebenito. Reliability sensitivity estimation of nonlinear structural systems under stochastic excitation: A simulation-based approach. *Computer Methods in Applied Mechanics and Engineering*, 289(Supplement C):1–23, 2015.

[65] H.A. Jensen, F. Mayorga, and C. Papadimitriou. Reliability sensitivity analysis of stochastic finite element models. *Computer Methods in Applied Mechanics and Engineering*, 296(Supplement C):327 – 351, 2015.

[66] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, jun 1953.

[67] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, apr 1970.

[68] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos. Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework. *The Journal of Chemical Physics*, 137(14):144103, 2012.

[69] M. Pellissetti. Parallel processing in structural reliability. *Structural Engineering and Mechanics*, 32:95–126, 1 2009.

[70] S.A. Zarghami, I. Gunawan, and F. Schultmann. Integrating entropy theory and cospanning tree technique for redundancy analysis of water distribution networks. *Reliability Engineering & System Safety*, 176:102 – 112, 2018.

[71] E. Simoen, C. Papadimitriou, and G. Lombaert. On prediction error correlation in Bayesian model updating. *Journal of Sound and Vibration*, 332(18):4136 – 4152, 2013.

[72] L.S. Katafygiotis, O. Sedehi, and F.R. Rofooei. Bayesian time-domain model updating considering correlation of prediction errors. In *12th International Conference on Structural Safety and Reliability*, Viena, Austria, 2017.

[73] E.J.M. Blokker, E.J. Pieterse-Quirijins, J.H.G. Vreeburg, and J.C. van Dijk. Simulating nonresidential water demand with a stochastic end-use model. *Journal of Water Resources Planning and Management*, 136(1):19–26, 2011.

[74] L. Liu, E.D. Brill, G. Mahinthakumar, J. Uber, E. M. Zechman, and S. Ranjithan. Considering demand variability and measurement uncertainties in adaptive source characterization in water distribution networks. In *ASCE World and Environmental Congress*, Reston, VA, 2007. ASCE.

[75] S. Bi, M. Broggi, and M. Beer. The role of the Bhattacharyya distance in stochastic model updating. *Mechanical Systems and Signal Processing*, 117:437 – 452, 2019.

[76] B. Goller, M. Broggi, A. Calvi, and G.I. Schuëller. A stochastic model updating technique for complex aerospace structures. *Finite Elements in Analysis and Design*, 47(7):739 – 752, 2011.

[77] H.A. Jensen, C. Vergara, C. Papadimitriou, and E. Millas. The use of updated robust reliability measures in stochastic dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 267(Supplement C):293 – 317, 2013.

[78] J. Ching and J.-S. Wang. Application of the transitional markov chain monte carlo algorithm to probabilistic site characterization. *Engineering Geology*, 203:151 – 167, 2016.

[79] J. Morrison. Managing leakage by district metered areas: A practical approach. *Water*, 21:44–46, 2004.

[80] J. Rougler. Probabilistic leak detection in pipelines using the mass imbalance approach. *Journal of Hydraulic Research*, 43(5):556–566, 2003.

[81] R. Puust, Z. Kapelan, D.A. Savic, and T. Koppel. A review of methods for leakage management in pipe networks. *Urban Water Journal*, 7(1):25–45, 2010.

[82] C. Papadimitriou, J.L. Beck, and S.-K. Au. Entropy-based optimal sensor location for structural model updating. *Journal of Vibration and Control*, 6(5):781–800, 2000.

[83] C. Malings and M. Pozzi. Value of information for spatially distributed systems: Application to sensor placement. *Reliability Engineering & System Safety*, 154:219–233, 2016.

[84] M. Potter and D.C. Wiggert. *Fluid Mechanics*. McGraw-Hill, 2008.

[85] P.-L. Liu and A. der Kiureghian. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, 1(2):105 – 112, 1986.

[86] L.D. Lutes and S. Sarkani. *Random vibrations: Analysis of structural and mechanical systems*. Elsevier Butterworth-Heinemann, Burlintong, MA, USA, 2004.

[87] K.M. Zuev, J.L. Beck, S.-K. Au, and L.S. Katafygiotis. Bayesian post-processor and other enhancements of subset simulation for estimating failure probabilities in high dimensions. *Computers & Structures*, 92:283 – 296, 2012.

[88] E. Todini and S. Pilati. A gradient method for the analysis of pipe networks. In *International Conference on Computer Applications for Water Supply and Distribution*, Leicester Polytechnic, UK, September 1987.