HAL
open science

A Spectral Approach to Uncertainty Quantification in Water Distribution Networks

Mathias Braun, Olivier Piller, Angelo Iollo, Iraj Mortazavi

To cite this version:

Mathias Braun, Olivier Piller, Angelo Iollo, Iraj Mortazavi. A Spectral Approach to Uncertainty Quantification in Water Distribution Networks. Journal of Water Resources Planning and Management, 2020, 146 (3), 04019080, 14 pp. 10.1061/(ASCE)WR.1943-5452.0001138 . hal-02511730

HAL Id: hal-02511730

https://hal.science/hal-02511730v1

Submitted on 19 Mar 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.

Author-produced version of the article published in J. Water Res. Plan. & Man., 2020, 146(3), 04019080.
The original publication is available at
https://ascelibrary.org/doi/abs/10.1061/(ASCE)WR.1943-5452.0001138

doi:10.1061/(ASCE)WR.1943-5452.0001138

A Spectral Approach to Uncertainty Quantification in Water Distribution Networks

Mathias Braun[1], Olivier Piller[2], Angelo Iollo[3], and Iraj Mortazavi[4]

[1]Research Scientist, INRAE, ETBX, F-33612 ,Cestas, France, & INRIA Bordeaux-Sud-Ouest, Team MEMPHIS, 200 Avenue de la Vieille Tour, 33405 Talence, France

[2]Senior Research Scientist, INRAE, ETBX, F-33612 ,Cestas, France, & Adjunct Senior Lecturer, School of Civil, Environmental, and Mining Eng., Univ. of Adelaide, Adelaide, SA 5005, Australia. Email: olivier.piller@irstea.fr

[3]Professor, Université Bordeaux, IMB, 351 Cours da la Libération, 33400 Talence, France, & Research director, INRIA Bordeaux - Sud-Ouest, Team MEMPHIS, 200 Avenue de la Vieille Tour, 33405 Talence, France

[4]Professor, CNAM, EA-7340-M2N-Modélisation Mathématique et Numérique, 2 rue Conté, 75003 Paris, France.

ABSTRACT

To date, the hydraulics of water distribution networks are calculated using deterministic models. Regarding the fact that many of the parameters in these models are not known exactly, it is important to evaluate the effects of their uncertainties on the results through Uncertainty Analysis. For the propagation of uncertain parameters, this article for the first time applies the Polynomial Chaos expansion to a hydraulic model and compares the results to classical approaches like the First Order Second Moment method and Monte Carlo simulations. Results presented in this article show that the accuracy of the Polynomial Chaos expansion is on the same level as the Monte Carlo simulation. Further, it is concluded that due to its computational efficiency the Polynomial Chaos Expansion is superior to the Monte Carlo simulation.

INTRODUCTION

High quality drinking water is one of our most valuable resources. Water distribution networks (WDN) make this resource available to the population in a convenient and efficient manner. Hydraulic modeling of this vital infrastructure is an important task for water utilities and has been used for a number of practical applications in the past. Some examples are: Network design Savic and Walters (1997); Ostfeld and Shamir (1996), modeling water leakage Van Zyl and Cassa (2013), optimal sensor placement Hart and Murray (2010); Ostfeld and Salomons (2004); Hagos et al. (2016) and water quality simulation Rossman et al. (1993). In all of these applications, parameters of the network model play an important role in determining accurate and reliable results. But, due to limited accessibility and sparse availability of sensors inside the network they cannot always be measured directly and some of the parameters may vary over time. This means that many parameters have to be considered as uncertain.

Figure 1 illustrates three of the most important sources of uncertainty in a water distribution network: The nodal demands, the roughness and diameter of the pipes. The nodal demand in water distribution networks is inherently uncertain and has an underlying aleatory variability on the scale of minutes, hours and days or even on monthly and annual time-scales Herrera et al. (2010); Buchberger and Wells (1996). Further, a number of modeling assumptions introduce epistemic errors to the demand. This includes the nodal agglomeration of demands that occur distributed along a pipe Walski et al. (2003) or due to simplification of the network graph Perelman et al. (2008) and unrealistic demands due to model deficiencies Braun et al. (2017). Pipe diameter and roughness are influenced by corrosive processes and will change over time Boulos et al. (2004). Several applications have shown that taking parameter uncertainties into account may have significant influence on the results. For example Pasha and Lansey (2010) investigate the effect of parameter uncertainties on water quality using the Monte Carlo method and Perelman et al. (2013) investigate the influence of uncertainties in the context of least-cost design for water distribution networks.

The interest of Uncertainty Quantification (UQ) is to evaluate and mitigate the influence of

parameter uncertainties. The first step in UQ is the estimation of parameter and their uncertainties. Parameter forecasting and estimation are active research topics for water distribution network modeling. For fast changing parameters like the nodal demand there exist models that try to predict the value for the next time step. This may be achieved either by data driven models Herrera et al. (2010); Braun et al. (2014) or physical models that simulate the stochastic nature of consumers Blokker and Van der Schee (2006). Other applications choose to solve the problem through calibration methods, which try to identify a set of parameters that give the best fit between measured and simulated data. The majority of calibration algorithms are based on optimization and least-squares approaches Savic et al. (2009); Piller et al. (2010). The most popular approach for deriving information on the probability distribution of the parameters uses the First Order Second Moment (FOSM) method, which is limited to linear systems or in the case of non-linear systems to small errors. An alternative approach that takes full account of the model non-linearities is presented by Kapelan et al. (2007). They use a shuffled complex evolution metropolis (SCEM-UA) algorithm in order to solve the inverse problem by the means of a Bayesian approach which results in a more realistic estimate for the probability distribution.

In the second step of UQ the parameter uncertainties are propagated. Popular approaches are perturbation methods like the FOSM Bush and Uber (1998); Lansey et al. (2001) which use the sensitivity of the network model Razavi and Gupta (2015); Lu and Vesselinov (2015) and Monte Carlo simulation Lansey (1997). Piller and Brémond (2002) investigate the influence of a stochastic consumer model that is expressed by a binomial distribution using both the FOSM method and Monte Carlo simulation. Similar to that Hwang et al. (2017) apply both methods to a number of different network models and evaluated with respect to the use in calibration, abnormality detection and system design.

The objective of this article is to evaluate current uncertainty quantification methods for steady state hydraulic models in order to get uncertainty measures for calculated flow rates and heads. This article explores an alternative to the FOSM and Monte Carlo simulations with the application of a spectral approach in the form of the Polynomial Chaos Expansion (PCE) for uncertainty

quantification in water distribution networks.

To evaluate the results two classical methods are chosen and compared to the spectral approach. One is the First Order Second Moment method as a representative for the perturbation methods. The second one is the Monte Carlo simulation for the sampling approaches.

The structure of this article is as follows. Section 2 introduces the basic concepts and terms from Uncertainty Analysis. This is followed by description of the hydraulic model and the application of the three propagation methods in Section 3. In Section 4 an illustrative example is presented using a tree shaped network graph. Finally, in Section 5, the method is applied and validated for a real medium size network with almost 4000 elements for a high dimensional parameter space.

UNCERTAINTY ANALYSIS

The objective of Uncertainty Analysis is to investigate and mitigate the effect of uncertainties on the Quantity of Interest (QoI). In this framework Uncertainty Quantification analyses and quantifies the sources for error linked to uncertainties. In general three classes of uncertainties can be identified in the numeric modelling and calculation process. They are given by Model Errors and Discrepancies, Input Uncertainties and Numerical Errors and Uncertainties.

- Model Errors and Discrepancies: Mathematical models are an approximation of the real systems. In the modeling process, errors are introduced due to a lack of knowledge or simplifications of actual structures. These discrepancies introduce bias and systematic uncertainties that are difficult to quantify using a probabilistic framework.
- Input Uncertainties: Inputs to a model can take the form of parameters or initial conditions. Typically they are estimated or calibrated but still uncertain.
- Numerical Errors and Uncertainties: These uncertainties are linked with the numerical evaluation of the system equations. Examples are for instance round-off, discretization or approximation errors.

In models that use real data further errors may be introduced through measurement uncertainties.

Random Parameters

In uncertainty analysis, parameters are modeled as random variables. Measurement theory defines a random variable 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 $\Omega = \{\omega\}$, where ω is one specific outcome. 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 and $A_j \in \mathcal{F}$ 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) is defined as $F_X \rightarrow [0, 1]$ by

$$F_X(x) = P\left\{\omega \in \Omega \mid X(\omega) \leq x\right\} \tag{1}$$

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) defined by $f_X(x)$ which describes directly the probability of a certain realization. The PDF and the CDF are linked by the integral

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt. \tag{2}$$

Uncertainty Propagation

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 (2013). 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 highly non-linear models Cacuci et al. (2005).
- 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, it is computationally expensive, as the variance of the mean estimate converges with a rate of $1/\sqrt{M}$, where M is the number of simulations Xiu (2010); Fishman (2013).
- Stochastic Spectral Methods: The objective of the polynomial chaos expansion is the calculation an approximation for random variables in a mathematical model. This truncated series uses orthogonal polynomials with increasing frequency as a basis. Utilizing the smoothness of these polynomials leads to an efficient convergence behaviour Xiu (2010); Smith (2013).

In reality, some methods are not strictly limited to one of these groups. For example, the Non-Intrusive Spectral Projection (NISP) is a spectral method that uses a sampling approach for the calculation of its coefficients Le Maître and Knio (2010).

Perturbation Methods: These 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 roughness. The perturbed parameters can then be represented by the generalized vector

$$\mathbf{P} = \bar{\mathbf{p}} + \delta\mathbf{P} = \left[\bar{d}_1 + \delta D_1, \ldots, \bar{d}_{nd} + \delta D_{nd}, \bar{r}_1 + \delta R_1, \ldots, \bar{r}_{np} + \delta R_{np}\right]^T. \tag{3}$$

Here, $\bar{\mathbf{p}}$ is taken as the expected value of the parameters and $\delta\mathbf{P}$ is a random perturbation or uncertainty on $\mathbf{P}$. To propagate the parameter perturbation the QoI G is developed in a Taylor expansion as a function of the perturbed parameters $\mathbf{P}$:

$$\mathbf{G} = \mathbf{g}(\mathbf{P}) = \mathbf{g}\left(\bar{\mathbf{p}} + \delta\mathbf{P}\right) \approx \mathbf{g}\left(\bar{\mathbf{p}}\right) + \mathbf{S}\delta\mathbf{P}. \tag{4}$$

where $\mathbf{S}$ is the sensitivity matrix evaluated at the mean value $\bar{\mathbf{p}}$ of the random variable. For this first order expansion the expected value of the QoI can be calculated by

$$\boldsymbol{\mu}_{\mathbf{G}} = \mathbb{E}\left[\mathbf{G}\right] = \mathbf{g}(\bar{\mathbf{p}}) \tag{5}$$

Using the Taylor expansion of the random parameter in equation (4) the variance can be determined as:

$$\Sigma_{\mathbf{G}} = \mathbb{E}\left[(\mathbf{G} - \boldsymbol{\mu}_{\mathbf{G}})(\mathbf{G} - \boldsymbol{\mu}_{\mathbf{G}})^T\right] = \mathbb{E}\left[\mathbf{S}\delta\mathbf{P}\delta\mathbf{P}^T\mathbf{S}^T\right] = \mathbf{S}\Sigma_{\mathbf{P}}\mathbf{S}^T. \tag{6}$$

where $\Sigma_{\mathbf{P}}$ the covariance matrix of the parameters Cacuci et al. (2005).

Monte Carlo: Stochastic collocation methods like Monte Carlo simulations (MCS) use random samples in order to obtain a stochastic approximation for the mean and variance of random variables. 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, random

samples are generated from the parameter space $\mathbf{P}^{(i)} = (d_1^{(i)}, \ldots, d_k^{(i)}, r_1^{(i)}, \ldots, r_l^{(i)})^T, i = 1, \ldots, M$ 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 $i = 1, \ldots, M$ from the parameter space $\mathbb{P}$ to obtain the solution ensemble $\mathbb{G} = \left\{g^{(1)} \ldots g^{(M)}\right\}$. 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

$$\mathbb{E}[G] \approx \hat{\mu}_G = \frac{1}{M} \sum_{i=1}^{M} g^{(i)} \tag{7}$$

and the sample variance as

$$\mathbb{E}[(G - \mu_G)^2] \approx \hat{\sigma}_G^2 = \frac{1}{M-1} \sum_{i=1}^{M} (g^{(i)} - \hat{\mu}_G)^2 \tag{8}$$

The Monte Carlo method is supported by two basic statistical principles: The Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). The Law of Large Numbers states that if the samples are iid the sample average $\hat{\mu}_G$ will converge to the true mean in the limit of $M \to \infty$. This also holds for the sample variance $\hat{\sigma}_G^2$ 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 $\mathcal{N}(\mathbb{E}[G], (\sigma_G^2)/M)$, with a standard deviation of $\sigma_G/\sqrt{M}$ and σ_G 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.

Polynomial Chaos Expansion: If $\boldsymbol{Z}$ is a vector of independent and identically distributed (i.i.d.) random variables on Ω, then a scalar random variable $X : \Omega \to \mathbb{R}$ can be represented by the infinite

expansion

$$x(\boldsymbol{Z}) = \sum_{k=0}^{\infty} x_k \Psi_k(\boldsymbol{Z}), \tag{9}$$

where Ψ_k are the orthogonal polynomials. In practical applications, a truncated expansion with a total number of $N+1$ elements is used to approximate random variables

$$x_N(\boldsymbol{Z}) = \sum_{k=0}^{N} x_k \Psi_k(\boldsymbol{Z}). \tag{10}$$

The number of terms in this finite polynomial chaos expansion is dependent on the dimension of the parameter space n and the order of the expansion l. It is calculated as:

$$N+1 = \frac{(n+l)!}{n!l!}. \tag{11}$$

Polynomial chaos expansion as introduced by Wiener (1938) uses orthogonal Hermite polynomials to model arbitrarily distributed random variables based on a Gaussian distribution. Xiu and Karniadakis (2002) generalized the method to the use of a wider variety of random processes by introducing broader classes of polynomials defined by the Askey scheme Xiu (2010). For each polynomial basis $\Psi(z)_k$ a weighted inner product $L^2(\Omega)$ is defined as

$$\langle \Psi_i, \Psi_j \rangle = \int_{\Omega} \Psi_i(z) \Psi_j(z) \, dF_Z(z) = \int_{\Omega} \Psi_i(z) \Psi_j(z) f_Z(z) \, dz = \delta_{ij} \langle \Psi_i^2 \rangle, \tag{12}$$

with respect to the probability density function $f_Z(z)$.

The numerical evaluation of such weighted inner products are usually done using quadrature rules that are based on polynomials interpolating functions. This quadrature rule transforms a general continuous integral as given in equation (12) into a discrete sum.

$$\int_{\Omega} \Psi_i(z) \Psi_j(z) f_Z(z) \, dz = \sum_{k=1}^{n+1} \Psi_i(z_k) \Psi_j(z_k) w_k \tag{13}$$

For the correct evaluation, the evaluation points on the abscissa z_k and the weight coefficients w_k

have to be determined. This can be either done using tables as the ones given in Zwillinger (2002) or using specific algorithms as presented in Press et al. (1988) . The evaluation of these sums is very efficient since for an integral over a polynomial with a maximum order of n the quadrature rule gives the exact result for a sum with $n + 1$ elements.

For the forward propagation of the parameter uncertainties, it is necessary to calculate their input parametrization in the orthogonal polynomials basis. Given a random input parameter P that is modeled as a function of the base random variable Z as:

$$p_N(Z) = \sum_{k=0}^{N} p_k \Psi_k(Z) \tag{14}$$

the $N + 1$ coefficients have to be determined. The basic framework for the calculation of the coefficients is given by the normalized projection:

$$p_k = \frac{\langle P, \Psi_k \rangle}{\langle \Psi_k^2 \rangle}. \tag{15}$$

In practice, there are two approaches. The strong approximation applies in cases where the parameter is explicitly known as a function P of the basis random variable Z. For that case, the evaluation of equation (15) can be done straightforward by the integral:

$$p_k = \frac{1}{\Psi_k^2} \int_\Omega P(z) \Psi_k(z) f_Z(z) \, dz \tag{16}$$

Such a case is for example given by a random variable $P \sim \mathcal{N}(\mu, \sigma)$, which is defined as $P(Z) = \mu + \sigma Z$.

Usually the direct definition of the random parameter is not known. Instead the parameter is characterized by its probability distribution function $F_P(p) = \mathbb{P}(P \leq p)$. For such cases, a weak approximation is performed. In contrast to the strong approximation, the convergence of the weak approximation is not guaranteed. Instead, it converges in probability. With $F_P : I_P \rightarrow [0, 1]$ and $F_Z : I_Z \rightarrow [0, 1]$ mapping the random variables P and Z to a uniform distribution in $[0, 1]$, equation

(16) may be rewritten. Assume that $U = F_P(P) = F_Z(Z)$, the random variables may be expressed as $P = F_P^{-1}(U)$ and $Z = F_Z^{-1}(U)$ respectively. This allows to give $P = F_P^{-1}(F_Z(Z))$ as a function of Z, which results in

$$p_k = \frac{1}{\Psi_k^2} \int_\Omega F_P^{-1}(F_Z(z))\, \Psi_k(z)\, f_Z(z)\, dz \tag{17}$$

For multivariate random variables the inversion of the probability density function is generally not possible. This means that multivariate input parameterizations usually use the tensorization of univariate random variables

$$p(\mathbf{Z}) = \prod_{i=1}^{n} p(Z_i)\,. \tag{18}$$

However, this way of modeling demands for the random parameters to be independent.

Currently a number of toolkits are under development that are either open source or free of use for academic users. UQLab is developed as a MatLab toolbox by the research group of Prof. Bruno Sudret, the chair of the department for Risk, Safety and Uncertainty Quantification at ETH Zurich Marelli and Sudret (2014). It is divided into the UQLabCore which is a proprietary content management system and the open source UQLab modules. OpenTURNS is an open source project for the development of a software platform dedicated to uncertainty quantification in simulation models. The development is lead by collaboration of EDF R&D, Airbus Group and Phimeca Engineering which was later joined by IMACS Baudin et al. (2017). It is licensed under the LGPL and is written as a C++ library and a Python Text User Interface (TUI), which works under Linux and Windows environments. Sandia National Laboratories developed the software toolkit Dakota for optimization and uncertainty quantification. It uses an advanced parametric analyses that enables, among other applications, the uncertainty quantification through computational models. The Dakota toolkit is also licensed under the GNU LGPL and written as a C++ application that provides a flexible, extensible interface to connect to external simulation codes Adams et al. (2009). Due to the implementation in MatLab, for this article the UQLab toolbox is chosen for the evaluation of the more complex use cases.

Evaluation of uncertainties

Result of the polynomial chaos expansion is a vector or matrix 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.

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

$$\mu_k = \mathbb{E}\left[|X|^k\right] = \int_{-\infty}^{\infty} (x)^k f_X(x) dx. \tag{19}$$

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:

$$\sigma_k = \mathbb{E}\left[|X - \mu_1|^k\right] = \int_{-\infty}^{\infty} (x - \mu_1)^k f_X(x) dx. \tag{20}$$

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, skewness and kurtosis.

Marginal distribution: 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

$$f_{X_j}(x_j) = \int_{D_{x_{\sim j}}} f_X(y_{\sim j}) dy_{\sim j} \tag{21}$$

with the simplified notation $y_{\sim j} = (y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_M)^T$. 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

$$x_N(Z_j) = \int_{D_{x_{\sim j}}} \sum_{k=0}^{N} x_k \Psi_k(Z) dZ_{\sim j}. \tag{22}$$

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

Estimation of the probability density and marginal 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

$$\tilde{f}(x) = \frac{1}{n} \frac{\text{Number of } x_j \text{ in same section as } x}{\text{Width of section}} \tag{23}$$

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 K

$$\tilde{f}(x) = \frac{1}{Mh} \sum_{i=1}^{M} K\left(\frac{x - x_i}{h}\right). \tag{24}$$

Here M is the sample size and $h > 0$ is a smoothing parameter. 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.

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.

Interval Estimation: The objective of an interval estimate is to determine the values f_L and f_R that bound the location of the true value $f_L \leq f \leq f_R$. The estimate is based on a set of realizations $f = [f_1, \ldots, f_M]$ of the random variable and the interval $[f_L, f_R]$ is called an interval estimator. A confidence interval is the combination of an interval estimator and a confidence coefficient α. The confidence coefficient can be interpreted as the probability that the interval estimator contains the true value f. The $(1-\alpha) \times 100\%$ confidence interval for $[f_L, f_R]$ is defined such that for all $f \in \mathbb{F}$,

$$P\left[f_R(X) \leq f\right] = \frac{\alpha}{2} \text{ and } P\left[f_L(X) \leq f\right] = 1 - \frac{\alpha}{2}. \tag{25}$$

MODEL AND APPLICATION

This section first introduces the hydraulic model that gives the basis for the subsequent investigations of the uncertainties, followed by details on the application of the Polynomial Chaos expansion.

Hydraulic Network Model

In hydraulic modeling the simplified topological structure of a water distribution network is described by a graph like the one shown in Figure 2.

In this graph the links represent pipe sections and nodes the pipe junctions. The mathematical description of this graph is given by the incidence matrix $\mathbf{A} \in \mathcal{M}^{nj \times np}(\mathbb{R})$, where nj is the number of nodes and np is the number of links. The coefficients are defined as follows:

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

Water distribution networks in general have a looped structure and the system state is described by

the potential at the nodes (head) and the current on the links (flow rate). The system equations are given by two sets of equations. First the mass balance at the nodes:

$$\mathbf{A}\mathbf{q} + \mathbf{d} = \mathbf{0}_{nj} \tag{26}$$

where $\mathbf{A}$ is linked to the part of the network that only contains junctions with free potential, $\mathbf{q} \in \mathbb{R}^{np}$ is the vector containing the flow rates and $\mathbf{d} \in \mathbb{R}^{nj}$ is a parameter defining the demands at consumption node. Second the energy equation:

$$\Delta\mathbf{h}(\mathbf{r}, \mathbf{q}) - \mathbf{A}^T\mathbf{h} - \mathbf{A}_f^T\mathbf{h}_f = \mathbf{0}_{np}. \tag{27}$$

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

$$\begin{aligned} \Delta\mathbf{h} \colon \mathbb{R}^{np} \times \mathbb{R}^{np} &\to \mathbb{R}^{np} \\ (\mathbf{r}, \mathbf{q}) &\mapsto \Delta\mathbf{h}(\mathbf{r}, \mathbf{q}). \end{aligned} \tag{28}$$

For turbulent and transitional flow regime, the head-loss in general is a non-linear function of the friction coefficient $\mathbf{r}$ and flow rate $\mathbf{q}$. In the following application the state vector $\mathbf{x}$ consists of the flow rates $\mathbf{q}$ and the head $\mathbf{h}$ and the system parameters are combined in the vector $\mathbf{P}$. An efficient solution algorithm for this non-linear system of equations is given by the Global Gradient Analysis Todini (2011). The sensitivities of head and flow rate with respect to demand can be expressed in a direct form as has been shown by Piller et al. (2016)

$$\mathbf{S} = \begin{pmatrix} \frac{\partial \mathbf{q}}{\partial \mathbf{d}} \\ \frac{\partial \mathbf{h}}{\partial \mathbf{d}} \end{pmatrix} = \begin{pmatrix} -\mathbf{D}^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T\right)^{-1} \\ -\left(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T\right)^{-1} \end{pmatrix}. \tag{29}$$

Here, $\mathbf{D}$ is the diagonal Jacobian matrix containing the derivatives of the head-loss $\Delta\mathbf{h}(\mathbf{r}, \mathbf{q})$ with respect to the flow rate $\mathbf{q}$.

A similar formulation is also defined for the sensitivities with respect to the pipe resistance.

$$\mathbf{S} = \begin{pmatrix} \frac{\partial \mathbf{q}}{\partial \mathbf{r}} \\ \frac{\partial \mathbf{h}}{\partial \mathbf{r}} \end{pmatrix} = \begin{pmatrix} \mathbf{D}^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B} - \mathbf{D}^{-1}\mathbf{B} \\ \left(\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B} \end{pmatrix}. \tag{30}$$

Matrix $\mathbf{B}$ is the diagonal Jacobian matrix for the derivatives of the head-loss $\Delta\mathbf{h}(\mathbf{r}, \mathbf{q})$ with respect to $\mathbf{r}$

The direct calculation of the sensitivities allows for an extremely efficient implementation of FOSM methods even for high dimensional parameter spaces.

Stochastic Hydraulic Model

For the polynomial chaos expansion, the objective of the stochastic hydraulic model is to calculate the coefficients for the uncertainties of the QoIs. To do so there exist two major frameworks with the intrusive and non-intrusive methods. Intrusive methods like the Galerkin projection require a reformulation of the system equations with the expansion coefficients as variables. This in general requires the creation of new solvers and is not easily applicable to existing software. Non-intrusive methods on the other hand use a stochastic collocation approach, which uses evaluations of the deterministic model for the calculation of the coefficients.

In applying the intrusive PCE to the hydraulic equations first all uncertain variables are replaced by the polynomial series expansion of the order $N + 1$. Then, this approximation is projected onto the $N + 1$ polynomial basis functions.

$$\langle \mathbf{A}\mathbf{q}_N + \mathbf{d}_N, \Psi_k \rangle = \mathbf{0} \tag{31}$$

$$\left\langle \Delta\mathbf{h}(\mathbf{r}_N, \mathbf{q}_N) - \mathbf{A}^T\mathbf{h}_N - \mathbf{A}_f^T\mathbf{h}_f, \Psi_k \right\rangle = \mathbf{0} \tag{32}$$

It can be easily seen, that through this projection a new augmented system of equations is formulated which contains $N + 1$ times the number of equations as the original system. This new system of

equations is used to directly evaluate the coefficients for the expansion of the QoIs. The direct calculation of the weighted inner product for the head-loss term is not possible, since the function contains a rational power of the flow rate vector $\mathbf{q}$. In order to render the evaluation of the inner product possible a polynomial approximation of the head-loss is introduced. Here, the Taylor series expansion is chosen for formulating the polynomial series expansion Le Maître and Knio (2010). The second order Taylor series expansion is given as:

$$\mathbf{\Delta h} \approx \mathbf{\Delta h}|_{\tilde{\mathbf{q}}} + \nabla \mathbf{\Delta h}|_{\tilde{\mathbf{q}}} \delta \mathbf{q} + \frac{1}{2} \delta \mathbf{q}^T \mathbf{H}|_{\tilde{\mathbf{q}}} \delta \mathbf{q}. \tag{33}$$

where $\mathbf{\Delta h}|_{\tilde{\mathbf{q}}}$ is the head loss and $\mathbf{H}|_{\tilde{\mathbf{q}}}$ is the Hessian matrix at the expansion point $\tilde{\mathbf{q}}$. The flow rate vector calculated for the mean values of all parameters is given by $\tilde{\boldsymbol{q}}$ and $\delta \boldsymbol{q}$ is the difference between the actual flow rate $\boldsymbol{q}$ and the development point $\tilde{\boldsymbol{q}}$. An application is given by Pecci et al. (2017) who introduce quadratic approximations for the head-loss function.

For the application of the non-intrusive projection, in this article, a matrix inversion approach is chosen. It uses multiple solutions of the deterministic system and the corresponding values of the orthogonal basis polynomials to estimate the coefficients using linear regression.

$$\begin{bmatrix} \Psi_0(Z^1) & \dots & \Psi_N(Z^1) \\ \vdots & \ddots & \vdots \\ \Psi_0(Z^M) & \dots & \Psi_N(Z^M) \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} u(d(Z^1)) \\ \vdots \\ u(d(Z^M)) \end{bmatrix} \tag{34}$$

The M collocation points $\{Z^m\}_{m=1}^M$ are given by the samples of the random parameter space. Here, a Monte Carlo sampling strategy is used for the non-intrusive calculation of the polynomial chaos expansion coefficients. It has to be mentioned that this approach is one of the most basic non-intrusive methods and that more efficient sampling strategies are available Xiu (2010) and that current research puts serious effort into improving the efficiency of sampling for high dimensional parameter spaces Blatman and Sudret (2011).

ILLUSTRATIVE MODEL

The illustrative model was used for the first implementation and validation of both the intrusive and non-intrusive PCE algorithms. Its topology is shown in Figure 2. The tree structure is convenient because the flow rates are calculated by a set of linear equations which facilitates the validation process. However, the application can be extended to more complex use cases like looped networks without changes. The parameters of the mathematical model have been introduced in Section 3 and are given by the incidence matrix $\mathbf{A}$, the demand vector $\mathbf{d}$, the friction vector $\mathbf{r}$ and the fixed head potential vector $\mathbf{h_f}$. The quantities of interest are the flow rate vector $\mathbf{q}$ and the head vector $\mathbf{h}$.

Scenario

The case presented in this scenario introduces a system with a single uncertain parameter. This parameter is defined by the demand multiplier at node 10 of the network depicted in Figure 2. The parameter uncertainty is modeled as Gaussian random variables $\mathcal{N} \sim (1, 0.3)$. This way the mean of the demand flow rate, which is defined as the base demand times the demand multiplier, has its mean at the base demand and is highly unlikely to become negative. A total number of $N_{samples} = 1e5$ samples are generated using Latin Hypercube Sampling (LHS) in order to ensure an error of approximately $e = 0.1\%$ for the Monte Carlo simulation. The generated distribution is shown in Figure 3

Results

This section compares the results from the Monte Carlo simulation, first order perturbation method and the PCE. The results for the quantities of interest are illustrated by the flow rate through pipe 5 and the head at node 5. These elements are representative for the results in the rest of the network.

On the left side of Figure 4 the resulting probability distribution for the flow rate through link 5. The abscissa gives the flow rate q_5 in l/s and the ordinate quantifies its probability $P(q_5)$. The result of the Monte Carlo simulation uses the full sample size of $N_{mc} = 1e5$ evaluations and is given by the histogram. The yellow curve gives the result from the First Order Second Moment method, which calculates the mean and the variance and is approximated as a Gaussian distribution.

Finally the red curve shows the result of a first order PC expansion, which has been evaluated using a sample size of $N_{pce} = 1e2$. All three methods are in good agreement for the quantification of uncertainty in the flow rate q_5 . Information on the convergence behavior of both the Monte Carlo and the PCE approach are given on the right hand side of Figure 4 in blue and red respectively. The graphs show the estimated mean and standard deviation as a function of the number of samples that have been used for the evaluation. A logarithmic scale is used for the abscissa.

Figure 5 shows the results for the head h_5 in m_{H_2O} at node 5 on the left hand side and is again completed by the convergence plots in mean and standard deviation on the right hand side. For the FOSM method the estimate probability density function is once again chosen to be Gaussian. However, in this case the approximation is not as accurate as the results from both the Monte Carlo simulations and the fourth order PCE that give a non-symmetric distribution that is skewed in the direction of lower heads.

Taking a close look at the methods used for quantifying the uncertainties and the structure of the model they have been applied to, it is possible to explain the difference in performance shown by Figures 4 and 5. In the special case of the hydraulic model defined in Section 3 the flow rates can be calculated by the set of linear Equations (26) due to the tree structure of the network. From this it follows for one that the FOSM method uses a first order or linear approximation of the random variable for the calculation of the first and second moment. Similar to this it can be shown that for a Gauss-Hermite PCE any Gaussian distribution can be described exactly by a first order expansion with the coefficient u_0 as mean value and u_1 as standard deviation. In conclusion it can be argued that both methods give the exact result in this special case and the result is confirmed by the Monte Carlo simulation.

As introduced in Section 3 the head is strongly influenced by the non-linear head-loss function. That explains in part the different results in Figure 5. Once again the Monte Carlo simulation with the full sample size $N_{mc} = 1e5$ gives the result for validation. The FOSM method produces a Gaussian distribution that is based on the sensitivity of the system and is not able to capture the asymmetry of the probability distribution. Figure 5 shows a good fit in the central part even for a

substantial variance in the demand. Major problems in the description are limited to the distribution tails. In contrast, a PCE of the order four is capable to give an accurate description of the head probability distribution that is in agreement with the Monte Carlo simulation.

Discussion

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 a superior convergence behavior Smith (2013). This is illustrated in the convergence plots for the mean and standard deviation on the right hand side of Figures 4 and 5. For a low dimensional problem as the one discussed in this scenario a very small number of points is sufficient to get a good estimation of the PCE coefficients, while the Monte Carlo method with pseudo-random LHS needs more than $1e4$ samples to give comparable accuracy.

Intrusive versus Non-Intrusive Methods: As introduced in Section 3 there exist two basic approaches to calculate the expansion coefficients with the intrusive and non-intrusive methods. For the application of the PCE in this article both have been tested and the results are shown in Table 1.

The coefficients for the Galerkin approach have been obtained by using a second order Taylor expansion for approximating the non-linear head-loss function. For the non-intrusive matrix inversion, the coefficients have been calculated using sample sizes of 5 and 10 . The table shows that the coefficients u_0 to u_2 calculated by the Galerkin projection are in good agreement with the NISP calculations. Coefficients of an order higher than the expansion of the head-loss function are effectively zero, suggesting that in future applications the expansion order should match that of the PCE. The coefficients from the NISP results no major difference can be identified between the calculations based on 5 and 10 samples underlining the efficiency of the stochastic collocation approach. However, under the circumstances of the example it is probable that the 10 sample NISP is more accurate due to the higher number of collocation points and with respect to the

second order approximation of the head-loss function for the Galerkin projection. Applying the intrusive approach is challenging since the newly created set of equations changes for the addition of new, uncertain input parameter 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 non-intrusive approach more flexible. Further, the non-intrusive method does not require the approximation of the non polynomial head loss term in the system equations, which means that it contains the correct non-linear behavior. In conclusion, it can be stated that in small applications like the one presented in this article the matrix inversion is the better approach due to the flexible application and efficient sampling.

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 modeled 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. As can be seen in Table 1 the coefficients decrease with order of the expansion and approach zero for higher orders, which justifies the series truncation.

MEDIUM SIZE MODEL

In this section the scenario is extended with respect to the network size as well as in the dimension of the parameter space. First, the subject network is defined by a realistic model that has been supplied by Veolia through the ResiWater project. This network contains 2175 pipes, 1822 nodes and one reservoir in the highly looped region. Its topology is illustrated in Figure 6. Second, parameter uncertainties are defined by a 12 dimensional model in demand flow rates and

pipe roughness.

Scenario

For the scenario the network is divided into the 6 different regions using the k-means algorithm. The regions are illustrated in Figure 6. For each of the regions an uncertain demand multiplier is applied to the nodes with their respective base demand and a roughness multiplier is applied to the design roughness of each pipe. All of the 12 parameters are independent and normally distributed random variables. The demand multipliers are defined as $\mathcal{N} \sim (2, 0.6)$ in order to peak demand period flow rates and avoid negative demands in the model. The roughnesses are chosen as $\mathcal{N} \sim (2, 0.3)$ in order to simulate the aging process of the pipes, as the mean roughness is expected to rise with age and it is unlikely to be lower than the design value. The parameter space has been chosen to be based on Gaussian distributions to make the results comparable to similar studies; however, using the PCE the design of more complex parameter spaces with multiple different probability distributions is possible without the loss of generality. Samples form the 12 dimensional distribution are created using a Latin Hypercube Sampler. In order to ensure a very low estimated error of $e = 0.1\%$ for the Monte Carlo simulation a maximum sample size of $N_{samples} = 1e5$ has been generated. Figure 7 shows the realizations for one of the demand multipliers and one of the roughness multipliers.

Results

As for the illustrative example the results are presented by the probability density function, which is tested for normality using the Pearson χ^2-test and the convergence plots in mean and standard deviation. The evaluation of the results is limited to the use of the Monte Carlo simulation and non-intrusive PCE method. For the Monte Carlo simulation represented by the histogram the full sample size of $N_{mc} = 1e5$ has been used. In comparison, the PCE with the NISP approach has been developed on the basis of a total number of $N_{pce} = 5e2$ evaluations of the full hydraulic system in order to calculate the coefficients for a 6th order expansion. The kernel density estimation uses $N_{kde} = 1e5$ evaluations of the resulting meta-model, which is computationally insignificant.

Figure 8 shows the estimated probability distribution of the flow rate through one of the pipes

in the highly looped section of the network that is located next to the reservoir. In the χ^2-test it is confirmed that the flow rate is distributed normally. The location of the pipe suggests that the demand flows from the all over the network are collected at this location which explains the high flow rates and the Gaussian distribution. For the convergence it can be seen that the PCE is able to give a very precise estimate in mean and variance of the flow while the Monte Carlo simulation starts converging after a number of $1e4$ samples. This fast convergence of the PCE could also be observed for the flow rate in the illustrative model.

In Figure 9 the flow is illustrated for one of the pipes in the highly looped area with a medium distance from the reservoir. The distribution is slightly skewed. This is confirmed by the rejection of the null hypothesis in the χ^2-test. For the mean and especially the standard deviation it can be seen, that convergence for the PCE is reached with about $1e3$ samples. Although the flow rate distribution in this part of the network is clearly non-Gaussian, in the same part of the network also contains pipes with similar characteristics where the flow rates were classified as normally distributed.

Figure 10 show the probability distribution for the head at one of the nodes in the looped region of the network. In contrast to the flow rate distribution this result is very characteristic for the head distributions all over the network. The form of the distribution can likely be explained from the head-loss function that is approximately quadratically dependent on the flow rate. For a normally distributed flow rate this leads to χ^2 distribution of the head-loss and the $1-\chi^2$ characteristic in the remaining head. The convergence in mean head and head variance once again shows that the PCE stabilizes around $n_{samples} = 1e3$ while the Monte Carlo simulation needs more than $1e4$ samples. As before for it can be seen that for non-Gaussian distributions the PCE needs considerably more samples than for close to Gaussian distributions.

Discussion

In the flow rates few generalizations can be made from the results of the high dimensional scenario. For one it can be seen that flow rates close to the reservoir follow a Gaussian distribution. This is most likely due to the effect of demand agglomeration. As the network is supplied by a

single reservoir all demands have to flow through these pipes. In contrast, for arbitrary pipes that are part of the loop structure in the network it is not possible to give any sort of generalization. There results show some flow rate distributions that are classified as Gaussian and others as non-Gaussian by the χ^2-test with no apparent difference to the topological position. However, the results from the illustrative network example has some added relevance, as it describes the behavior of the tree structures in the network.

For the head on the other hand, it can be seen that the distribution generally has a characteristic form similar to the one shown in Figure 10. As stated before, this form can be directly explained by the close to quadratic relation in the head-loss formula which leads to a χ^2 distribution. Similar results have been reported for a study by Piller and Brémond (2002) who introduced a stochastic consumer model based on a binomial distribution and propagate the uncertainties using MCS and the FOSM. In contrast, Hwang et al. (2017) report that the probability density function for the nodal heads behave more Gaussian for looped network. This can be explained through the reduced sensitivity. The existence of alternative water paths in a looped system makes the nodal pressure in the network grid less sensitive to changes in the parameters. Even though the probability density function follows a Gaussian form in cases of low sensitivity, from a mathematical point of view this should not be interpreted as a linearity in the system.

For the application of the PCE with the high dimensional parameter space a number of issues have to be mentioned. First, for an efficient evaluation of the Monte Carlo approach the samples have been generated using Latin Hypercube sampling. While this makes the MCS more efficient, this benefit is not limited to the method as the PCE collocation approach profits from this in the same way. Second, the Model and Application section shows that the coefficients are estimated based on the linear regression Equation 34.

The calculation of the coefficients involves the inversion of a $M \times N$ matrix, where M is the number of collocation points and N is the number of coefficients in the meta-model. This matrix inversion can become very demanding or even prohibitive for memory resources in cases of huge sample sizes or a very big number of coefficients. This is not the case for the intrusive PCE

approach. However, as stated in Le Maître and Knio (2010) the development of special solvers is necessary for such applications.

CONCLUSION

Polynomial chaos expansion is a relatively new and efficient tool for the propagation of uncertainties in mathematical models. In this article it has been applied to the hydraulic model of a water distribution network with a high dimensional parameter space for the first time. The application of intrusive and non-intrusive Polynomial Chaos expansion methods to the hydraulic model has shown that it is capable to capture non-linearities in the quantities of interest with comparable accuracy to Monte Carlo simulations. But, in comparison to these classical Monte Carlo simulations the Polynomial Chaos expansion is by far computationally more efficient. In the application of the intrusive Galerkin projection, the reformulation of the system equations for the calculation of the expansion coefficients together with the necessary approximation of the non-polynomial head-loss function has been found to be challenging. Especially since this new set of equations is only applicable to one specific instance of the parameter space. The application of non-intrusive methods has proven to be more generic. It allows for an easy adaption to new input parameters and deals with non-polynomial elements directly through the sampling of parameter space.

In the application to high dimensional parameter spaces experience has shown, that the PCE profits from efficient sampling strategies in the way as the MCS. However, in the fitting of the metamodel current implementations have too calculate the inverse of a matrix whose size is determined by the number samples times the number of elements in the polynomial expansion. Due to this, system memory becomes a limiting factor in practical applications.

Applying the First Order Second Moment method to the hydraulic equations has been shown to be limited since it is not possible to capture non-linear effects. But, due to the direct formulation of the sensitivities and highly efficient calculation of the covariance matrix, this method may still be relevant in cases with high dimensional input spaces and concentrated variance in the uncertain parameters.

In this study, it is assumed that the product from the number of nodal consumers and the

probability of consuming is greater than 5 so that the normal distribution approximation holds. However, the framework of the generalized Polynomial Chaos Expansion also allows to model parameter uncertainties that follow distributions like the Binomial and the Poisson distribution which should be tested in the future. Further, the results presented in this article all use a one dimensional parameter-space. However, the PCE can also be applied to multiple uncertain parameters using a multidimensional orthogonal polynomial basis. First results show that this may produce distributions of the QoI that contain multiple peaks, which are accurately represented by the PCE.

The PCE can also be incorporated in more complex tasks as for example optimization applications. As the PCE is more efficient than MCS for a reasonable number a parameters or group of parameters it provides a computational benefit. The software Dakota developed by Sandia National Laboratories already implements these capabilities.

Going on, there are two major topics of interest with respect to the Polynomial Chaos Expansion. The first one concerns the application of the non-intrusive method which requires a repeated solution of the hydraulic system. For high dimensional parameter spaces the computational complexity may be alleviate through the use of reduced order models like graph decomposition Zheng et al. (2013). Further speed up may also be achieved through the use of graphical processing units as shown by Crous Crous et al. (2012). The second topic is the investigation of the influence of parameter uncertainties on water quality.

Data Availability Statement

The code generated the study are available from the corresponding author by request. The propagation code used during the study is available on-line. (https://www.uqlab.com) The data and model used during the study of the branched network are available from the corresponding author by request. The data and model used during the study of the real network are proprietary in nature and may only be provided with restrictions.

Acknowledgments

The work presented in the paper is part of the French-German collaborative research project ResiWater that is funded by the French National Research Agency (ANR; project: ANR-14-PICS-

0003) and the German Federal Ministry of Education and Research (BMBF; project: BMBF-13N13690).

REFERENCES

Adams, B. M., Bohnhoff, W., Dalbey, K., Eddy, J., Eldred, M., Gay, D., Haskell, K., Hough, P. D., and Swiler, L. P. (2009). "Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: version 5.0 user's manual." *Sandia National Laboratories, Tech. Rep. SAND2010-2183*.

Baudin, M., Dutfoy, A., Iooss, B., and Popelin, A.-L. (2017). "Openturns: An industrial software for uncertainty quantification in simulation." *Handbook of Uncertainty Quantification*, 2001–2038.

Blatman, G. and Sudret, B. (2011). "Adaptive sparse polynomial chaos expansion based on least angle regression." *Journal of Computational Physics*, 230(6), 2345–2367.

Blokker, E. and Van der Schee, W. (2006). "Simulation of water demands provides insight." *Water Supply and Drainage for Buildings, CIB W*, 62.

Boulos, P. F., Lansey, K. E., and Karney, B. W. (2004). *Comprehensive Water Distribution Systems Analysis Handbook For Engineers and Planners*. MWH Soft Inc.

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.

Braun, M., Piller, O., Deuerlein, J., and Mortazavi, I. (2017). "Limitations of demand-and pressure-driven modeling for large deficient networks." *Drinking Water Engineering and Science*, 10(2), 93.

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.

Cacuci, D. G., Ionescu-Bujor, M., and Navon, I. M. (2005). *Sensitivity and uncertainty analysis, volume II: applications to large-scale systems*, Vol. 2. CRC press.

Crous, P., Van Zyl, J., and Roodt, Y. (2012). "The potential of graphical processing units to solve hydraulic network equations." *Journal of Hydroinformatics*, 14(3), 603–612.

Fishman, G. (2013). *Monte Carlo: concepts, algorithms, and applications*. Springer Science & Business Media.

Hagos, M., Jung, D., and Lansey, K. E. (2016). "Optimal meter placement for pipe burst detection in water distribution systems." *Journal of Hydroinformatics*, 18(4), 741–756.

Hart, W. E. and Murray, R. (2010). "Review of sensor placement strategies for contamination warning systems in drinking water distribution systems." *Journal of Water Resources Planning and Management*, 136(6), 611–619.

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.

Hwang, H., Lansey, K., and Jung, D. (2017). "Accuracy of first-order second-moment approximation for uncertainty analysis of water distribution systems." *Journal of Water Resources Planning and Management*, 144(2), 04017087.

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

Lansey, K. (1997). "Uncertainty in water distribution network modeling." *Journal of Contemporary Water Research and Education*, 103(1), 5.

Lansey, K., El-Shorbagy, W., Ahmed, I., Araujo, J., and Haan, C. (2001). "Calibration assessment and data collection for water distribution networks." *Journal of Hydraulic Engineering*, 127(4), 270–279.

Le Maître, O. and Knio, O. M. (2010). *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media.

Lu, Z. and Vesselinov, V. V. (2015). "Analytical sensitivity analysis of transient groundwater flow in a bounded model domain using the adjoint method." *Water Resources Research*, 51(7), 5060–5080.

Marelli, S. and Sudret, B. (2014). *UQLab: A Framework for Uncertainty Quantification in Matlab*. ASCE, 2554–2563.

Ostfeld, A. and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *Journal of Water Resources Planning and Management*, 130(5), 377–385.

Ostfeld, A. and Shamir, U. (1996). "Design of optimal reliable multiquality water-supply systems." *Journal of Water Resources Planning and Management*, 122(5), 322–333.

Pasha, M. and Lansey, K. (2010). "Effect of parameter uncertainty on water quality predictions in distribution systems-case study." *Journal of hydroinformatics*, 12(1), 1–21.

Pecci, F., Abraham, E., and Stoianov, I. (2017). "Quadratic head loss approximations for optimisation problems in water supply networks." *Journal of Hydroinformatics*, jh2017080.

Perelman, L., Housh, M., and Ostfeld, A. (2013). "Robust optimization for water distribution systems least cost design." *Water Resources Research*, 49(10), 6795–6809.

Perelman, L., Maslia, M. L., Ostfeld, A., and Sautner, J. B. (2008). "Using aggregation/skeletonization network models for water quality simulations in epidemiologic studies." *Journal (American Water Works Association)*, 100(6), 122–133.

Piller, O. and Brémond, B. (2002). "A stochastic model for peak period analysis of pipe networks." *ASCE Environmental and Water Resources Systems Analysis*.

Piller, O., Elhay, S., Deuerlein, J., and Simpson, A. R. (2016). "Local sensitivity of pressure-driven modeling and demand-driven modeling steady-state solutions to variations in parameters." *Journal of Water Resources Planning and Management*, 04016074.

Piller, O., Gilbert, D., and Van Zyl, J. E. (2010). "Dual calibration for coupled flow and transport models of water distribution systems." *Water Distribution Systems Analysis*, 722–731.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1988). "Numerical recipes in c." *Cambridge University Press*, 1, 3.

Razavi, S. and Gupta, H. V. (2015). "What do we mean by sensitivity analysis? the need for comprehensive characterization of "global" sensitivity in earth and environmental systems models." *Water Resources Research*, 51(5), 3070–3092.

Rossman, L. A., Boulos, P. F., and Altman, T. (1993). "Discrete volume-element method for network water-quality models." *Journal of Water Resources Planning and Management*, 119(5), 505–517.

Savic, D. A., Kapelan, Z. S., and Jonkergouw, P. M. (2009). "Quo vadis water distribution model calibration?." *Urban Water Journal*, 6(1), 3–22.

Savic, D. A. and Walters, G. A. (1997). "Genetic algorithms for least-cost design of water distribution networks." *Journal of water resources planning and management*, 123(2), 67–77.

Smith, R. C. (2013). *Uncertainty quantification: theory, implementation, and applications*, Vol. 12. SIAM.

Todini, E. (2011). "Extending the global gradient algorithm to unsteady flow extended period simulations of water distribution systems." *Journal of Hydroinformatics*, 13(2), 167–180.

Van Zyl, J. and Cassa, A. (2013). "Modeling elastically deforming leaks in water distribution pipes." *Journal of Hydraulic Engineering*, 140(2), 182–189.

Walski, T. M., Chase, D. V., Savic, D. A., Grayman, W., Beckwith, S., and Koelle, E. (2003). *Advanced water distribution modeling and management*. Haestad press.

Wiener, N. (1938). "The homogeneous chaos." *American Journal of Mathematics*, 60(4), 897–936.

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

Xiu, D. and Karniadakis, G. E. (2002). "The wiener–askey polynomial chaos for stochastic differential equations." *SIAM journal on scientific computing*, 24(2), 619–644.

Zheng, F., Simpson, A. R., Zecchin, A. C., and Deuerlein, J. W. (2013). "A graph decomposition-based approach for water distribution network optimization." *Water Resources Research*, 49(4), 2093–2109.

Zwillinger, D. (2002). *CRC standard mathematical tables and formulae*. Chapman and Hall/CRC.

List of Tables

1 Calculated coefficients for the polynomial chaos expansion using the intrusive Galerkin projection and the non-intrusive matrix inversion approaches for the flow rate through pipe 5 q_5 and the head at node 5 h_5. 33

	Galerkin		Non-Intrusive Spectral Projection			
	2nd-order Taylor		5 Samples		10 Samples	
	q_5	h_5	q_5	h_5	q_5	h_5
u_0	21	845.2393	21	845.2393	21	844.8645
u_1	10	-80.6156	10	-80.6156	10	-80.2202
u_2	0	-15.0044	1e-8	-15.0044	3e-13	-15.7426
u_3	0	-3e-18	1e-9	0.4011	2e-13	0.5413
u_4	0	0	1e-11	-0.0202	2e-15	-0.1009

TABLE 1. Calculated coefficients for the polynomial chaos expansion using the intrusive Galerkin projection and the non-intrusive matrix inversion approaches for the flow rate through pipe 5 q_5 and the head at node 5 h_5.

List of Figures

1 Sources for parameter uncertainties in hydraulic water distribution network models. 35

2 Illustrative network graph with a tree structure. 36

3 Parameter uncertainties for demand $Z \sim \mathcal{N}(1, 0.3)$. 37

4 Estimated probability density function for the flow rate through link 5 using FOSM, MCS and a 1st-order PCE. Comparison of convergence in mean and variance as a function of sample size. 38

5 Estimated probability density function for the head at node 5 using FOSM, MCS and a 8th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 39

6 Clusters for parameter uncertainties in the medium size network. 40

7 Exemplary parameter uncertainties for demand multiplier $Z_{1,\ldots,6} \sim \mathcal{N}(2, 0.6)$ and roughness multiplier $Z_{7,\ldots,12} \sim \mathcal{N}(2, 0.3)$. 41

8 Estimated probability density function for flow rate in looped network area close to the reservoir using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 42

9 Estimated probability density function for flow rate in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 43

10 Estimated probability density function for head in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size. 44

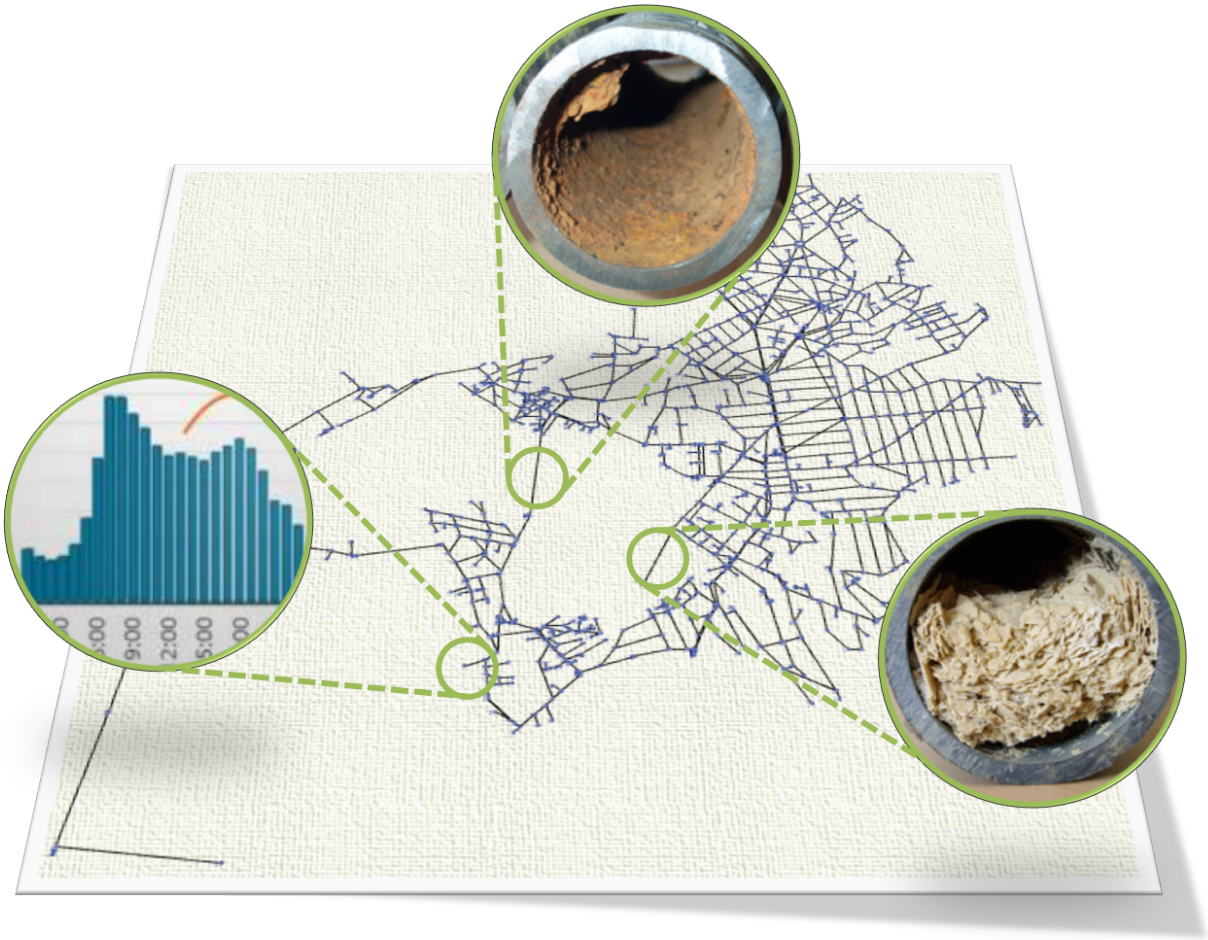

Fig. 1. Sources for parameter uncertainties in hydraulic water distribution network models.

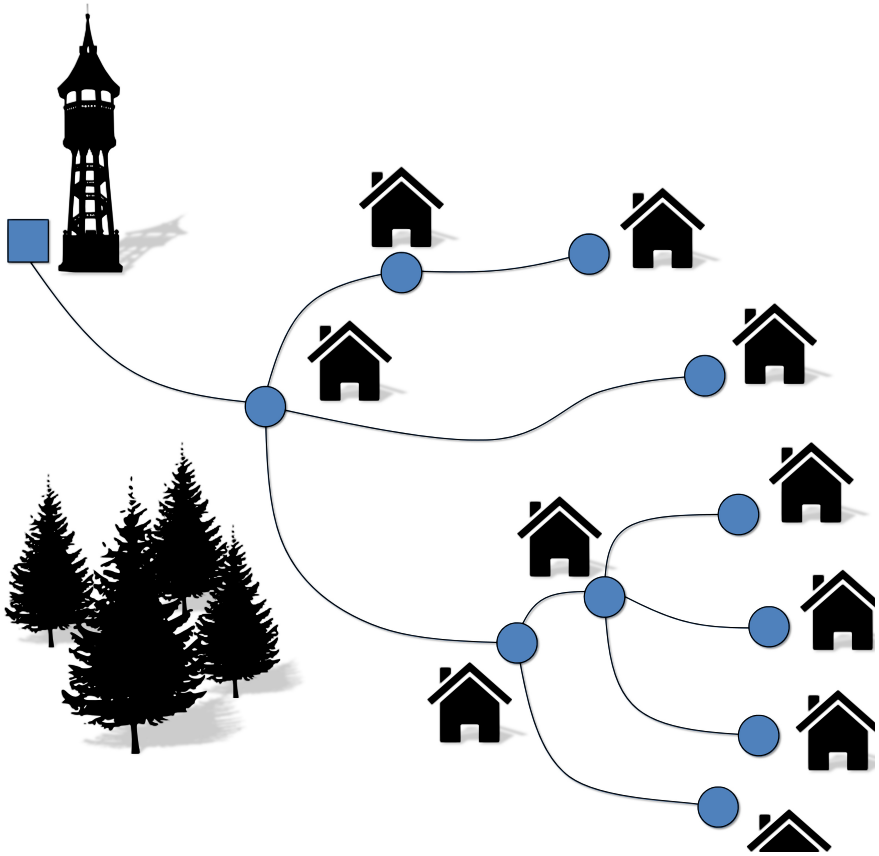

Fig. 2. Illustrative network graph with a tree structure.

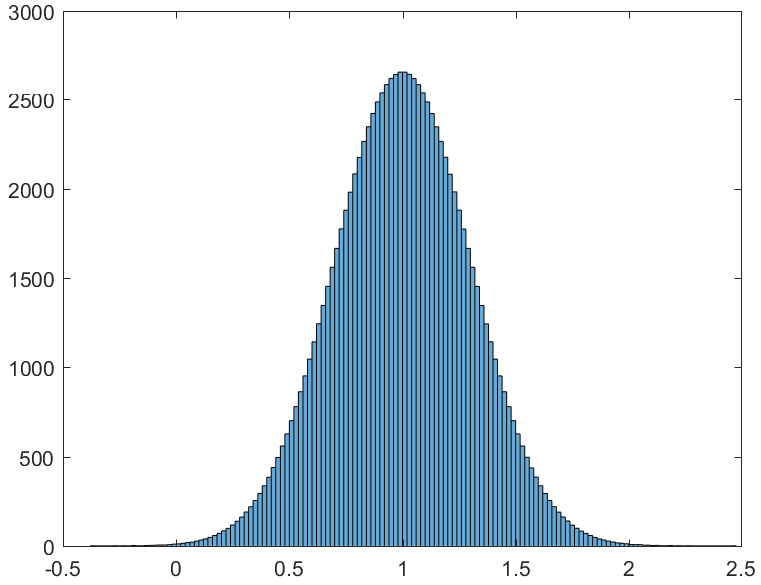

Fig. 3. Parameter uncertainties for demand $Z \sim \mathcal{N}(1, 0.3)$.

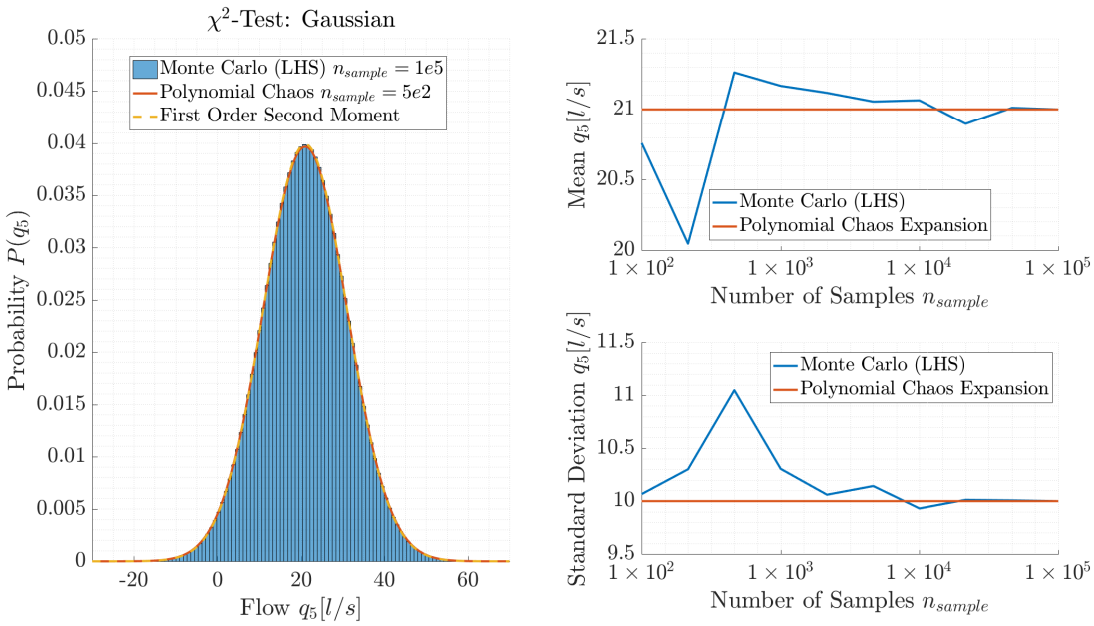

Fig. 4. Estimated probability density function for the flow rate through link 5 using FOSM, MCS and a 1st-order PCE. Comparison of convergence in mean and variance as a function of sample size.

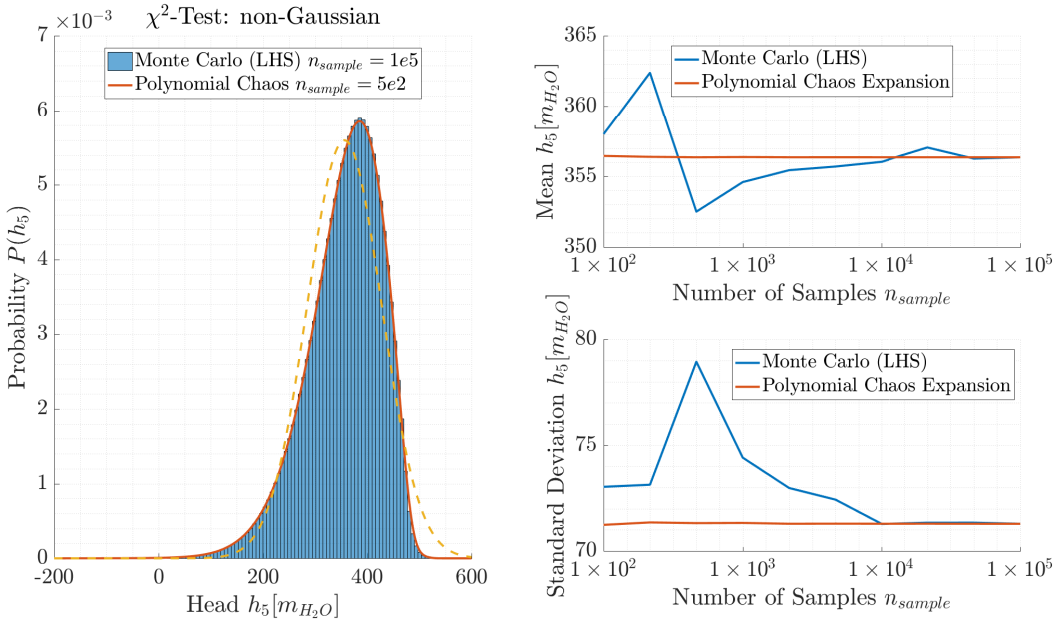

Fig. 5. Estimated probability density function for the head at node 5 using FOSM, MCS and a 8th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

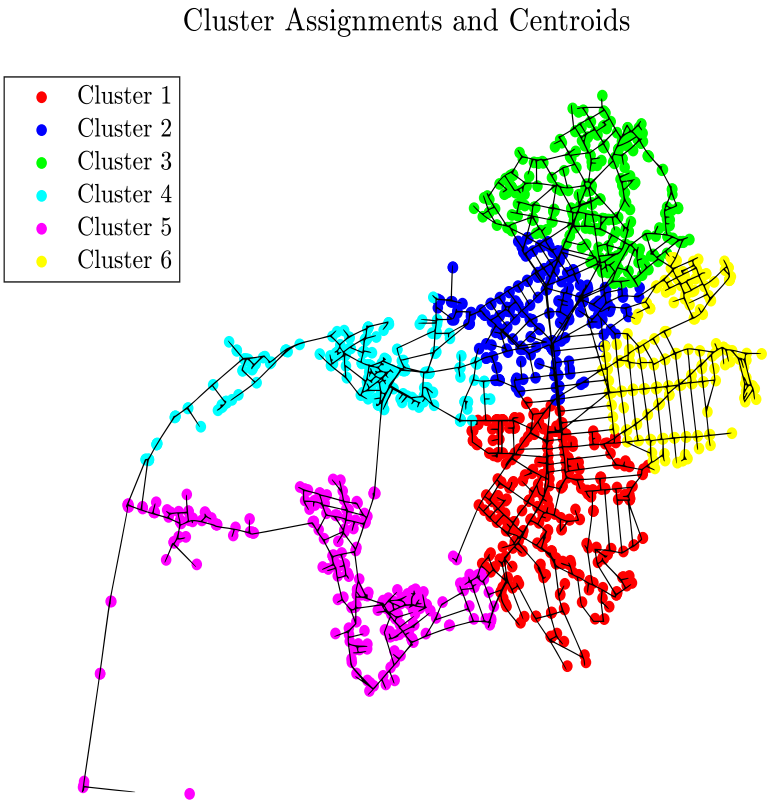

Fig. 6. Clusters for parameter uncertainties in the medium size network.

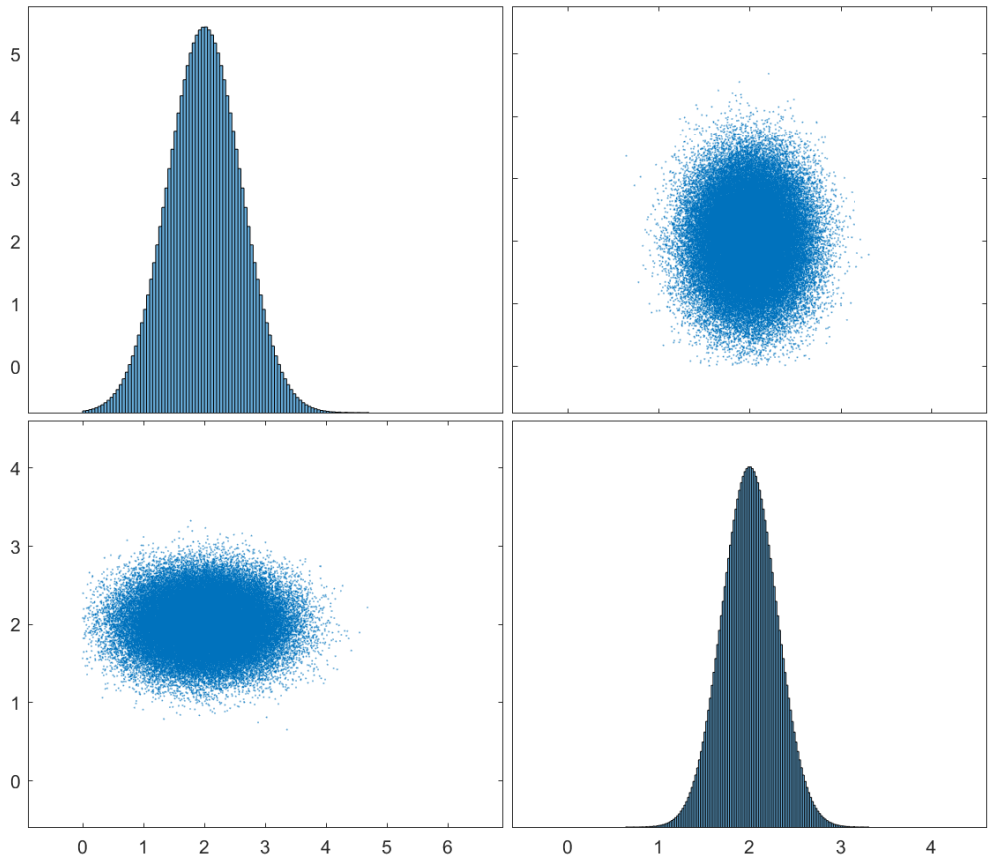

Fig. 7. Exemplary parameter uncertainties for demand multiplier $Z_{1,\ldots,6} \sim \mathcal{N}(2, 0.6)$ and roughness multiplier $Z_{7,\ldots,12} \sim \mathcal{N}(2, 0.3)$.

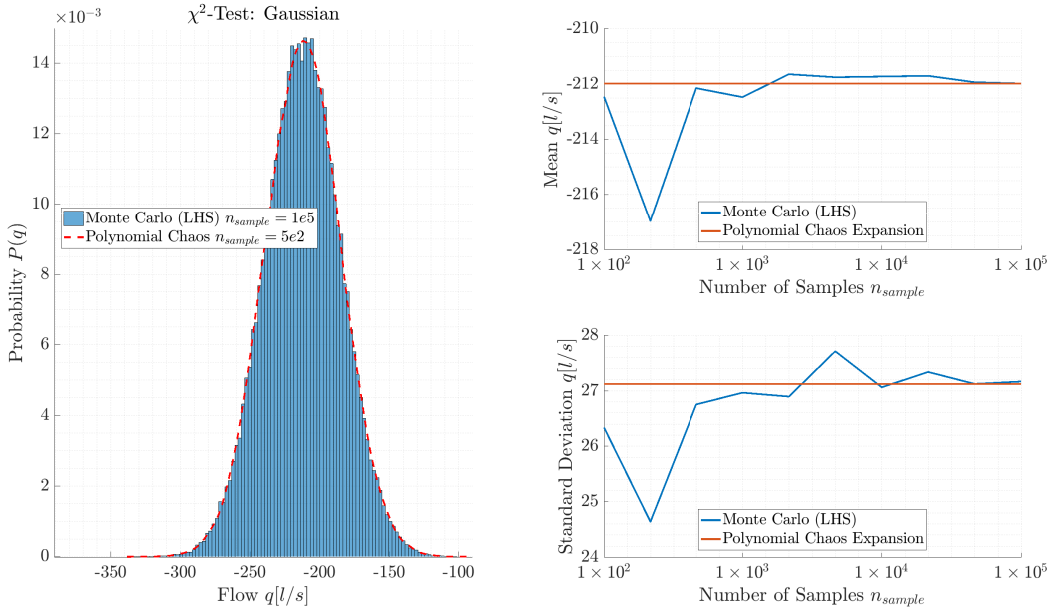

Fig. 8. Estimated probability density function for flow rate in looped network area close to the reservoir using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

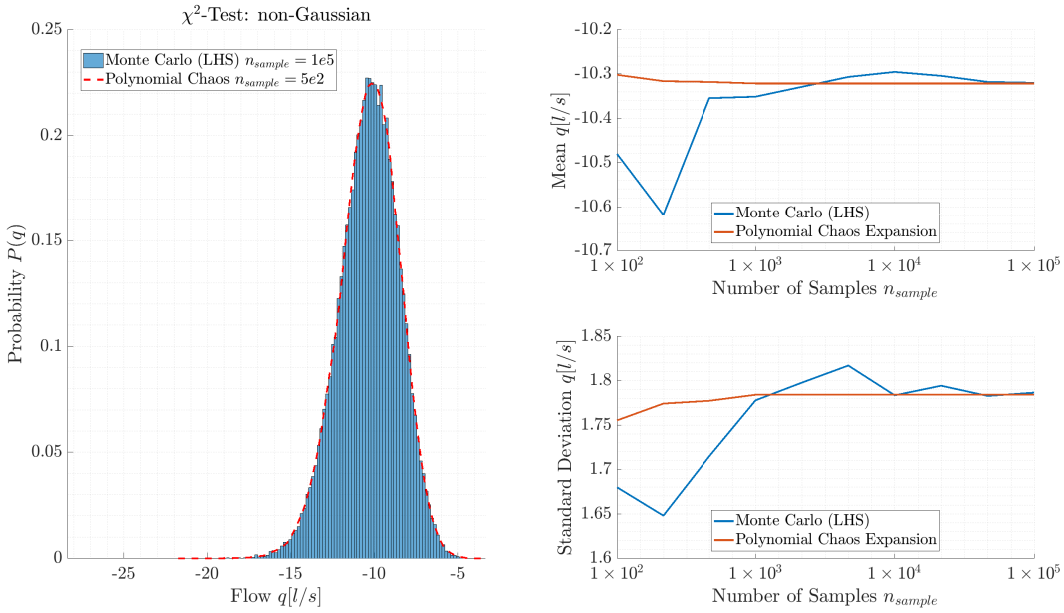

Fig. 9. Estimated probability density function for flow rate in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.

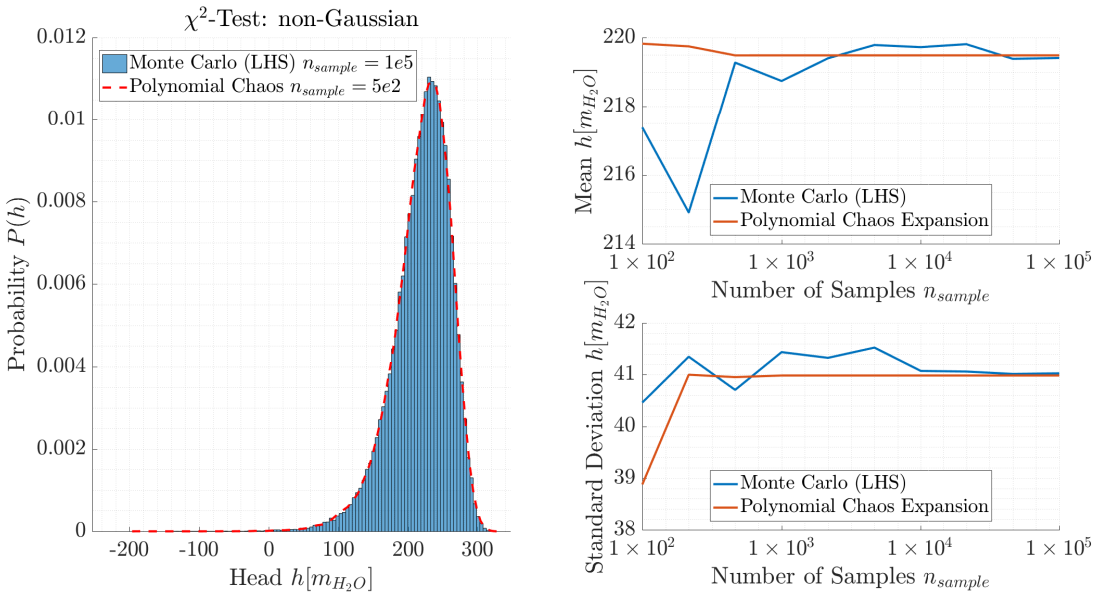

Fig. 10. Estimated probability density function for head in looped network area using Monte Carlo simulations and a 6th-order PCE. Comparison of convergence in mean and variance as a function of sample size.