

ResiWater deliverable report D4.3: Addressing parameter and measurement uncertainties for water supply networks under abnormal operational conditions

Mathias Braun, Olivier Piller, Jochen Deuerlein, Fabrizio Parisini

▶ **To cite this version:**

Mathias Braun, Olivier Piller, Jochen Deuerlein, Fabrizio Parisini. ResiWater deliverable report D4.3: Addressing parameter and measurement uncertainties for water supply networks under abnormal operational conditions. [Research Report] irstea. 2017, pp.33. hal-02608635

HAL Id: hal-02608635
https://hal.inrae.fr/hal-02608635

Submitted on 16 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

http://www.resiwater.eu

RESIWATER

Deliverable Report D4.3

Written by Braun Mathias

INNOVATIVE SECURE SENSOR NETWORKS AND MODEL-BASED ASSESSMENT TOOLS FOR INCREASED RESILIENCE OF WATER INFRASTRUCTURES

Deliverable 4.3

Addressing parameter and measurement uncertainties for water supply networks under abnormal operational conditions

Dissemination level: Public

WP4

Robust Hydraulic Simulation Tools

12th December 2017

Contact persons:
Olivier PILLER Olivier.Piller@irstea.fr
Fereshte SEDEHIZADE Fereshte.Sedehizade@bwb.de

Project reference for France & for Germany: ANR-14-PICS-0003 & BMBF-13N13690

WP 4 – Robust Hydraulic Simulation Tools

D4.3 Study of Uncertainties and attainable Accuracy

List of Deliverable 4.3 contributors:

From Irstea

Olivier Piller (olivier.piller@irstea.fr)

Mathias Braun (mathias.braun@irstea.fr)

From 3S Consult

Jochen Deuerlein (deuerlein@3sconsult.de) [WP4 leader]

From EMS

Jean-Marc Weber (jean-marc.weber@strasbourg.eu)

Work package number	4.3	**Start date:**		01/12/2015
Contributors	Irstea	3S Consult	EMS	
Person-months per partner	30	3	1	
Keywords Parameter Uncertainty – Uncertainty Quantification – Confidence Limits				
Objectives Development of concept for estimation of the parameter uncertainty impacts on reliability of the simulation results.				

TABLES OF CONTENTS

1 Summary 5
2 Literature 6
3 Parameter Uncertainties 8
3.1 Random Variables 8
3.2 Sampling Strategies for Random Variables 8
4 Uncertainty propagation 11
4.1 Hydraulic Equations and Models 11
4.2 Propagation Algorithms 12
4.2.1 Perturbation Methods 13
4.2.2 Monte Carlo 14
4.2.3 Spectral Stochastic System 15
5 Evaluation of Uncertainties 17
5.1 Stochastic Moments 17
5.2 Marginalization 17
5.3 Estimation of the Probability Density Function 18
5.4 Confidence Intervals 18
6 Application 20
6.1 Network Models 20
6.2 Uncertainty Quantification Scenarios 21
6.2.1 Linear Tree Network 21
6.2.2 Small Looped Network 22
Medium Sized Network 26
7 Lessons learned 29
7.1 Evaluating the Expansion Order 29
7.2 Monte Carlo versus Non-Intrusive Spectral Projection 30
7.3 Intrusive versus Non-Intrusive Methods 30
8 Conclusion and next steps 31
9 References 32

LIST OF FIGURES:

Figure 1: (a) Hermite polynomials order N=4. **(**b) First order PCE with Gaussian germ. **10**
Figure 2: Tree graph network. **12**
Figure 3: Small looped network. **20**
Figure 4: Medium sized network model. **21**
Figure 5: Flow rate and head PDFs for the tree network in 1D demand uncertainty. **22**

Figure 6: Flow rate and head PDFs based on 1D uncertain demand for the small loop network. **23**
Figure 7: Flow rate and head PDFs based on 2D uncertain demands for the small loop network. **24**
Figure 8: Comparison of flow rate and head DDM and PDM PDFs based on 1D uncertain demand for the small loop network. **24**
Figure 9: Min, average and max water ages predictions for two nodes in the small looped network. **25**
Figure 10: Water age distributions for small network: (a) Meshed part of the system (b) at Branch in the network. **26**
Figure 11: Flow rate PDF based on 1D uncertain demand for VEDIF medium-sized network. **27**
Figure 12: Water age distributions for the large Network: (a) Meshed part of the system after 116 hours simulation time; (b) Meshed part of the system after 130 hours simulation time. **28**
Figure 12: Probability density function for the pressure at node 7 based on the uncertain demand. **29**

1 SUMMARY

The objective of work package 4.3 is the development of a concept for estimation of the parameter uncertainty impacts on the reliability of the simulation results. The major steps of work package 4.3 are:

- Survey and classification of parameter/measurement uncertainties (sources, range, importance);
- Development of a concept for estimation of the parameter uncertainty impacts on the reliability of the simulation results;
- Estimation of confidence intervals for hydraulic and water quality model results.

This deliverable describes the treatment and modelling of parameter uncertainties based on the available information, the propagation of these uncertainties using the mathematical model of water distribution networks and the evaluation with respect to the confidence intervals. After that the effectiveness will be illustrated in a number of example applications.

The deliverable is structured as follows: First a review on uncertainty quantification applications to water distribution networks is given, then the modelling of parameter uncertainties will be described, followed by an overview of the different options for propagation. The application is demonstrated using two example networks and the results are evaluated. The deliverable closes with the conclusions that could be drawn and the next steps in the development.

2 LITERATURE

The process of modelling and calibration is an important task in managing the distribution of potable water in urban networks. In general these models still contain uncertain parameters that are not further treated in the classical approach. The objective of uncertainty analysis (UA) is to quantify the parametric uncertainties and to evaluate their influence on the quantities of interest (QoI). In mathematical models of physical systems, parameter uncertainties are usually divided into two groups: aleatory and epistemic uncertainties (Smith, 2014):

- Aleatory uncertainties: These are stochastic uncertainties, as they are inherent to the problem and cannot be reduced by further physical knowledge. Typically, they are unbiased and naturally defined in a probabilistic framework.
- Epistemic uncertainties: Also known as systematic uncertainties are due to incomplete knowledge or simplification of the physical process. Epistemic uncertainties are often biased and in general less naturally defined in a probabilistic framework.

Water distribution network models contain a number of sources for uncertainty. This includes epistemic errors like insufficient information on the network topology or valve states and aleatory uncertainty associated with consumer demands. The nodal demands in water distribution networks are inherently uncertain and have underlying variability on the scale of minutes, hours and days or even on monthly and annual timescales (Buchberger and Wells, 1996; van Zyl et al., 2008; Herrera et al., 2010). Due to a number of unknown parameters in predictive models and the lumping of withdrawals at the network nodes, epistemic uncertainty is added to the demand. Pipe diameter and roughness are dependent on corrosive processes and will change over time. In order to reduce the epistemic error introduced by unknown parameters and processes, calibration methods are applied by several researchers (Savic et al. 2009; Piller et al., 2010). However, due to the large number of parameters, most approaches rely on some sort of grouping algorithm (Bascia et al., 2003; Kumar et al., 2010). Pumps, valves and tanks are key components in managing the water distribution. With age, the performance of these components may change. For pumps, the pump curve relationship may change as a result of corrosive processes and cavitation. Valves are also subject to corrosion which may diminish the ability to fully close a valve. Even worse, since most valves are operated manually the true state may even be misrepresented.

From a modeling point of view, parameters may be divided into two groups. There are fast changing parameters like the demand, which may be different for every realization of a scenario, fix or slow changing parameters like pipe diameter and roughness and other parameters where the exact value is unknown like pipe length. Fast changing parameters need frequent estimates for the value to get realistic results. Depending on the application different methods may be considered to be advantageous. Data driven models are very popular in real-time scenarios where the model is used to create realistic boundary conditions. They use the available historic data in order to estimate the nodal demands. Artificial Neural Networks (ANN), Support Vector Regression and other machine learning algorithms have been applied by a number of authors (e.g., Herrera et al. 2010; Braun et al. 2014). Although these models perform reasonably well under the chosen performance measure most of these approaches depend heavily on consumer aggregation in district metered areas (DMAs). Other models like SIMDEUM (Blokker et al., 2010) and Opointe (Piller and Bremond, 2002) use stochastic consumers that are represented in probability distributions. Through the development of appropriate scaling laws,

Vertommen et al. (2015) derive an uncertainty measure for nodal demands that is based on the collected, aggregated data.

Slow changing network parameters like pipe roughness or diameters are in general estimated through model calibration. The majority of calibration algorithms are based on optimization and least-squares approaches (Piller et al., 2010). These algorithms are designed to find deterministic parameter values that give the best fit between measured and simulated data with respect to a cost function. These approaches may consider measurements and model errors by weighting the residuals in an attempt to give each data point its proper amount of influence on the estimation. It is also possible to limit the influence of outliers on the point-wise estimation by using more robust estimators as the least absolute deviation criterion or the Huber function (Piller et al., 2015). Also, with these approaches, it is still possible to estimate confidence intervals for the predicted values for uncertainty that is introduced through measurement and model errors (Do et al., 2017). Typically, first-order second moment (FOSM) methods are applied (Bush and Uber, 1998; Lansey et al., 2001). Also, it is possible to cluster the nodal demand by defining a membership function based on the SVD decomposition of the sensitivity matrix for the head with regard to the nodal demand (Sanz and Perez, 2015). An additional benefit is the use of sensitivity coefficients for sampling design for calibration goal (Kapelan et al., 2003; Do et al., 2016). The sensitivity coefficients can be evaluated very efficiently based on explicit local sensitivity formulations as given in (Piller et al., 2017) even for large systems (Deuerlein et al., 2017). Nevertheless, the FOSM method is limited in scope to interval estimation and normal pdf distribution. Although the FOSM can be evaluated very efficiently based on explicit local sensitivity formulations as given in (Piller et al., 2017), it results in symmetrical confidence intervals. The linearization further limits the accuracy in cases with high variance and nonlinearity. A more complex approach for the calibration is introduced in (Kapelan et al., 2007) using the shuffled complex evolution metropolis (SCEM-UA) algorithm in a Bayesian type algorithm. It results in an approximation for more general probability distribution of the parameters.

With an accurate description of the parameter uncertainties, it is possible to evaluate their influence on the QoIs. This is done by propagating them through the mathematical model. Classical approaches use once again the FOSM, but they suffer from the same limitations as for the parameter uncertainty quantification. One of the most popular approaches that have been applied in numerous studies is the Monte Carlo simulation (MCS). Kang and Lansey (2007) use it to validate the results for the FOSM. While MCS are producing more accurate results for non-linear systems, they are also computationally more demanding than the FOSM. This deliverable shows that with applications of the Polynomial Chaos Expansion (PCE) it is possible to obtain results for the uncertainty propagation with similar quality as the MCS, but for much lower computational cost.

The objective of our work on WP4.3, presented in this deliverable, is to apply the more efficient spectral Polynomial Chaos Expansion to water distribution network models. Similar applications in the literature generally deal with small network models. By introducing a real medium sized network for the analysis of parameter uncertainties for the first time the scalability of the algorithms is demonstrated.

3 PARAMETER UNCERTAINTIES

In order to handle the parameter uncertainties in mathematical models, they are expressed as random processes. This section gives a formal definition of random variables and stochastic processes followed by details on random number generation.

3.1 Random Variables

In measurement theory, a random variable is defined by the triple $(\Omega, \mathcal{F}, P)$ containing the sample space Ω, the σ-field or σ-algebra $\mathcal{F}$ and the probability measure P. The sample space Ω of an experiment is defined as the set of all possible outcomes $\Omega = \{\omega\}$. The σ-algebra $\mathcal{F}$ is a subset of the sample space that contains all relevant events. In this context an event may be defined as a set of outcomes, including the empty set $\emptyset$ and all combinations of other events in the σ-field. Probability is a concept to measure the likelihood of occurrence for a certain event $P: \mathcal{F} \rightarrow [0,1]$. It has to satisfy the definitions $P(\emptyset) = 0, P(\Omega) = 1$ and if $A_i \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$, $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.
On this basis a random variable $X = X(\omega)$ assigns a number to each outcome ω of a random experiment with a quantifiable probability. Based on the nature of the experiment the sample space may be defined by a discrete set of abstract outcomes like in a coin toss or as in the parameters of a water distribution network by a continuous range of values, which can be used directly as the random variable.

There are a number of different ways for characterizing random variables. A common tool is the distribution as a function of the random variable. Also known as the Cumulative Distribution Function (CDF), which is defined for $F_X \rightarrow [0,1]$ by

$$F_X(x) = P\{\omega \in \Omega | X(\omega) \leq x\}$$

and describes the probability that a realization of the random variable has a value lower than x. An illustrative derivation of the CDF is the Probability Density Function (PDF) $f_X(x)$ which describes directly the probability of a certain realization. The PDF and the CDF are linked by the integral

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

3.2 Sampling Strategies for Random Variables

In many stochastic applications it is important to generate values of a random variable. For the generation of a basic independent and identically distributed (iid) random variable there exist two general approaches. They are known as Pseudo Random Number Generators (PRNG) and Quasi Random Number Generators (QRNG).

Pseudo Random Number Generators use an algorithm that produces a sequence of random numbers that has similar characteristics as the random variables. It is called pseudo random since the sequence of numbers is determined by the initial value which is often called the seed. Although the sequences are not truly random they are used because of their speed in the generation of random numbers and the reproducibility of particular sequences.

Quasi Random Number Generators use low-discrepancy sequences to generate iid uniformly distributed random numbers. The low-discrepancy requirement of these sequences generates more evenly distributed values however the points are highly correlated. Common sequences are for example the *Sobol Sequence*, the *Halton Sequence* or the *Mannersley Set.*
These quasi random sequences for uniform iid random variables are also helpful in the generation of more advanced random numbers.

Inverse Transform Sampling (ITF) is a method for generating samples of a random variable with a given cumulative distribution function. For a realization u of the uniform random variable $U \sim \mathcal{U}(0,1)$ the realization of the a random variable x can be defined as

$$x = F_X^{-1}(u).$$

In cases where the CDF is unknown and cannot be easily derived from a known PDF, ITF can be used as the base random number generator for assessing the sampling distribution.

Rejection Sampling (RS). The main idea of RS is to use an easy-to-sample distribution that is scaled to completely envelope the desired PDF. For the creation of the random numbers the enveloping distribution is sampled and the value is accepted with a probability that is determined by the ratio between the enveloping and the desired distribution function. Obviously, in order for the algorithm to be efficient the enveloping distribution should be chosen to closely follow the desired distribution in order to reduce the ratio of rejected samples.

Polynomial Chaos Expansion (PCE) offers an attractive alternative to the more complicated and sometimes inefficient algorithms like Rejection Sampling. The PCE models a random variable by using a polynomial chaos expansion that is expressed in an easy to sample basic random variable Z using a set of joint orthogonal polynomials $\Psi_i(Z)$. The polynomials have to hold under the inner product that is defined by

$$\langle \Psi_i, \Psi_j \rangle = \mathbb{E}[\Psi_i \Psi_j] = \int \Psi_i(z)\, \Psi_j(z) dP_Z(z) = \int \Psi_i(z)\, \Psi_j(z) f_Z(z) dz = \delta_{ij} \|\Psi_i\|^2$$

with respect to the probability density function f_Z.
The approximation for a random variable X is expressed by the truncated series:

$$X_N(Z) = \sum_{k=0}^{N} x_k \Psi_k(Z)$$

with the coefficients x_k and the order N of the expansion. The coefficients are calculated by the projecting the random variable X on the basis polynomials:

$$x_k = \frac{\langle X, \Psi_k \rangle}{\|\Psi_k\|^2}.$$

The choice of the polynomial basis generally depends on the distribution of the random variable. Gaussian variables are efficiently approximated by the Hermite polynomials

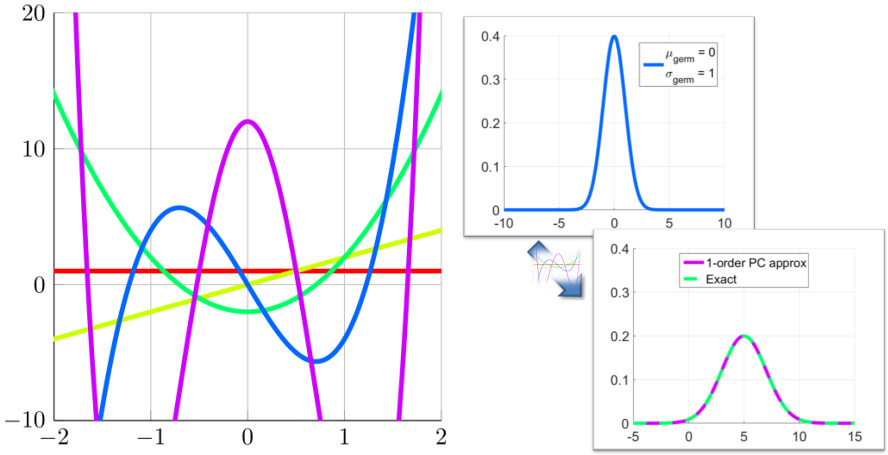


Figure 1: (a) Hermite polynomials order N=4. **(**b) First order PCE with Gaussian germ.

4 UNCERTAINTY PROPAGATION

4.1 Hydraulic Equations and Models

In hydraulic modelling, the simplified topological structure of a water distribution network is described by a directed graph corresponding to Figure 2. In this graph, links represent pipe sections, valves and pumps, and nodes the resource nodes, tanks, demand nodes and connections. The mathematical description of this graph is given by the node-link incidence matrix $\boldsymbol{A} \in \mathcal{M}^{\mathrm{nj}\times\mathrm{np}}(\mathbb{R})$, where nj is the number of junction nodes and np is the number of links. It is defined as $\mathbf{A} = \{A_{i,j}\}_{1\le i\le nj;1\le j\le np}$. The coefficients are defined as follows:

$$A_{i,j} = \begin{cases} -1 & \text{, if node } i \text{ is the end node of link } j \\ 0 & \text{, if link } j \text{ is not connected to node } i \\ +1 & \text{, if node } i \text{ is the start node of link } j. \end{cases}$$

Water distribution networks in general have a looped structure and the system state is described by the potentials at the nodes (heads) and the currents on the links (flow rates). The system equations are given by two sets of equations. First the mass balance at the junction nodes:

$$\mathbf{Aq} + \mathbf{c} = \mathbf{0} \tag{1}$$

where $\mathbf{A}$ is linked to the part of the network that only contains junctions with known demands, $\mathbf{q} \in \mathbb{R}^{np}$ is the vector containing the flow rates and $\boldsymbol{c} \in \mathbb{R}^{nj}$ is the vector of demands at junction nodes. Second the energy conservation equation:

$$\Delta\mathbf{h}(\mathbf{r},\mathbf{q}) - \mathbf{A}^{\mathrm{T}}\mathbf{h} - \mathbf{A}_f^T \mathbf{h}_f = \mathbf{0}$$

where $\mathbf{A}_f$ is the node-link incidence matrix reduced to nodes with known potential like reservoirs or tanks, and $\mathbf{h} \in \mathbb{R}^{nj}$ is the vector containing the piezometric heads at junction nodes. Parameters are given by the potential vector $\mathbf{h}_f \in \mathbb{R}^{nf}$ describing fixed heads at source nodes like reservoirs or tanks and the vector $\mathbf{r} \in \mathbb{R}^{np}$ containing the resistance coefficients for each link. The function $\Delta\mathbf{h}(\mathbf{r},\mathbf{q})$ describes the loss in head along a pipe and is defined by:

$$\begin{aligned} \Delta\mathbf{h}\colon \mathbb{R}^{np} \times \mathbb{R}^{np} &\longrightarrow \mathbb{R}^{np} \\ (\mathbf{r},\mathbf{q}) &\longmapsto \Delta\mathbf{h}(\mathbf{r},\mathbf{q}). \end{aligned}$$

It is usually termed the head-loss function. For medium and large Reynolds numbers the head-loss in general is a non-linear function of flow $\mathbf{q}$ and is linear in $\mathbf{r}$, but $\mathbf{r}$ can depend of $\mathbf{q}$. In the following application, the state vector $\mathbf{x}$ consists of the unknown flow rates $\mathbf{q}$ and the head $\mathbf{h}$ and the system parameters are combined in the vector $\mathbf{y}$.

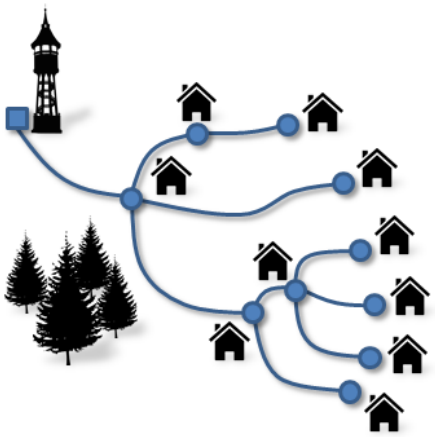


Figure 2: Tree graph network.

The demand boundary condition in the hydraulic model is set using one of two paradigms. In demand driven models (DDM), the consumption is defined as a fixed demand vector **d** that gives the actual consumption at each demand node in the network (in Eq. (1) **c** = **d**). In contrast, pressure driven models (PDM) define the nodal demand as a function of the available pressure $\mathbf{h} - \boldsymbol{e}$ (e is the elevation vector at junction nodes), a predefined service pressure head $\mathbf{h}_s$ and the minimum pressure head necessary for a discharge $\mathbf{h}_m$.This function is also referred to as the pressure outflow relation (POR). Literature gives several possible definitions for the POR, but one of the most common and realistic definitions and the one used here, was introduced by Wagner et al. (1988):

$$c(h_i) = \begin{cases} 0 & \text{, if } h_i \le e_i + h_{m,i} \\ \left(\dfrac{h_i - e_i - h_{m,i}}{h_{s,i} - e_i - h_{m,i}}\right)^{\frac{1}{2}} & \text{, if } h_{m,i} \le h_i - e_i \le h_{s,i} \\ d_i & \text{, if } h_{s,i} + e_i \le h_i. \end{cases}$$

For the PDM paradigm $\mathbf{c} = \mathbf{c}(\mathbf{h})$ in Eq. (1).

4.2 Propagation Algorithms

Central part of the Uncertainty Analysis is the propagation of errors and uncertainties by means of the mathematical model. To do so, a multitude of algorithms are available that have been tested and proven in numerous applications (Smith, 2014). It is possible to classify the majority of these methods in three groups. The perturbation or sensitivity methods, the sampling methods and the spectral methods.

- Perturbation Methods: These methods calculate the moments for the distribution of the quantity of interest directly from the system equations by means of a truncated Taylor expansion. Typically, the expansions employed are limited to first- or second-order expansions. This limits their accuracy for non-linear models with major variance in the parameters.
- Sampling Methods: With Monte Carlo Simulations as one of the most prominent representatives for this group, sampling methods are often applied for the propagation of uncertainties in non-linear models. Although, in general, implementation of the method is a straightforward task, its rate of convergence is asymptotically $1/\sqrt{M}$, where M is the number of simulations (Xiu, 2010; Fishman 2013).
- Stochastic Spectral Methods: The objective of spectral approaches like stochastic Galerkin and stochastic collocation methods is the calculation of a spectral representation of the random quantity of interest. Utilizing the smoothness requirement of the basis leads to an efficient convergence behaviour. The Polynomial Chaos Expansion is an efficient spectral method that has become popular in recent years (Xiu, 2010; Smith, 2014).

4.2.1 Perturbation Methods

Perturbation methods are especially efficient for the application in water distribution network modelling. The basic idea is to model the uncertain parameters symmetrical about their nominal value which in case of an uncertain demand would be $\bar{d}_i$ and $\bar{r}_i$ for an uncertain resistance. The perturbed parameters can then be represented by the generalized vector

$$\mathbf{Y} = \bar{\mathbf{y}} + \delta\mathbf{Y} = \left[\bar{d}_1 + \delta D_1, \dots, \bar{d}_{nd} + \delta D_{nd}, \bar{r}_1 + \delta R_1, \dots, \bar{r}_{np} + \delta R_{np}\right]^T.$$

Here $\delta\mathbf{Y}$ is the perturbation or uncertainty on the random vector $\mathbf{Y}$. Typically, $\overline{y_i}$ is taken as the expected value of the parameter and $\delta\mathbf{Y}_i$ as two standard deviations. To propagate the parameter perturbation the QoI vector $\mathbf{X}$ is developed in a first-order Taylor series expansion as a function of the perturbed parameter $\mathbf{Y}$.

$$\mathbf{X} = \mathbf{x}(\mathbf{Y}) = \mathbf{x}(\bar{\mathbf{y}} + \delta\mathbf{Y}) \approx \mathbf{x}(\bar{\mathbf{y}}) + \mathbf{S}\delta\mathbf{Y}$$

where the matrix $\mathbf{S}$ gives the sensitivity of the QoIs with respect to the parameters. The expected value of the QoI can be calculated as

$$\boldsymbol{\mu}_X = \mathbb{E}[\mathbf{X}] = \mathbb{E}[\mathbf{x}(\bar{\mathbf{y}})] + \mathbf{S}\mathbb{E}[\delta\mathbf{Y}] = \mathbf{x}(\bar{\mathbf{y}})$$

Using the same expansion the variance may be expressed as:

$$\boldsymbol{\Sigma}_X = \mathbb{E}\left[\left(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X}\right)\left(\mathbf{X} - \boldsymbol{\mu}_X\right)^T\right] = \mathbb{E}[\mathbf{S}\delta\mathbf{Y}\delta\mathbf{Y}^\mathbf{T}\mathbf{S}^T] = \mathbf{S}\boldsymbol{\Sigma}_\mathbf{Y}\mathbf{S}^T$$

where $\boldsymbol{\Sigma}_Y$ is the covariance matrix of the parameters. The sensitivities with respect to demand and roughness may be directly evaluated by:

$$\mathbf{S}_d = \begin{pmatrix} \frac{\partial \mathbf{q}}{\partial \mathbf{d}} \\ \frac{\partial \mathbf{h}}{\partial \mathbf{d}} \end{pmatrix} = \begin{pmatrix} -\mathbf{D}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1} \\ -(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1} \end{pmatrix}$$

$$\mathbf{S}_r = \begin{pmatrix} \frac{\partial \mathbf{q}}{\partial \mathbf{r}} \\ \frac{\partial \mathbf{h}}{\partial \mathbf{r}} \end{pmatrix} = \begin{pmatrix} \mathbf{D}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1}\mathbf{A}^T\mathbf{D}^{-1}\mathbf{B} - \mathbf{D}^{-1}\mathbf{B} \\ (\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1}\mathbf{A}^T\mathbf{D}^{-1}\mathbf{B} \end{pmatrix}$$

Here, **D** is the diagonal Jacobian matrix containing the derivatives of the head-loss with respect to the flow rate and **B** is the diagonal Jacobian matrix for the derivatives of the head-loss with respect to the resistance.

This formulation for the sensitivities can either be derived directly by deriving the system equations with respect to the demand and resistance or through the evaluation of the adjoint sensitivities. Although an adjusted formulation for the sensitivities exists for the pressure driven model (Piller et al., 2017) its usefulness for uncertainty quantification is limited to symmetrical confidence intervals or pdf as normal or uniform distribution.

4.2.2 Monte Carlo

Monte Carlo simulations (MCS) are stochastic collocation algorithms that rely on random sampling to obtain an approximate result for the problem at hand. In uncertainty quantification applications, it is used to repeatedly evaluate the deterministic system equations for a random sample of the uncertain parameter with the objective to obtain an approximate representation of the PDF of the QoIs.

Monte Carlo methods are very popular due to their straightforward implementation of the general procedure. In the first step iid random samples are generated from the parameter space $\mathbf{Y}^{(i)} = \left(d_1^{(i)}, \dots, d_k^{(i)}, r_1^{(i)}, \dots, r_l^{(i)}\right)$, $i = 1, \dots, M$ according to their respective distributions. This step makes heavy use of random number generation algorithms like the ones described in Section 3.2. In the second step the deterministic system is evaluated for each sample $i = 1, \dots, M$ from the parameter space **Y** to obtain the solution ensemble **X**. In the last step, the solution ensemble is used to evaluate the solution statistics defined in Section 3.1 where the mean is approximated by the sample average

$$\mathbb{E}[\mathbf{X}] \approx \widehat{\boldsymbol{\mu}}_X = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}^{(i)}$$

and the sample variance for each component as

$$\mathbb{E}\left[\left(X_k - \mu_{X_k}\right)^2\right] \approx \hat{\sigma}_{X_k}^2 = \frac{1}{M-1}\sum_{i=1}^{M}\left(x_k{}^{(i)} - \hat{\mu}_{X_k}\right)^2$$

The Monte Carlo method is driven by two basic statistical principles: The Law of Large Numbers (LLN) and the Central Limit Theorem (CLT) (Billing et al., 1979). The Law of Large Numbers states that if the samples are iid the sample average $\hat{\mu}_{X_k}$ will converge to the true mean in the limit of $M \to \infty$. This also holds for the sample variance $\hat{\sigma}^2_{X_k}$ and higher moments. Although the LLN guarantees the convergence of the MCS it does not evaluate the accuracy of the approximation. To do so the CLT has to be applied. Under the condition that the sample size justifies the LLN and the solution ensemble is iid, the Central Limit Theorem states that the sample distribution of the sample average converges to a Gaussian distribution $\mathcal{N}\left(\mathbb{E}[X_k], \sigma^2_{X_k}/M\right)$, with a standard deviation of $\sigma_{X_k}/\sqrt{M}$ and σ_{X_k} as the standard deviation of the true solution. This relation justifies the concept that the MCS converges proportional to the inverse of the square root of the sample size. It is obvious that the MCS can be easily generalized to more complex and even high dimensional applications, but due to its slow convergence with $1/\sqrt{M}$ it is prone to suffer from the curse of dimensionality.

4.2.3 Spectral Stochastic System

In the application of the Polynomial Chaos Expansion random variables are substituted by their spectral series expansion. This includes parameters for which the coefficients are known and the QoI for which the coefficients still have to be determined. This substitution is formulated in the stochastic system equations.

$$\mathbf{A}\mathbf{q}_N + \mathbf{d}_N = \mathbf{0}$$
$$\Delta\mathbf{h}(\mathbf{r}_N, \mathbf{q}_N) - \mathbf{A}^T\mathbf{h}_\mathbf{N} - \mathbf{A}_f^T\mathbf{h}_f = \mathbf{0}$$

In this system of equations, the random variables have been replaced by their PCE of the order N, following the description in section 3.2. There exist two general approaches to evaluate the coefficients of the QoI with their respective benefits and drawbacks. The first option is an intrusive approach that requires a reformulation of the system equations to allow for the direct solution. Although computationally very efficient it may pose certain complications since it is not possible to use existing implementations for the solution of the problem. Further, the handling of nonlinearities may require additional steps that introduce additional errors. The second option is given by non-intrusive algorithms which use a number of samples similar to the Monte Carlo simulation to evaluate the coefficients. This means that in regards to the computation current software and models may be used; on the other hand, though more efficient than the MCS, at some point the curse of dimensionality will make the calculations infeasible.

For the *Intrusive Methods*, the most prominent approach is the Galerkin Projection. It projects the stochastic system equations on the polynomial basis functions $\mathbf{\Psi}_k$ to create a new augmented system of equations that has $N + 1$ times the number of equations to determine the expansion coefficients of the QoI.

$$\langle \mathbf{A}\mathbf{q}_N + \mathbf{d}_N, \boldsymbol{\Psi}_k \rangle = \mathbf{0}$$
$$\langle \Delta\mathbf{h}(\mathbf{r}_N, \mathbf{q}_N) - \mathbf{A}^T\mathbf{h}_N - \mathbf{A}_f^T\mathbf{h}_f, \boldsymbol{\Psi}_k \rangle = \mathbf{0}$$

Non-intrusive methods follow more the spirit of stochastic collocations like Monte Carlo simulation does in the sense that it uses collocation points for which the deterministic system equations are evaluated. In contrast to the MCS, the direct evaluations in Intrusive Methods are used to fit the coefficients of the Polynomial Chaos Expansion to the data. For a QoI $\mathbf{X}$ vector:

$$\begin{bmatrix} \Psi_0\left(\mathbf{z}^{(1)}\right) & \dots & \Psi_N\left(\mathbf{z}^{(1)}\right) \\ \vdots & \ddots & \vdots \\ \Psi_0\left(\mathbf{z}^{(M)}\right) & \dots & \Psi_N\left(\mathbf{z}^{(M)}\right) \end{bmatrix} \begin{bmatrix} \mathbf{x}_1{}^T \\ \vdots \\ \mathbf{x}_N{}^T \end{bmatrix} = \begin{bmatrix} \mathbf{x}\left(\mathbf{y}\left(\mathbf{z}^{(1)}\right)\right)^T \\ \vdots \\ \mathbf{x}\left(\mathbf{y}\left(\mathbf{z}^{(M)}\right)\right)^T \end{bmatrix}$$

Where the $\mathbf{z}^{(i)}$ are realizations of the $\mathbf{Z}$ random vector and $\Psi_k\left(\mathbf{z}^{(i)}\right)$ is the value of the k-th polynomial basis function at the realization $\mathbf{z}^{(i)}$. The right-hand side is determined by the simulation result $\mathbf{x}$ of the deterministic system at the parameter vector $\mathbf{y}$ that is associated with the realization $\mathbf{z}^{(i)}$ of the germ distribution. $\mathbf{x}_k$ are the coefficients for the PCE of the state vector, which allows to evaluate the approximation from section 3.2.

5 EVALUATION OF UNCERTAINTIES

Result of the polynomial chaos expansion is a vector containing the coefficients for the spectral expansion of all Quantities of Interest in the network. This is by far more powerful than a simple estimation of the confidence intervals used in frequentist hypotheses testing. The following sections will show how the spectral expansion can be used to construct an estimate of the probability density function which gives the full information on the probability of every value for a QoI. Further, it will be shown how the stochastic moments like mean and covariance can be calculated directly.

5.1 Stochastic Moments

In many applications the probability distribution of random variable is characterized by a number of derived parameters called stochastic moments. The k-th moment of a distribution is defined by

$$\mu_k = \mathbb{E}[|X|^k] = \int_{-\infty}^{\infty} t^k f_X(t)\, dt.$$

The first moment is also known as the mean and gives the balance point of the distribution. With the use of the first moment it is possible to define the k-th central moment as:

$$\sigma_k = \mathbb{E}[|X - \mu_1|^k] = \int_{-\infty}^{\infty} (t - \mu_1)^k f_X(t)\, dt.$$

The central moments give a characterization for the shape of a distribution. For simple distributions a good characterization may be given by the mean and the second to fourth central moments also known as the variance, kurtosis and skewness.

5.2 Marginalization

One is often interested in the marginal density distribution of a QoI. The marginal distribution can be interpreted as a projection of the multivariate distribution on one of the output variables. This allows for a more comprehensible evaluation, however additional information like the covariance is lost in this representation. The marginal density is defined as

$$f_{X_j}(x_j) = \int_{\mathcal{D}_{\mathbf{x}_{\sim j}}} f_X(\mathbf{y}_{\sim j})\, d\mathbf{y}_{\sim j}$$

with the simplified notation $\mathbf{y}_{\sim j} = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_M)^T$. Their spectral representation of the integral may be formulated using the PCE and the marginal of a QoI in a M dimensional parameter space is given by

$$X_N(Z_j) = \int_{\mathcal{D}_{\mathbf{Z}_{\sim j}}} \sum_{k=0}^{N} x_k \Psi_k(\mathbf{Z})\, d\mathbf{Z}_{\sim j}.$$

In general, this integral is evaluated using the Monte Carlo algorithm by sampling the multivariate basic random variable $\mathbf{Z}_{\sim j}$.

5.3 Estimation of the Probability Density Function

The uncertainty propagation gives a characterization for the result random variables that allows for further evaluation and the estimation of the confidence intervals. One of the most common ways to visualize sampling data, which is generated by Monte Carlo type algorithms or from the Polynomial Chaos Expansion, is a histogram. In the histogram the parameter domain x is divided into N equidistant sections and the density for each section is approximated by

$$\tilde{f}(x) = \frac{1}{N} \frac{\text{Number of } x_i \text{ in same section as } x}{\text{Width of section}}$$

Kernel Density Estimation: A more general approach is the Kernel Density Estimation (KDE). It achieves a smooth and continuous approximation for the probability density function based on a chosen kernel function

$$\tilde{f}(x) = \frac{1}{Mh} \sum_{i=1}^{M} K\left(\frac{x - x_i}{h}\right)$$

Here, M is the number of samples, K is the chosen kernel function and h is a smoothing factor. The choice of the kernel function greatly influences the final result. One of the most common examples is the KDE with a Gaussian kernel function. In this case the only parameter is the standard deviation which is chosen based on the number of samples.
Pearson distributions: A theoretic way for the reconstruction of a probability density function that does not depend on sampling is given by the Pearson distributions. The Pearson distributions are a set of five functions that, based on the first four moments of a random variable, give a direct expression for the probability density function. Although this approach works quite well for random variables that have been modelled on a one dimensional parameter space, the application to more complex distribution of a random variable modelled on a two dimensional parameter space fails to give an accurate description of the real probability distribution.

5.4 Confidence Intervals

The objective of an interval estimate is to determine the values x_L and x_R that bound the location of the true value $x_L \leq x \leq x_R$. The estimate is based on a set of realizations $\{x_1, \dots, x_M\}$ of the random variable and the interval $[x_L, x_R]$ is called an interval estimator. A confidence interval is the combination of an interval estimator and a confidence coefficient. The confidence coefficient can be interpreted as the probability that the interval estimator contains the true value x. The $(1 - \alpha) \times 100\%$ symmetrical confidence interval for $[x_L, x_R]$ is defined such that for all $x \in \mathbb{F}$,

$$P[x_L \leq x \leq x_R] = 1 - \alpha, with\ P[x_L \geq x] = \frac{\alpha}{2}.$$

6 APPLICATION

This section introduces the example networks and presents the results for the application of the uncertainty propagation methods to different scenarios.

6.1 Network Models

All scenarios will be applied to one of the following networks. The first network is used for the simple, linear structure. The second one introduces the looped topology and the third network gives a realistic graph from the Veolia network.

Tree Network

The first network shown in Figure 2 is defined by tree shaped graph with a reservoir at the root node. Tree networks are special since the continuity equations are a determined system of linear equations (same number of unknowns and number of equations), which means that the flows can be calculated using the continuity equations and the demand vector **d** (for the DDM case). In contrast, the head is calculated by the energy equation, which is a nonlinear system of equations. The state vector $\mathbf{x} \in \mathbb{R}^{20}$ contains the flow rate of 10 pipes and the head at 10 junction nodes. The network is supplied by one tank at the tree root.

Small Looped Network

For the application of the uncertainty propagation algorithms in a fully nonlinear model, the small exemplary network illustrated in Figure 3 is used. The network contains three reservoirs and one loop. With the state vector $\mathbf{x} \in \mathbb{R}^{21}$ containing 12 flow rate values and 9 head values, the mass conservation equation is underdetermined and may no longer be solved directly. As for the tree network, uncertainties are directly imposed on nodal demands or the resistance of a specific pipe.

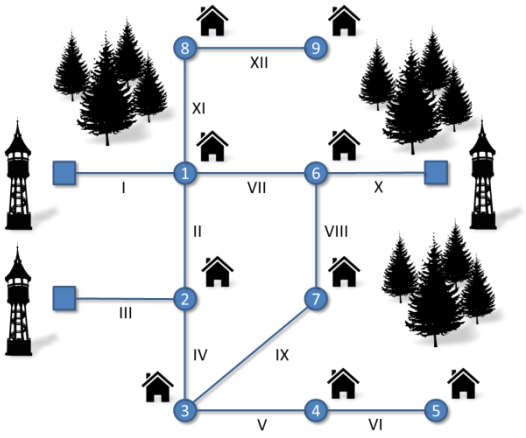


Figure 3: Small looped network.

Realistic Medium Sized Network

For testing the scalability of the propagation algorithms for bigger more complex networks, the second model illustrated in Figure 4 is introduced. It is based on a part of a real network that is managed by Veolia and contains a total of 2,175 pipes, 1,822 junction nodes and one reservoir node. In contrast to the small looped network, the uncertainty is imposed on the multiplier of the demand pattern so the effect is introduced on the whole network. The model currently contains two demand patterns.

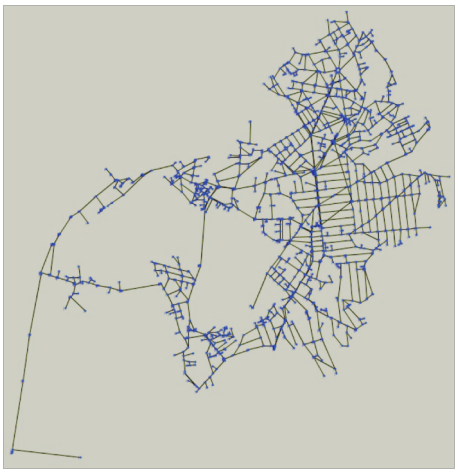


Figure 4: Medium sized network model**.**

6.2 Uncertainty Quantification Scenarios

In this section, the test scenarios are described for each network followed by a short discussion of the results.

6.2.1 Linear Tree Network

In this simple scenario, one of the demands at node 9 is uncertain and is modelled by a Gaussian distribution. For the propagation, three approaches are chosen for comparison. The FOSM, MCS and PCE are applied to compare their performance. Applying the FOSM the sensitivities have to be calculated. For the Monte Carlo simulation the complete system was evaluated a total number of $M = 1e4$ times. Due to the Gaussian form chosen for the uncertain parameters, the PCE uses the orthogonal Hermite polynomials as basis functions and in order to model the non-linear effects for the head, the expansion order has been chosen to be $N = 2$. The results are shown for the flow rate through pipe 9 in Figure 5 (a) and for the head at node 7 in Figure 5 (b).

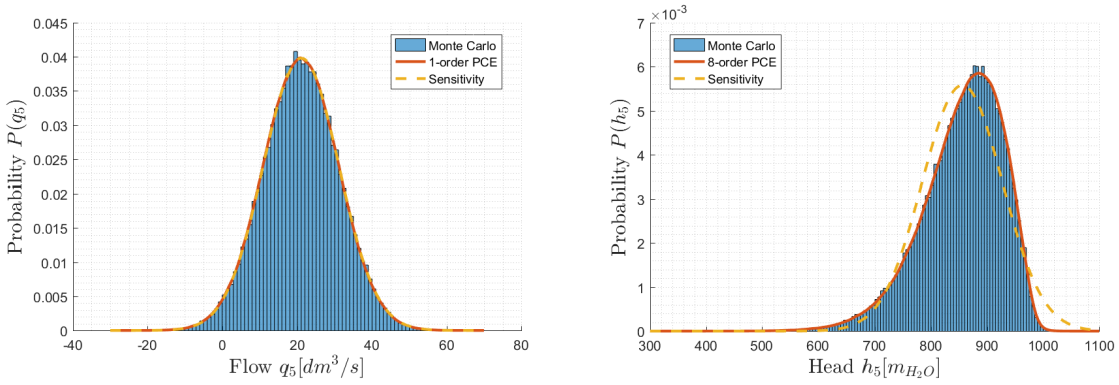


Figure 5: Flow rate and head PDFs for the tree network in 1D demand uncertainty.

For the flow, it is obvious that the three methods are in good agreement. This is due to the determined linear system of equations for the mass conservation. That means that the FOSM is able to calculate the exact Gaussian solution using the sensitivities. The MCS is able to reproduce a fairly accurate representation of the PDF. Actually, the PCE would be able to produce the same result using a first order expansion as this would also be a perfect Gaussian distribution. For the head, the result of the FOSM is different from the MCS and the PCE, since FOSM is not able to capture the non-linear effects of the energy equation. However, depending on the variance in the uncertain parameter the FOSM solution comes relatively close to the more complex methods in respect to the mean and the variance.
This scenario has also been tested with an application of the intrusive Galerkin projection. In general, the results were comparable, but the intrusive method adds further error due to the non-polynomial head-loss function which here was approximated by a Taylor series expansion.

6.2.2 Small Looped Network

For the small looped network, four scenarios have been applied. First, a one-dimensional parameter space is chosen for one of the demand nodes. In order to test the multidimensional polynomial expansion the second scenario introduces a two dimensional parameter space with two uncertain demands. In the third scenario, a comparison is made between the demand-driven and the pressure-driven model. The fourth scenario deals with the influence of demand uncertainties on water age.

1D Parameter Space DDM

The application of the one-dimensional parameter space to the small looped network is in fact very similar to the one chosen for the tree network. As uncertain parameter, the demand at node 5 is chosen. Since the branched network has shown that the FOSM is of limited use in non-linear applications, the evaluation is performed using only the PCE and the MCS. The result of the PCE is evaluated using two different methods. The first one samples the polynomial series expansion of the QoIs and approximates the PDFs using the kernel density estimation. To do so the PCE has been sampled a total number of $M = 1e4$ times. This result is represented by the solid red line in **Error! Reference source not**

found. Figures 6a and 6b. In the second approach, the moments of the PDFs are evaluated directly from the polynomial series through numerical integration. With the estimates for the first four moments, it is possible to draw the PDFs using the Pearson distributions. This result is given by the dashed yellow line. The result of the MCS is given by the blue histogram. As before, the histogram is based on $M = 1e5$ total evaluations of the hydraulic equations.

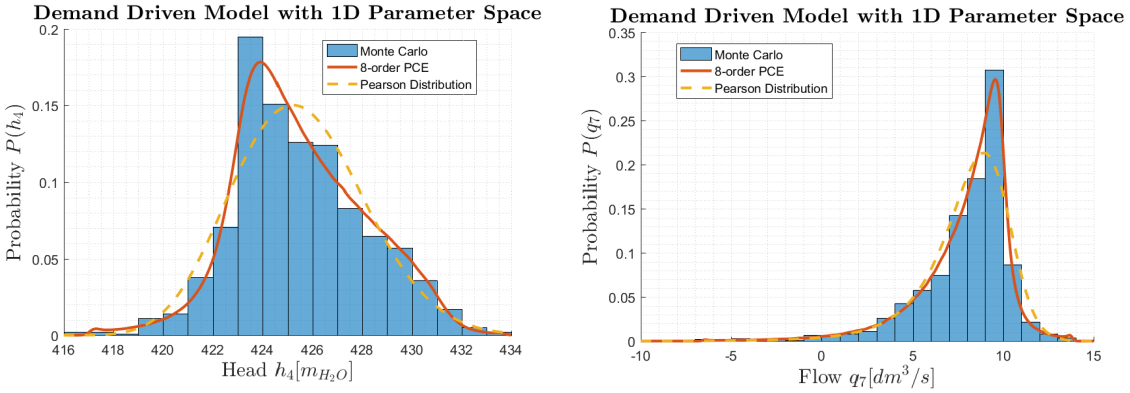


Figure 6: Flow rate and head PDFs based on 1D uncertain demand for the small loop network.

Through the introduction of the loop in the network graph, the mass conservation equations are underdetermined and need the solution in combination with the energy conservation equations to find a unique solution. This means that the flow rates are no longer determined by a linear system of equations and explains the clearly non-linear behaviour of the flow in **Error! Reference source not found.**Figure 6a. Apart from this, all three methods give very similar results and the conclusions from the linear 1D case still hold.

2D Parameter Space DDM

Expanding the previous example, a second uncertain demand is added to the parameter space. Now the demands at nodes 5 and 6 are represented by Gaussian random variables. The evaluation methods are unchanged using the PCE with the kernel density estimation and the Pearson distribution using the first four moments. As before MCS is used for the validation of the results.

From a numerical point of view the MCS shown in Figure 7 uses a total number of $M = 1e5$ simulations in order to converge while the PCE still uses $M = 1e4$ evaluations. However, this time with respect to the multidimensional parameter space the PCE is chosen to be 8^{th} order expansion.

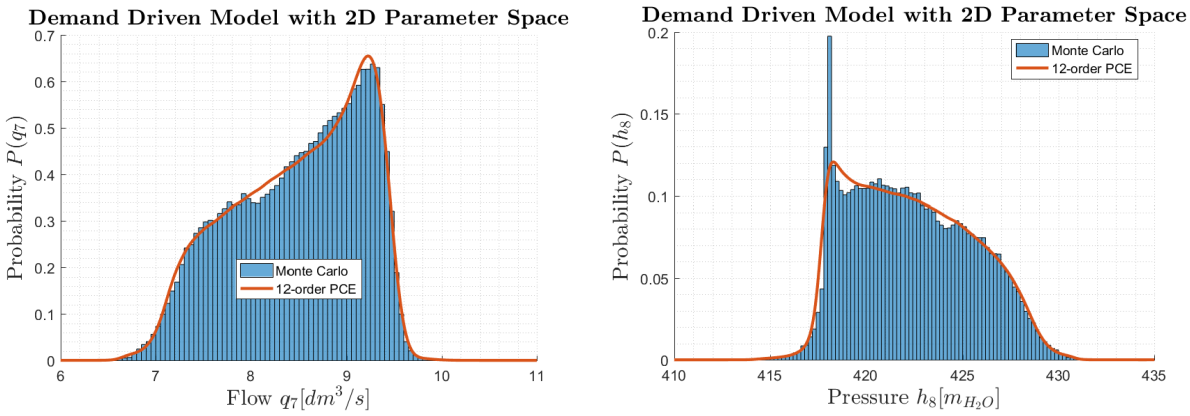


Figure 7: Flow rate and head PDFs based on 2D uncertain demands for the small loop network.

In Figure 7, once again, the MCS is given by the histogram and the kernel density estimation of the PCE in red. Since the Pearson distribution only uses the first four stochastic moments it is not able to capture the accurate shape of this distribution.

1D Parameter Space PDM

The third scenario for the looped network investigates the influence of PDM and compares the results to the classical DDM approach. In order to compare both modeling paradigms the same boundary conditions are applied in both cases and have been chosen to resemble those of the 1D DDM scenario. The only difference is the introduction of the POR for modeling the pressure dependence for demand.

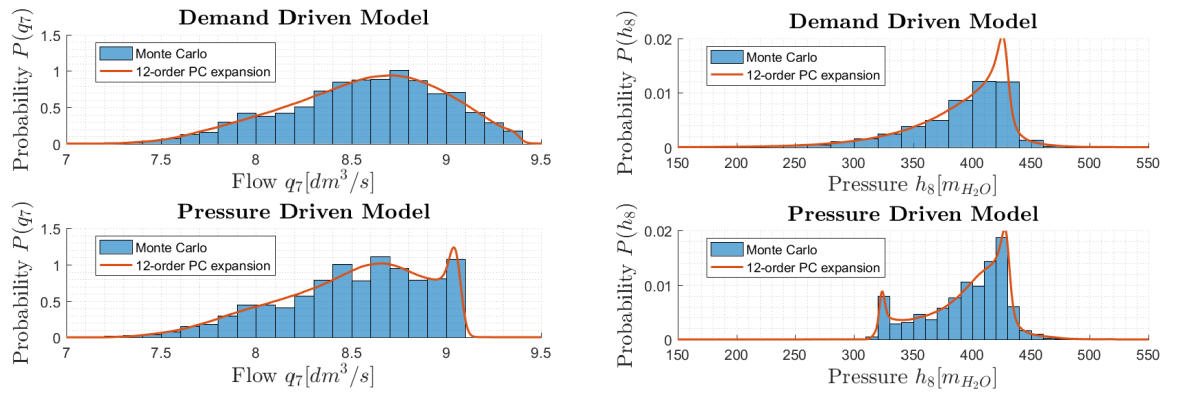


Figure 8: Comparison of flow rate and head DDM and PDM PDFs based on 1D uncertain demand for the small loop network.

The difference between the DDM and the PDM becomes apparent in Figure 8. Instead of a distribution with one peak as for the one-dimensional parameter space in the DDM, the PDM introduces a second

peak for elements where there is insufficient pressure to satisfy the requested demand. This pressure boundary condition also helps to explain the phenomenon. The flow displayed in Figure 8 (a) has a similar PDF for lower flows. If the flow rate reaches a certain value it is impossible for the PDM to exceed the flow based on the available pressure. Similar, Figure 8 (b) shows that the flow reduction restricts the pressure to a minimum value below which it cannot fall. The second peak in both figures may be explained by shifting the insufficient cases into the viable pressure region.

1D Parameter Space Water Quality

This scenario is designed to evaluate the influence of demand uncertainties on water quality, which in this case is represented by the water age (or water residence time). The uncertainty is given by a normal distribution ($\mu = 1, \sigma = 0.3$) on the demand multiplier of the domestic demand pattern. The results are obtained with the transport module of the software Porteau (from Irstea Partner). Porteau calculates a flow-weighted average age together with the min and max ages, which gives more insight for the water quality assessment than the average value alone.

The propagation is done using an extended period simulation (EPS) over the course of one week to get rid of the initial conditions. Figure 9 shows the mean water age together with the 95% confidence intervals as a function of the simulated time. It can be seen that after roughly one day the result stabilizes and is not dependent of the initial conditions. In Figure 10, the estimate PDF is visualized for the same nodes after a simulation period of 24 hours. The PCE and MCS are in good agreement for both distributions, but while the MCS uses a total number of $M = 1e5$ evaluations the PCE uses $M = 1e3$ evaluations for an expansion of the order $N = 8$.

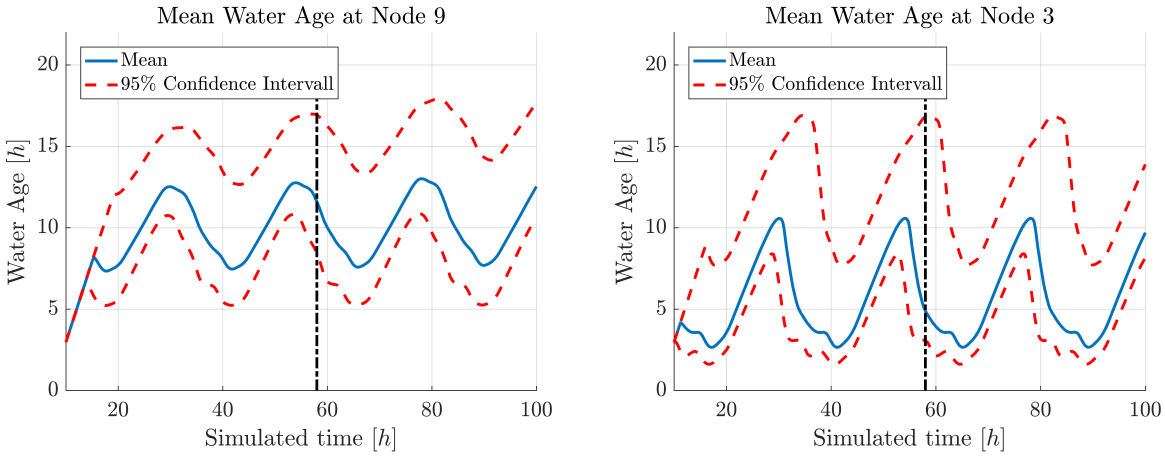


Figure 9: Min, average and max water ages predictions for two nodes in the small looped network.

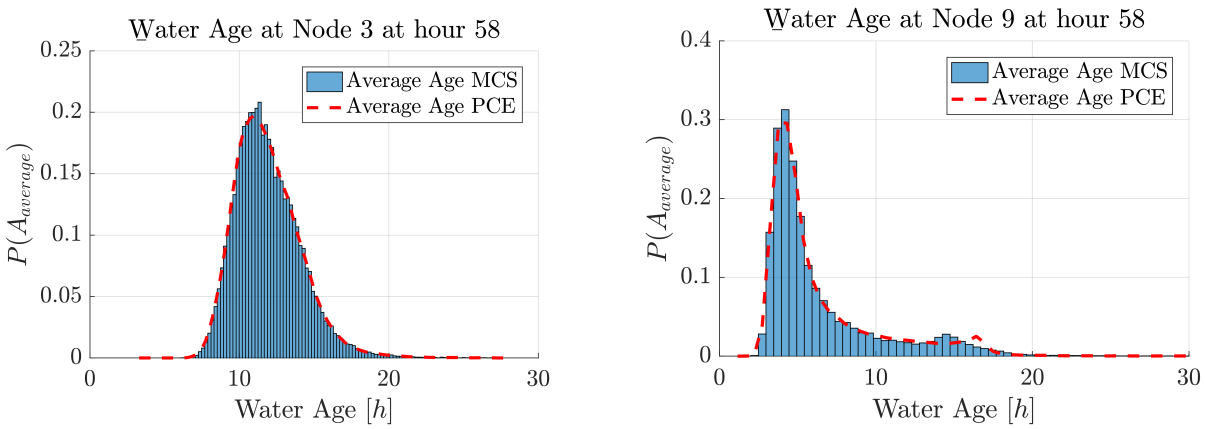


Figure 10: Water age distributions for small network: (a) Meshed part of the system (b) at Branch in the network.

Medium Sized Network

1D Parameter Space DDM

In the medium sized network the demand on each node is defined by the product of a base demand and a demand multiplier. For the application of the uncertainty propagation the demand multiplier is chosen as the uncertain parameter. Since this multiplier is applied to all demand nodes in the network, the uncertainty influences the complete network and not as before single elements. In general a multidimensional extension of this approach is possible by adding additional demand multipliers that are applied to groups of demand nodes.

Figure 11 shows the result of a normally distributed demand multiplier on the flow in the looped part of the network. The histogram represents the MCS with a total number of $M = 1e5$ evaluations and is compared to a PCE of the order $N = 12$ based on a total number of $M = 1e3$ evaluations.

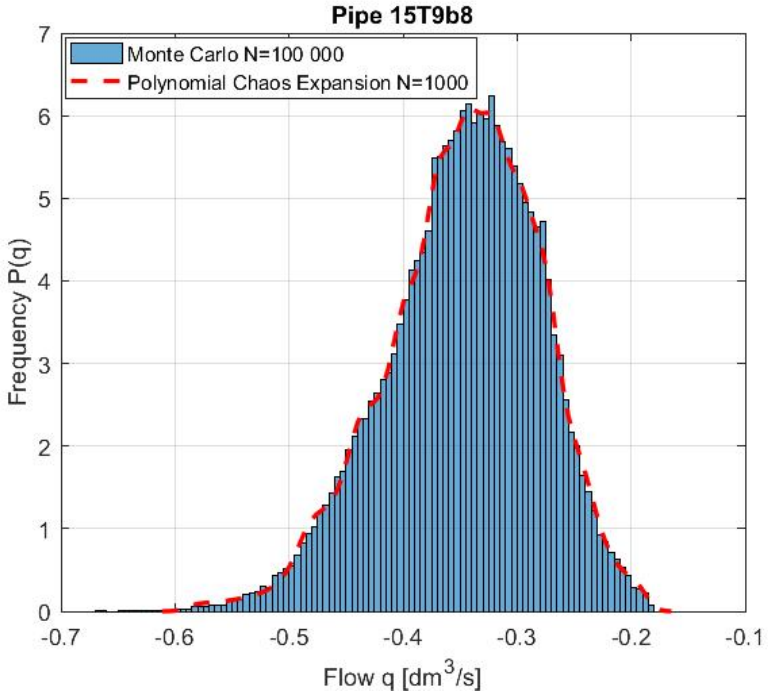


Figure 11: Flow rate PDF based on 1D uncertain demand for VEDIF medium-sized network.

Figure 10 shows the estimate for the resulting flow rate PDF for one of the pipes in the looped part of the network. It becomes clear that the distributed influence of the demand uncertainty has a strong nonlinear influence on the system. Further, it can be concluded that in this case, still the computationally much more efficient PCE gives comparable results to the MCS within a reasonable margin.

1D Parameter Space Water Quality

Similar to the small network the influence of uncertain demand is investigated for the medium size network. The uncertainty is again given by a normal distribution ($\mu = 1, \sigma = 0.3$) on the demand multiplier of the domestic demand pattern. In contrast to the small network where there were minimal differences in minimum, average and maximum water age Figure 12 shows all three at node 15Nf79 after the simulated time of 116 hours and 130 hours. These correspond to the maximum and minimum of the periodic age pattern. The PCE and MCS are again in good agreement with a total number of $M = 1e5$ MCS evaluations and a total of$M = 1e3$ evaluations for the PCE with an expansion of the order $N = 12$.

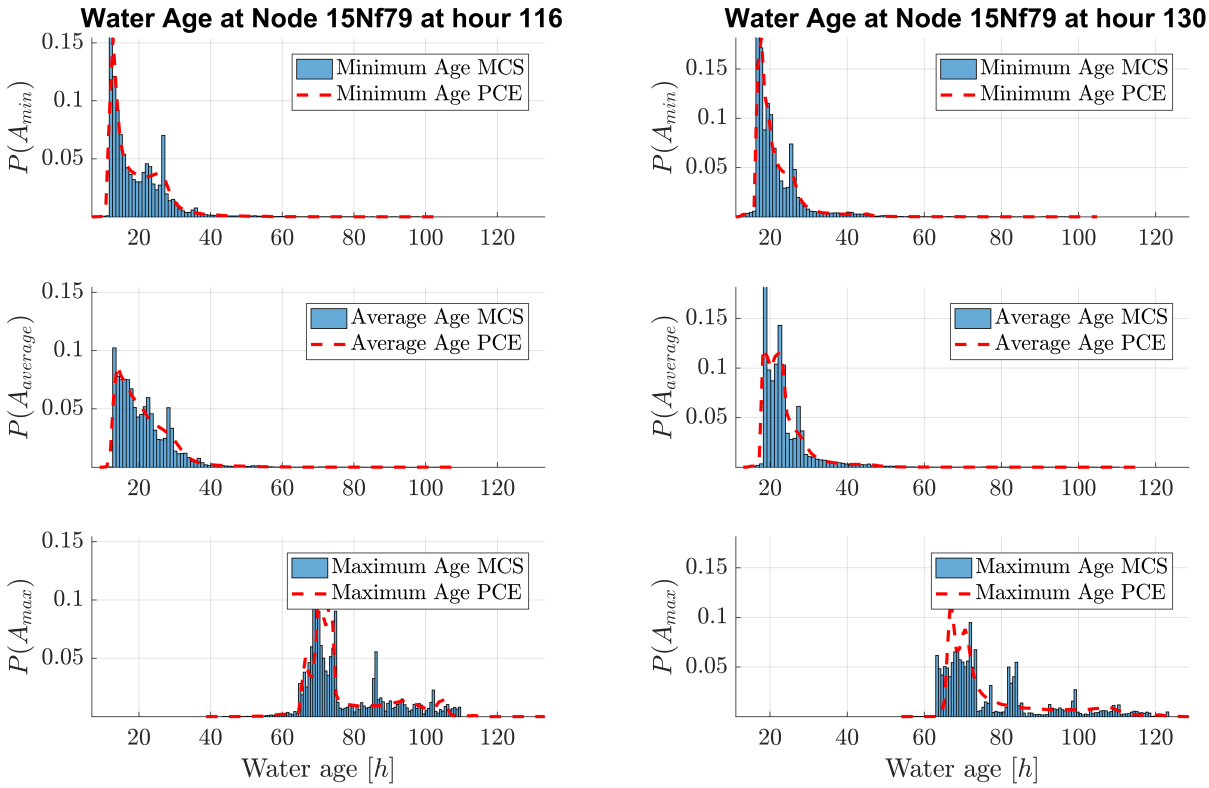


Figure 12: Water age distributions for the large Network: (a) Meshed part of the system after 116 hours simulation time; (b) Meshed part of the system after 130 hours simulation time.

7 LESSONS LEARNED

This section takes a closer look on technical questions that came up during the numerical experiments. First the necessary expansion order of the PCE is discussed. This is followed by a comparison of the efficiency between Monte Carlo methods and the PCE. The section closes with a comparison of the feasibility of intrusive and non-intrusive methods for the PCE. This section is based on the scenario in section 6.2.1, but the conclusions are representative for all applications.

7.1 Evaluating the Expansion Order

An important task for any application of an expansion approach is the evaluation of accuracy for the chosen development order. Since it is not possible to do so a priori this section shows the measures that have been taken based on the estimated coefficients. In a first iteration, the expansion order is chosen due to experience. Based on the evaluation it has to be adapted. The appropriate expansion order depends on factors like the non-linear properties of the modelled system and the desired accuracy for the application. Similar to other examples from polynomial approximation theory, it is assumed that the expansion converges to the true solution and that the theoretical infinite series may be represented by a truncated series of order N. From this it follows that coefficient values of higher order polynomials should be small and go to zero. Figure 13 (a) illustrates the convergence behaviour of the coefficients for a fourth order PC expansion. The coefficients are shown for the flow through pipe 5 and the pressure at node 8. As expected, their values decline rapidly and are close to zero for higher orders. Figure 13 (b) shows the probability density function for the pressure at node 5 up to a PC expansion of order four. While there are big differences between the first and the higher orders it is obvious that the distribution tails converge quickly for the orders three and higher.

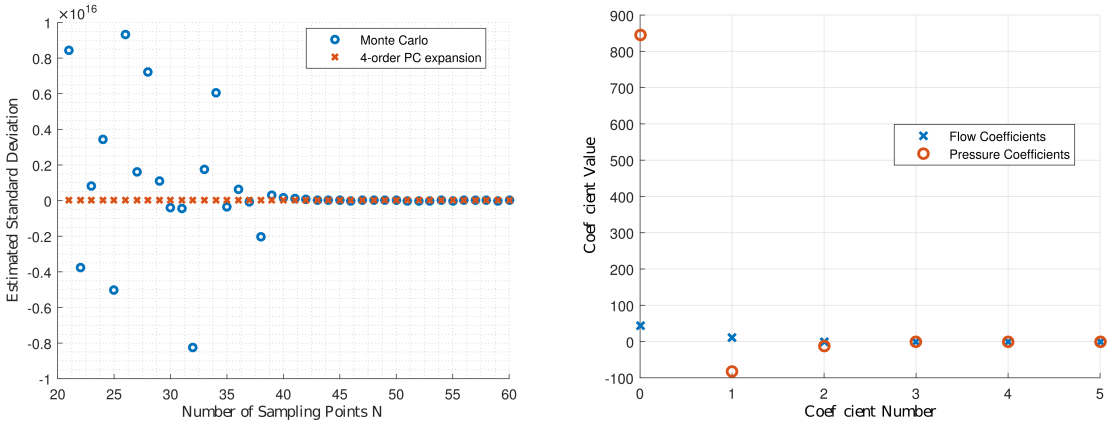


Figure 13: Probability density function for the pressure at node 7 based on the uncertain demand.

7.2 Monte Carlo versus Non-Intrusive Spectral Projection

Looking at the fact that both the Monte Carlo simulation and the non-intrusive spectral projection may be classified as sampling methods one may ask why the application of the Polynomial Chaos expansion is beneficial. From literature the answer to this question lays in the fact that PC methods use the smoothness of the orthogonal basis polynomials and in effect have superior convergence behaviour. To illustrate this, Figure 13 (a) shows the value for the standard deviation of the pressure at node 5 over the number of sampling points evaluated. For a low dimensional problem as the one discussed in this deliverable report a very small number of points is sufficient to get a good estimation of the PCE coefficients. For higher dimensional parameter spaces the number of sampling points will obviously increase.

7.3 Intrusive versus Non-Intrusive Methods

As introduced in Section 4.2.3 there exist two basic approaches to the calculation of the expansion coefficients with the intrusive and non-intrusive methods. As discussed, the intrusive Galerkin approach has to use a polynomial approximation for the non-linear head-loss function which introduces additional model errors into the system and for a low dimensional parameter space as in the presented scenarios the non-intrusive matrix inversion can reach very good results with as few as of 5 to 10 evaluations of the full system. Further, the application of the intrusive PCE is challenging, since the newly created set of equations changes for the addition of new, uncertain input parameters or with a change in the expansion order. This means it is not easily adaptable to new network models. The adaptation of non-intrusive methods to higher order expansions and a bigger parameter space on the other hand is relatively easy. This makes the approach more flexible. It follows that for reasonable sizes of the parameter space the non-intrusive approach gives an efficient and flexible framework, especially since methods like sparse grid collocation may reduce the amount of sampling points considerably.

8 CONCLUSION AND NEXT STEPS

The deliverable has introduced the theoretical framework for uncertainty quantification with the modelling of parameter uncertainties, the propagation of uncertainties and the evaluation for the QoIs. For the uncertainty propagation in addition to the two well-known approaches of the FOSM and MCS the new PCE approach has been introduced to water distribution network modelling.

In section 6, the performance of the different propagation methods has been compared in a number of different scenarios using network models of increasing complexity. It could be established that, as long as the uncertainty distribution has an inherent smoothness, the PCE is by far superior to the FOSM and even the MCS.

Further work will consider the use of reduced order models for a more efficient sampling in the parameter space. This will permit to deal with large-sized network as the Strasbourg network. In the cases presented in this deliverable, the parameter uncertainties have, in general, been assumed to follow a Gaussian distribution. However, in reality this is a strong assumption that is not always well supported by the data. Inverse modelling in the Bayesian sense gives a strong platform in order to infer the value of the parameters using measured data in the form of a probability distribution. The PCE can play a major role in making these computations more efficient. One shortcoming of the non-intrusive PCE is the repeated solution of the original network model for slightly changed boundary conditions. For high dimensional parameter spaces this may lead to a high computational load. In this respect, the introduction of reduced order models is beneficial to lower the burden and allow for a more efficient solution.

Among future applications that may be forecast is using uncertainty quantification as a guide for future data collection, calibration, state estimation, reliability studies incorporating probability assessment for the state to be in normal good condition and robust optimisation.

9 REFERENCES

Bascia, A., and Tucciarelli, T. (2003). "Simultaneous zonation and calibration of pipe network parameters." Journal of Hydraulic Engineering-ASCE,
129(5), 394-403.

Billingsley, P. (1979). "Probability and measure." John Wiley & Sons, 593pp.

Blokker, E. J. M., Vreeburg, J. H. G., and van Dijk, J. C. (2010). "Simulating Residential Water Demand with a Stochastic End-Use Model." Journal of Water Resources Planning and Management, 136(1), 19-26.

Braun, M., Bernard, T., Piller, O., and Sedehizade, F. (2014). "24-Hours Demand Forecasting Based on SARIMA and Support Vector Machines." Procedia Engineering, 89, 926-933.

Buchberger, S. G., and Wells, G. J. (1996). "Intensity, Duration, and Frequency of Residential Water Demands." Journal of Water Resources Planning and Management, 122(1), 11-19.

Bush, C. A., and Uber, J. G. (1998). "Sampling Design Methods For Water Distribution Model Calibration." Journal of Water Resources Planning and Management, 124(6), 334-344.

Deuerlein, J., Piller, O., Elhay, S., and Simpson, A. (2017). "Sensitivity Analysis of Topological Subgraphs within Water Distribution Systems." Procedia Engineering, 186, 252-260.

Efromovich, S. (2010). "Orthogonal series density estimation." Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 467-476.

Do, N. C., Simpson, A., Deuerlein, J., and Piller, O. (2016). "Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms." Journal of Water Resources Planning and Management, 142(11), 04016044, 04016013 pages.

Do, N. C., Simpson, A. R., Deuerlein, J. W., and Piller, O. (2017). "Particle Filter-Based Model for Online Estimation of Demand Multipliers in Water Distribution Systems under Uncertainty." Journal of Water Resources Planning and Management, 143(11), 04017065.

Elhay, S., Piller, O., Deuerlein, J., and Simpson, A. R. (2016). "A Robust, Rapidly Convergent Method That Solves the Water Distribution Equations For Pressure-Dependent Models." Journal of Water Resources Planning and Management, 142(2), 10 pages.

Fishman G. (2013). "Monte Carlo: concepts, algorithms, and applications.", Springer Science & Business Media, 4th printing 2013, 698 pages.

Herrera, M., Torgo, L., Izquierdo, J., and Pérez-García, R. (2010). "Predictive models for forecasting hourly urban water demand." Journal of Hydrology, 387(1), 141-150.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2003). "Multiobjective Sampling Design for Water Distribution Model Calibration." Journal of Water Resources Planning and Management, 129(6), 466-479.

Kapelan, Z. S., Savic, D. A., & Walters, G. A. (2007). "Calibration of water distribution hydraulic models using a Bayesian-type procedure." Journal of Hydraulic Engineering, 133(8), 927-936.

Kumar, S. M., Narasimhan, S., and Murty Bhallamudi, S. (2010). "Parameter Estimation in Water Distribution Networks." Water Resources Management, 24(6), 1251-1272.

Lansey, K. E., El-Shorbagy, W., Ahmed, I., Araujo, J., and Haan, C. T. (2001). "Calibration Assessment and Data Collection for Water Distribution Networks." Journal of Hydraulic Engineering, 127(4), 270-279.

Piller, O., and Brémond, B. "A Stochastic Model for Peak Period Analysis of Pipe Networks." Proc., ASCE Environmental & Water Resources Systems Analysis (EWRSA).

Piller, O., Gilbert, D., and Zyl, J. E. V. (2010). "Dual Calibration for Coupled Flow and Transport Models of Water Distribution Systems." Water Distribution Systems Analysis 2010, E. L. Kevin, Y. C. Christopher, O. Avi, and L. P. Ian, eds., ASCE, 722-731.

Piller, O., Braun, M., Montalvo, I., Deuerlein, J., and Bernard, T. (2015). "SMaRT-OnlineWDN D5.4: Development of a concept for online model calibration." Deliverable report for the ANR/BMBF research project SMaRT-Online^{WDN} on the CSOSG call, 34 pages.

Piller, O., Elhay, S., Deuerlein, J., and Simpson, A. R. (2017). "Local Sensitivity of Pressure-Driven Modeling and Demand-Driven Modeling Steady-State Solutions to Variations in Parameters." Journal of Water Resources Planning and Management, 143(2).

Sanz, G., and Pérez, R. (2015). "Sensitivity Analysis for Sampling Design and Demand Calibration in Water Distribution Networks Using the Singular Value Decomposition." Journal of Water Resources Planning and Management, 141(10), 04015020.

Savic, D. A., Kapelan, Z. S., and Jonkergouw, P. M. (2009). "Quo vadis water distribution model calibration?" Urban Water J., 6, 3.

Smith R. C. (2014). “Uncertainty quantification: theory, implementation, and applications”, book, CS12 SIAM series on Computational Science & Engineering, 382 pages.

Vertommen, I., Magini, R., and Cunha, M. d. C. (2015). "Scaling Water Consumption Statistics." Journal of Water Resources Planning and Management, 141(5), 04014072.

Wagner, J. M., Shamir, U., and Marks, D. H. (1988). "Water Distribution Reliability: Simulations Methods." Journal of Water Resources Planning and Management, 114(3), 276-294.

Xiu D. (2010). “Numerical methods for stochastic computations: a spectral method approach.,” Princeton University Press, 144 pp.

Xu, C., & Goulter, I. C., (1998). "Probabilistic model for water distribution reliability." Journal of Water Resources Planning and Management, 124(4), 218-228.