

# Uncertainty quantification of water age in water supply systems by use of spectral propagation

Mathias Braun, Olivier Piller, Jochen Deuerlein, Iraj Mortazavi and Angelo Iollo

## ABSTRACT

Water distribution networks are critical infrastructures that should ensure the reliable supply of high quality potable water to its users. Numerical models of these networks are generally governed by many parameters for which the exact value is not known. This may be due to a lack of precise knowledge like for consumer demand or due to a lack of accessibility as for the pipe roughness. For network managers, the effect of these uncertainties on the network state is important information that supports them in the decision-making process. This effect is generally evaluated by propagating the uncertainties using the mathematical model. In the past, perturbation, fuzzy and stochastic collocation methods have been used for uncertainty propagation. However, these methods are limited either in the accuracy of the results or the computational effort of the necessary calculations. This paper uses an alternative spectral approach that uses the polynomial chaos expansion and has the potential to give results of comparable accuracy to the Monte Carlo sampling through the definition of a stochastic model. This approach is applied to the hydraulic model of two real networks in order to evaluate the influence of uncertain demands on the water age.

**Key words** | multimodal distribution, real network, uncertainty quantification, water age, water distribution systems

**Mathias Braun** (corresponding author)  
**Olivier Piller**  
Irstea, Dept. of Water,  
Bordeaux Regional Centre, UR ETBX,  
50 Ave. de Verdun, Gazinet, F-33612 Cestas,  
France  
E-mail: [mathias.braun@irstea.fr](mailto:mathias.braun@irstea.fr)

**Mathias Braun**  
**Angelo Iollo**  
Inria Bordeaux-Sud-Ouest, Team MEMPHIS,  
200 Avenue de la Vieille Tour, 33405 Talence,  
France

**Olivier Piller**  
**Jochen Deuerlein**  
School of Civil, Environmental and Mining  
Engineering,  
University of Adelaide,  
South Australia 5005, Australia

**Jochen Deuerlein**  
3S Consult GmbH,  
Albtalstrasse 13, 76137 Karlsruhe, Germany

**Iraj Mortazavi**  
CNAM Paris, EA-7340-M2N-Modélisation  
Mathématique et Numérique,  
75003 Paris, France

**Angelo Iollo**  
Université de Bordeaux, IMB,  
33400 Talence, France

## INTRODUCTION

Evaluating the quality of drinking water supplied by a distribution system is usually done by using quality indicators like water age or chlorine concentration. These indicators may be determined from mathematical network models. The model of a water distribution network depends on a great number of parameters, like the nodal water demand, pipe roughness and reservoir or tank levels. However, the exact values of these parameters are rarely known and thus introduce parameter uncertainty to the model.

Uncertainties in mathematical models are generally divided into aleatory and epistemic uncertainties. Aleatory uncertainties are the result of inherently random processes and are often described by probability distributions that are

modelled on observations. These uncertainties are regarded as irreducible since the underlying process cannot be influenced. In contrast, epistemic uncertainties follow from incomplete insight into the system, which in theory could be eliminated through perfect knowledge. However, epistemic uncertainties are often the necessary effect of simplifications in the modelling process and the numerical evaluation (Der Kiureghian & Ditlevsen 2009; Smith 2013).

Parameter uncertainties are one of the most important aspects of hydraulic modelling as they can result in large prediction errors. Hidden parameters like the pipe roughness that cannot be observed directly are often estimated through calibration (Savic *et al.* 2009; Piller *et al.* 2010).

But, even for a well-calibrated network some uncertainty remains. Other parameters, like for example, the demand, are clearly aleatory in nature.

The objective of uncertainty analysis (UA) is the evaluation and mitigation of such parameter uncertainties on the quantities of interest (QoI). The central part of this process is the uncertainty propagation by means of the mathematical model. Classical approaches contain perturbation, stochastic collocation and fuzzy methods. Stochastic spectral methods are a relatively new addition to the field that has been introduced by [Xiu \(2010\)](#).

Perturbation 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 ([Savic \*et al.\* 2009](#)). This limits their accuracy for highly non-linear models. Stochastic sampling methods like the Monte Carlo simulation are some of the most prominent algorithms for uncertainty propagation. This is due to the easy implementation and the ability to be used for non-linear systems. However, it can be computationally demanding as its rate of convergence is proportional to  $1/\sqrt{M}$ , where  $M$  is the number of simulations ([Cacuci \*et al.\* 2005](#); [Savic \*et al.\* 2009](#)). The objective of stochastic spectral methods like stochastic Galerkin and non-intrusive spectral projection is the calculation of a spectral representation of the random variables. Utilizing the smoothness requirement of the polynomial basis leads to a more efficient convergence behaviour ([Xiu 2010](#); [Smith 2013](#)).

The effect of uncertainties in the context of water distribution networks has been investigated in a number of publications with a wide variety of interests. Special interest lies in the inclusion of parameter uncertainties during the network design. [Tsakiris \*et al.\* \(2017\)](#) use a fuzzy approach to uncertainty propagation in order to include the uncertainty in parameters like the pipe roughness. The application leads to useful recommendations for the network design, in the case of medium to low flow velocities. [Fu & Kapelan \(2011\)](#) present a fuzzy probabilistic approach for optimal design and rehabilitation of water distribution systems. The approach is tested on two example networks and the results show the solution to be more reliable than the deterministic Pareto optimal solutions. [Sumer & Lansley \(2009\)](#) propose a method for pipe network extension

optimization and calibration of pipe roughness at the same time. Uncertainties in the pipe roughness values are specified using FOSM analysis and embedded using a D-optimality criterion. The authors show that the mean values of the unknown parameters converged to their true values and also the cost changes showed a monotonic trend.

A second focus lies in the inclusion of uncertainty information into the management of the network.

[Xu & Qin \(2014\)](#) propose an integrated fuzzy programming and decision analysis approach for water distribution system management under uncertainty. The results indicate that the method helps analyse trade-offs between minimization of operation cost and reliability of running the system and linking optimization model outputs with decision analysis. [Raei \*et al.\* \(2018\)](#) apply multi-objective optimization for the design of an optimal pressure sensor network for leak detection. Uncertainty in the pressure measurements is considered by application of a range of thresholds. The detection times are calculated for leaks at any node with varying starting times. However, in this approach, pressure dependency is neither considered for the leakage outflows nor for the demands which lead to unrealistic results in real applications.

Studies on the effect of uncertainties on the water quality have been performed with respect to a number of objectives. [Di Cristo \*et al.\* \(2015\)](#) investigate the effect of parameter uncertainties in the kinetic models for the formation of trihalomethanes (THM). The parameter uncertainties in the chlorine decay parameters and the THM formation parameters are propagated using the first order second moment (FOSM) perturbation approach and Monte Carlo simulation for validation purposes. They conclude that the FOSM is sufficiently accurate for the application and well worth the benefit in computational effort. The study by [Hart \*et al.\* \(2019\)](#) performs a sensitivity analysis on a water quality model in order to identify important factors for the reduction in uncertainty for the selection of water quality sampling points in the case of a contamination event. They employ a full factorial design of experiments where the influences of parameters like valve status, reaction rate and injection location is investigated as well as the effect of stochastic demand variability. The results show that the injection location is the main source for uncertainty for the size and total area of the contamination plume, whereas the effect of the demand

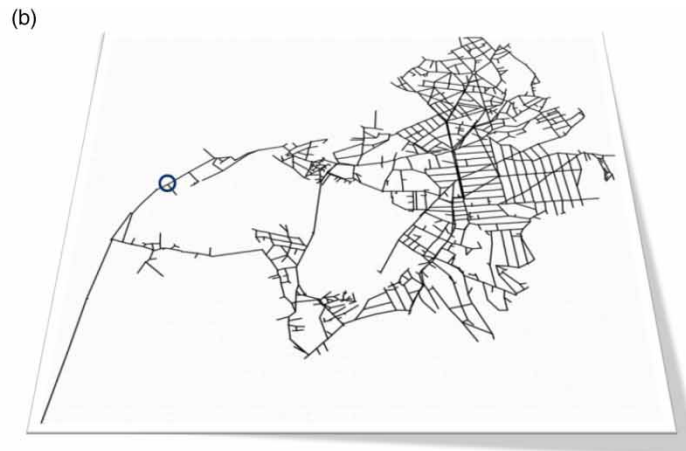
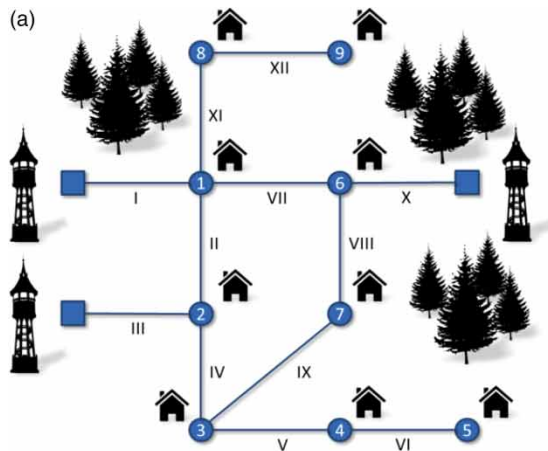
variability is not significant in comparison. The objective of this paper is to analyse the influence of the demand uncertainties on the water age in a network model that has been provided by the partner Veolia in the framework of the French–German project *ResiWater* (2019). Following this introduction, the example network will be introduced together with the basic equations for the hydraulic state and water age in the next section. This is followed by a short introduction of the PCE. Then the two use-cases are discussed after which the results of the investigation are presented. Finally, the conclusions drawn from the application are presented.

## MATERIAL

### Hydraulic model

In hydraulic modelling, the simplified topological structure of a water distribution network is described by a directed graph corresponding to Figure 1. 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  $\mathbf{A} \in \mathbb{M}^{nj \times np}$ , where  $nj$  is the number of junction nodes and  $np$  is the number of links (Jungnickel 2005). The coefficients are defined as follows:

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



**Figure 1** | (a) Strongly aggregated looped network; (b) medium-sized realistic network with the marked evaluation node 15Nf79.

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 classical hydraulic model is defined by the first and second Kirchhoff laws (Cross 1936). First, the mass balance at the nodes:

$$\mathbf{A}\mathbf{q} + \mathbf{d} = \mathbf{0}_{nj} \quad (2)$$

where the node-link incidence matrix  $\mathbf{A}$  includes only nodes with unknown heads,  $\mathbf{q} \in \mathbb{R}^{np}$  is the vector of link flow rates and  $\mathbf{d} \in \mathbb{R}^{nj}$  is the vector of demands at consumption nodes. Second, the energy equation:

$$\Delta\mathbf{h}(\mathbf{r}, \mathbf{q}) - \mathbf{A}^T\mathbf{h} - \mathbf{A}_f^T\mathbf{h}_f = \mathbf{0}_{np} \quad (3)$$

where  $\mathbf{A}_f$  describes the incidence matrix of nodes with fixed potential like reservoirs or tanks and  $\mathbf{h} \in \mathbb{R}^{nj}$  is the vector containing the unknown piezometric heads. Parameters are given by the head vector  $\mathbf{h}_f \in \mathbb{R}^{mf}$  describing fixed heads at special nodes like reservoirs or tanks and the 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 loss in head due to the friction between the fluid and the pipe wall.

### Water quality indicators

In the literature, the water age or residence time in the network is commonly used as an indicator for water quality. Dandy et al. (2013) studied the advantages and the

limitations of the use of water age as a surrogate for water quality. They tried to determine the chlorine concentration based on the residence time. They conclude that in certain situations, water age gives an approximate estimate of water quality and disinfection by-product formation. However, this estimate may be inaccurate due to mixing in tanks and at nodes and is inappropriate for systems with several water sources and/or re-chlorination.

The water age  $c$  is modelled on the basis of the convection–diffusion equation:

$$\frac{\partial c}{\partial t} = \nabla(D\nabla c) - \nabla(\vec{v}c) + R(t, \vec{x}) = 0 \quad (4)$$

The variable of interest  $c$  is usually assumed to be some sort of concentration,  $D$  is the diffusion coefficient,  $\vec{v}$  the velocity field and  $R$  a source term which may be a non-linear function in time and space. For water distribution systems two major simplifications are made. First, an incompressible water flow is assumed, for which extended period simulation (EPS) for the hydraulic is accurate enough (no water hammer or mass oscillation); it follows the hydraulic model is one-dimensional and the velocity vector  $\vec{v}$  simplifies to  $U(t, x) = U_{i,j}$  which is calculated for every hydraulic time step  $t_i$  and constant along  $x$  for each pipe  $p_j$ . Second, it is assumed that the regime is turbulent as a consequence that the longitudinal dispersion can be neglected (Rossman & Boulos 1996; Fabrie et al. 2010). This flow state is often characterized as plug-flow. With these adaptations, the transport equation for water age  $a$  may be derived by replacing  $R$  with a zero-order source term:

$$\frac{\partial c(t, x)}{\partial t} + U_{i,j} \frac{\partial c(t, x)}{\partial x} - 1 = 0 \quad (5)$$

During each hydraulic time step the water age is resolved by a succession of quality time steps. The equation is solved on the domain  $x \in [0, L_j]$  for pipe  $p_j$ , where  $L_j$  is the pipe length and for the time  $t \in [t_i, t_{i+1}]$ . The initial condition  $c(t_i, x_j)$  is defined by the previous time step. The boundary conditions are updated with the perfect mixing assumption for the average water age, or with the min or max value entering.

In the software Porteau (Piller et al. 2011), water age is evaluated in three variables: min, max and flow-averaged

ages. Since the software keeps track of these characteristic water ages during the mixing process at the pipe junctions, it is possible to evaluate the extreme values for the residence time as defined by Equation (6):

$$\begin{aligned} c_n^{upper} &= \max_{i \in S_m} c_i \\ c_n^{lower} &= \min_{i \in S_m} c_i \end{aligned} \quad (6)$$

The third variable, average water age is calculated by the weighted sum of the water age in the water columns mixing at the upstream junction (Piller & Tvard 2014):

$$c_n = \frac{\sum_{i \in S_m} |q_i| c_i}{\sum_{i \in S_m} |q_i|} \quad (7)$$

## UNCERTAINTY PROPAGATION

A recent and promising approach for the propagation of parameter uncertainties is presented in the polynomial chaos expansion (PCE), where a random variable  $X$  is expressed as a truncated spectral series expansion of the order  $N$  as a function of the basic random variable  $Z$  and a set of orthogonal basis polynomials  $\Psi_k$ :

$$X_N(Z) = \sum_{k=0}^N x_k \Psi_k(Z) \quad (8)$$

where the coefficients of the series are given by  $x_k$ . The choice of the basic random variable, also often termed as the germ, and the orthogonal polynomials is joint by the definition of the inner product:

$$\begin{aligned} E[\Psi_i(Z)\Psi_j(Z)] &= \int \Psi_i(Z)\Psi_j(Z)dP_Z(z) \\ &= \int \Psi_i(z)\Psi_j(z)f_Z(z)dz = \delta_i \|\Psi_i\|^2 \end{aligned} \quad (9)$$

Here,  $E$  defines the expected value,  $P_Z$  is the cumulative distribution function and  $f_Z$  is the probability density function of the germ distribution. The Kronecker delta  $\delta_i$  is equal to 1 if the indices are equal and 0 for the product of two different polynomials. The coefficients  $x_k$  of a

random parameter  $X$  are calculated by projecting it on the orthogonal basis polynomials:

$$x_k = \frac{\langle X, \Psi_k \rangle}{\|\Psi_k\|^2} \quad (10)$$

As Equation (7) suggests, the choice of the polynomial basis generally depends on the distribution of the basic random variable. In this paper, uncertainties are assumed to follow a Gaussian distribution which allows for the application of the Wiener chaos with a spectral expansion on the basis of the Hermite polynomials (Xiu & Karniadakis 2002). The use of the PCE is not limited to use with Gaussian distributed variables.

First, it is possible to approximate other distributions on the basis of the Hermite polynomials. Second, the extension to the general PCE defines further sets of orthogonal polynomial basis that can be used with different probability distributions (Xiu 2010). One such example would be the spectral expansion of a uniform distribution on the basis of the Legendre polynomials.

Application of the PCE, in general, follows an intrusive or a non-intrusive procedure. The intrusive approach requires a reformulation of the system equations and allows for the direct solution in the parameters of the QoI. 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 non-linearities may require additional steps that introduce additional errors. Non-intrusive algorithms use a number of samples similar to the Monte Carlo simulation to evaluate the coefficients. This means that in regard to the computation current software and models may be used. Even though the non-intrusive PCE is more efficient than the MCS for high dimensional parameter spaces the problem may become computationally infeasible for applications with high dimensional parameter spaces.

For the application of the PCE to the hydraulic equations, random variables are substituted by their spectral series expansion. This includes parameters (nodal demands  $\mathbf{d}_N$ , pipe resistance  $\mathbf{r}_N$ ) for which the coefficients are determined on the basis of their probability density function and the QoI (flow rates  $\mathbf{q}_N$ , nodal heads  $\mathbf{h}_N$ ) for which the coefficients still have to be determined. This substitution is

formulated in the stochastic system equations:

$$\begin{aligned} \mathbf{A}\mathbf{q}_N + \mathbf{d}_N &= \mathbf{0}_{nj} \\ \Delta\mathbf{h}(\mathbf{r}_N, \mathbf{q}_N) - \mathbf{A}^T\mathbf{h}_N - \mathbf{A}_f^T\mathbf{h}_f &= \mathbf{0}_{np} \end{aligned} \quad (11)$$

In this system of equations, the random variables have been replaced by their PCE of the order  $N$ , following the description in the section 'Material'. There exist two general approaches to evaluate the coefficients of the QoI with their respective benefits and drawbacks.

For the *intrusive* methods, the most prominent approach is the Galerkin projection. It projects the stochastic system equations on the polynomial basis functions  $\Psi_k$  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:

$$\begin{aligned} \langle \mathbf{A}\mathbf{q}_N + \mathbf{d}_N, \Psi_k \rangle &= \mathbf{0}_{nj} \\ \langle \Delta\mathbf{h}(\mathbf{r}_N, \mathbf{q}_N) - \mathbf{A}^T\mathbf{h}_N - \mathbf{A}_f^T\mathbf{h}_f, \Psi_k \rangle &= \mathbf{0}_{np} \end{aligned} \quad (12)$$

Similar to the hydraulic equations the partial differential equation for the coefficients of the water age may be determined by the following projection:

$$\left\langle \frac{\partial A^K}{\partial t} + U_{ij}^N \frac{\partial A^K}{\partial x} - 1, \Psi_m \right\rangle = 0 \quad (13)$$

Here, the uncertain parameter is given by  $U_{ij}^N$  which is derived from the PCE of the flow  $q_{ij}^N$  from the stochastic hydraulic equations and the QoI is water age  $A^K$ .

The *non-intrusive* method is a stochastic collocation like Monte Carlo simulation. 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 PCE to the data. For a QoI  $X$  vector, which could be either velocity or water age, the coefficients are determined by the following equation:

$$\begin{bmatrix} \Psi_0(\mathbf{z}^{(1)}) & \dots & \Psi_N(\mathbf{z}^{(1)}) \\ \vdots & \ddots & \vdots \\ \Psi_0(\mathbf{z}^{(M)}) & \dots & \Psi_N(\mathbf{z}^{(M)}) \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} = \begin{bmatrix} \mathbf{x}(\mathbf{y}(\mathbf{z}^{(1)}))^T \\ \vdots \\ \mathbf{x}(\mathbf{y}(\mathbf{z}^{(M)}))^T \end{bmatrix} \quad (14)$$

where  $\mathbf{z}^{(i)}$  are realizations of the  $\mathbf{Z}$  random vector and  $\Psi_k(\mathbf{z}^{(i)})$  is the value of the  $k$ -th polynomial basis function at realization  $\mathbf{z}^{(i)}$ . The right-hand side is determined by the simulation result



$\mathbf{x}$  of the deterministic system at the random parameter vector  $\mathbf{y}$  that is associated with the realization  $\mathbf{z}^{(i)}$  of the germ distribution.  $\mathbf{x}_k$  vectors ( $k = 1, \dots, N$ ) are the coefficients for the PCE of the state vector, which allows evaluation of the spectral coefficients (Le Maître & Knio 2010).

Like all sampling-based methods, the complexity of the parameter-space influences the amount of evaluations for the collocation points. For the application of Monte Carlo sampling, the Latin hypercube sampling has proven to be far more efficient for multidimensional applications (Helton & Davis 2003) and sparse grid sampling methods are frequently used in the application of the PCE (Xiong *et al.* 2009). Another approach was introduced by Blatman & Sudret (2011) with the least angle regression sampling, which strives to delete insignificant polynomials from the spectral expansion.

## NETWORK MODELS

In this article, classical propagation methods are applied to the water quality model in order to evaluate the influence of parameter uncertainties and compare them to the PCE. This is done in two case studies. The first one is the small, simplified graph of a strongly aggregated distribution network, as shown in Figure 1(a). It contains one loop and three elevated tanks. It has a total number of 21 free elements given by 12 pipes and 9 nodes. The model uses one domestic demand pattern for all nodes and the number of consumers per node is given in Table 1. The topology of the second distribution network is illustrated by the graph in Figure 1(b). In order to test the algorithms under practical conditions, this realistic network model has been provided by the project partner Veolia in the context of the ResiWater project. It is made up of 2,175 pipes, 1,822 nodes and one reservoir in the densely looped part of the network. In all, the model contains almost 4,000 graph elements. The model uses two demand patterns that have been defined and calibrated on operational data.

The general setup that is investigated in this paper is applied to both example networks. In each model, a diurnal

**Table 1** | Consumer distribution in the strongly aggregated looped network

Node	1	2	3	4	5	6	7	8	9
Consumers	72	108	84	172	6	51	180	83	116

demand pattern is defined for a selection of free nodes. This pattern contains a series of demand multipliers. For each hydraulic time step, the according demand factor is multiplied by the number of consumers that is defined individually for each node. Through this behavioural aggregation a large portion of the network is influenced by a concentrated number of variables.

The inherent aleatory uncertainty in the demand is modelled by a normal distribution with a mean of  $\mu = 1$  and standard deviation of  $\sigma = 0.3$ . This way the variation is applied on the deterministic demand estimate and the generation of negative demand values is highly unlikely. From this distribution, a sample of  $N = 1e5$  elements is taken to modify the parameters for the Monte Carlo simulation. From the ensemble of these simulations a reduced number of  $N = 1e4$  elements are taken to evaluate the PCE. The propagation uses an EPS over the course of 1 week.

## RESULTS

Figure 2 shows the mean water age at nodes 3 and 9 in the network, together with the 95% confidence intervals as a function of the simulated time. It can be seen that after roughly 1 day the result stabilizes and is not dependent on the initial conditions any more. In Figure 3, the estimate PDF is visualized for the same nodes after a simulation period of 58 hours with the histogram for the Monte Carlo approach, the dashed line for the PCE and the point dashed line for the FOSM.

Figure 3(a) gives the distribution at node 3 and Figure 3(b) the one for node 9. The values of the mean values and standard deviations are given in Table 2 along with the PCE order. In the evaluation of the results it has become apparent that for the small network no significant difference can be observed between the minimum, maximum and average water age. The behaviour at the two nodes is very different. The distribution at node 3, while clearly being non-linear, has still a single peak and a limited standard deviation. This observation is supported by the coefficients in Table 3 which decrease rapidly after the second coefficient. In contrast, the distribution at node 9 shows two peaks and the expansion coefficients decrease more slowly. This means that the probable water age in the loop node 3 is higher

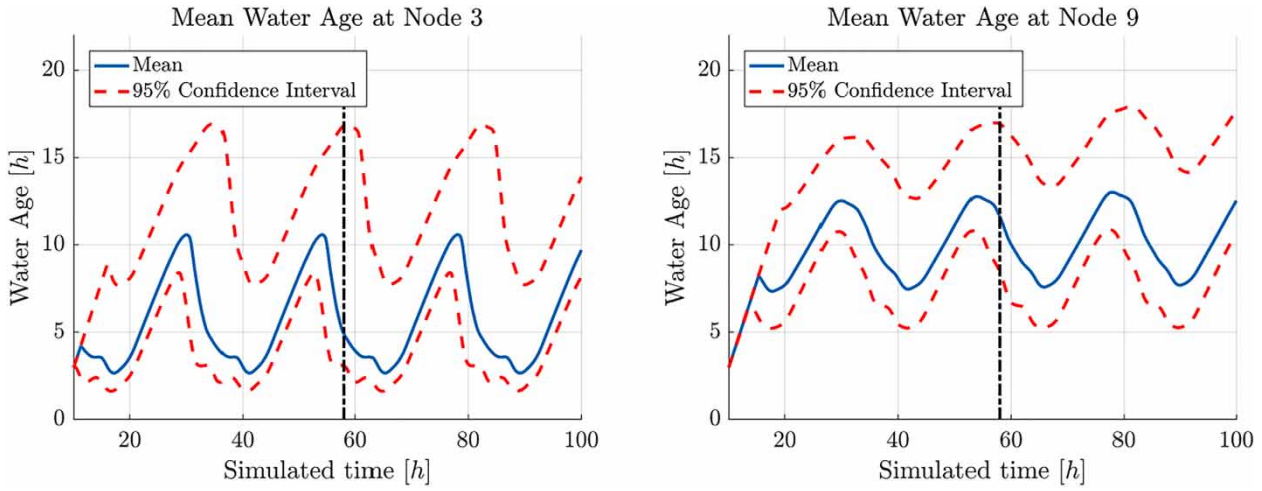


Figure 2 | Temporal development of mean water age at nodes 3 and 9 in the highly aggregated looped network with the 95% confidence interval.

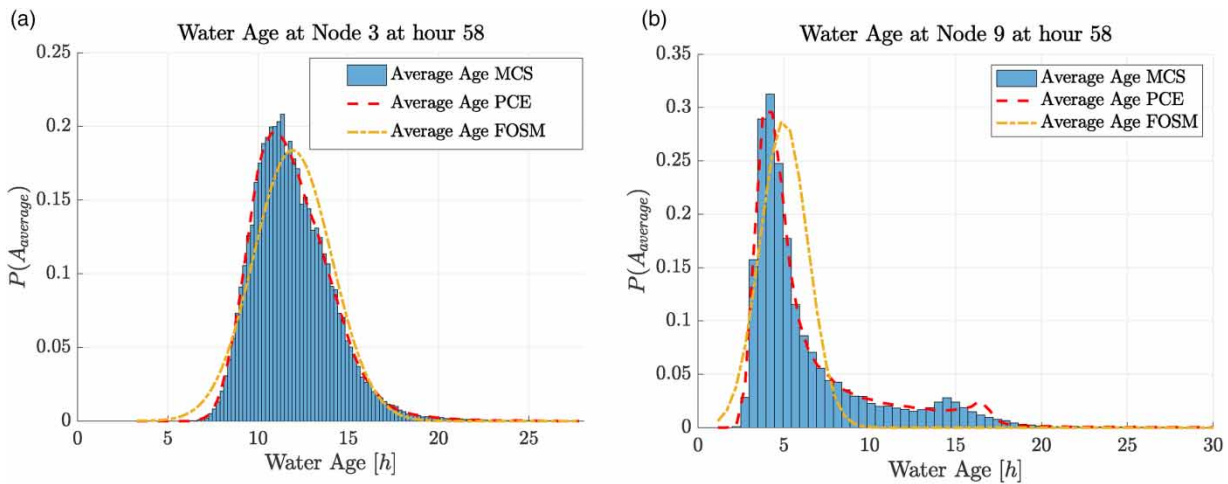


Figure 3 | Water age distributions for small network: (a) node 3 in the meshed part of the system; (b) at node 9 at a branch of the network.

Table 2 | Stochastic moments and PCE order for the probability distribution function of average water age at nodes 3 and 9 of the small network

		Avg water age $A_{avg}$ [h]
Node 3	Mean	$\mu_3 = 11.9$
	Standard deviation	$\sigma_3 = 4.7$
	Expansion order	$O_3 = 7$
Node 9	Mean	$\mu_9 = 6.5$
	Standard deviation	$\sigma_9 = 7.8$
	Expansion order	$O_9 = 9$

than the one in the tree node 9; however, due to the lower standard deviation the level of uncertainty is lower at node 3. The development of the two peaks in Figure 3(b) suggests that, depending on the demand multiplier, the flow structure

changes and water arrives from a different node, takes a longer route or flows more slowly. Remixing of old water is unlikely due to the structure of the network with three reservoirs that are connected to the loop. A detailed analysis of the origin reservoir has not been performed as a part of this study but may offer better insight. The results from the Monte Carlo approach and the PCE are in good agreement. For the sensitivity-based FOSM the limitation in capturing the non-linear behaviour becomes obvious. While it gives a good approximation in the mean and variance for node 3, it is unable to approximate the bimodal distribution in node 9.

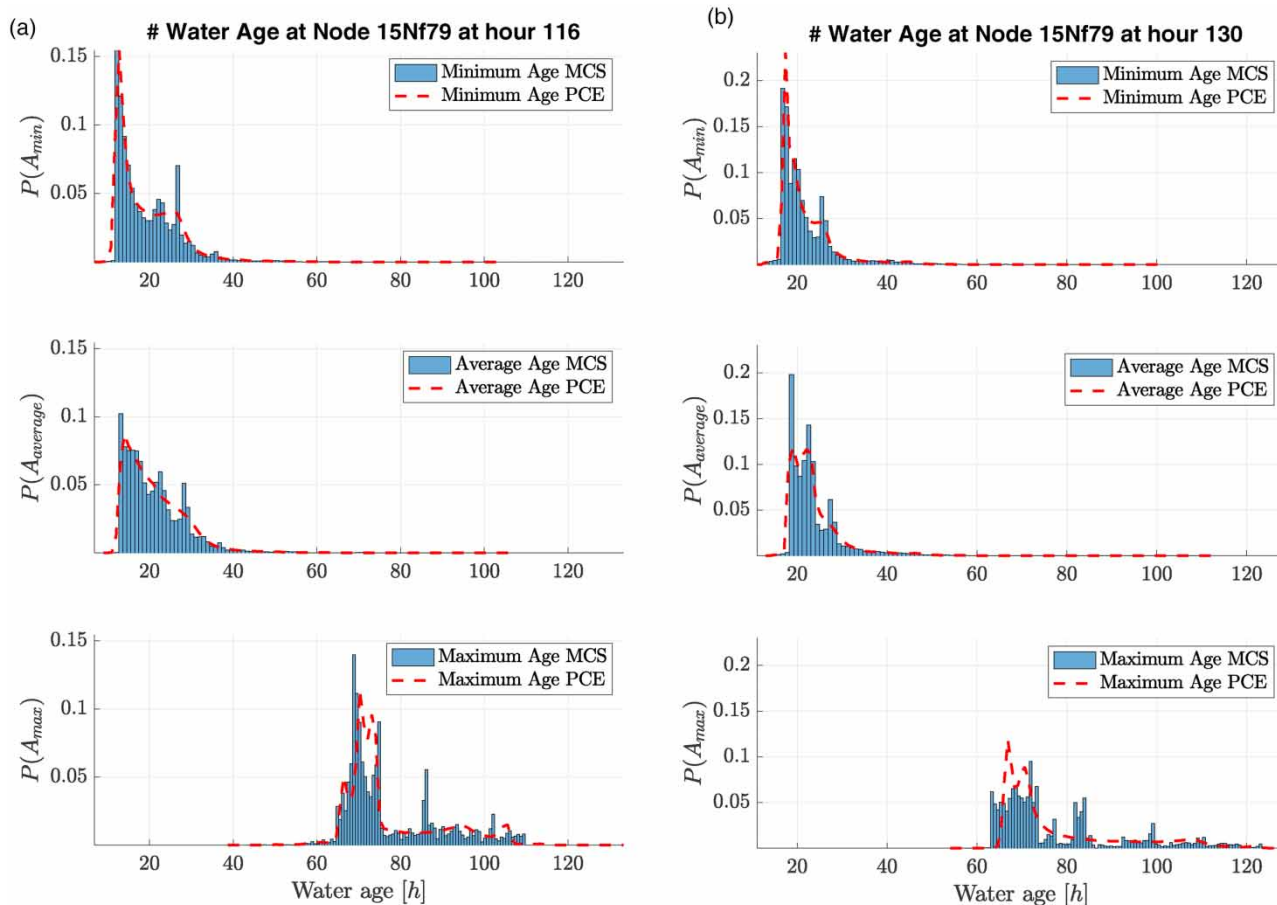
Figure 4 shows the probability distribution of the water age at node 15Nf79 of the medium-sized network which is indicated by a circle in Figure 1(b). Figure 4(a) displays the

**Table 3** | The expansion coefficients for the PCEs of average water age at nodes 3 and 9 of the small network

Order $N$	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$
Node 3	11.90	-2.12	0.40	0.011	0.05	0.09	0.03	0.04		
Node 9	6.52	-3.67	2.10	-1.52	-0.03	-1.70	0.00	-1.96	0.06	-0.590

distribution after a simulated duration of 116 hours whereas Figure 4(b) gives the distribution after 130 hours. For both times, the water quality has reached a periodicity. However, after a simulated time of 116 hours the water age is at the peak of the water age cycle while at 130 hours it reaches a valley. In contrast to the small network, there is a considerable difference in the three quality indicators of minimum, average and maximum water age and for this reason they are illustrated separately. The first two stochastic moments are given in Table 4 along with the expansion order  $N$ .

From the stochastic data a few trends can be observed. First, even though the simulation is periodic, it is not cyclic since the mean of all indicators rises with the simulation duration. Second, the distributions of the mean and average water age are very close in the stochastic moments as well as in the expansion coefficients up to the order  $N=5$ , which is shown in Table 5. From this it may be inferred that the average water age is dominated by the minimum water age and the amount of older water in the network is relatively low. Third, due to the fact that the mean and the variance of the maximum water age are considerably

**Figure 4** | 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.



**Table 4** | Stochastic moments for the probability distribution function and PCE order of minimum, average and maximum water age at node 15Nf79 for 116 and 130 hours of simulated time in the medium network

Node 15Nf79		Min water age $A_{min}$ [h]	Avg water age $A_{avg}$ [h]	Max water age $A_{max}$ [h]
116 hours	Mean	$\mu = 19.2$	$\mu = 21.1$	$\mu = 77.4$
	Standard deviation	$\sigma = 53.7$	$\sigma = 53.1$	$\sigma = 133.0$
	Expansion order	$O=20$	$O=17$	$O=20$
130 hours	Mean	$\mu = 21.6$	$\mu = 23.6$	$\mu = 77.5$
	Standard deviation	$\sigma = 36.3$	$\sigma = 36.7$	$\sigma = 178.3$
	Expansion order	$O=19$	$O=16$	$O=14$

**Table 5** | The first ten expansion coefficients for the PCEs of minimal, average and maximal water age at node 15Nf79 for 116 and 130 hours of simulated time in the medium network

Order $N$	116 hours			130 hours		
	$A_{min}$	$A_{avg}$	$A_{max}$	$A_{min}$	$A_{avg}$	$A_{max}$
$a_0$	19.15	21.11	77.41	21.61	23.60	77.51
$a_1$	-6.67	-6.38	-10.11	-5.12	-4.67	-11.76
$a_2$	2.81	3.25	3.63	2.86	3.47	5.57
$a_3$	-0.24	-0.02	0.00	-1.18	-1.18	0.43
$a_4$	0.30	0.63	-2.82	0.21	-0.04	-2.33
$a_5$	-0.86	-1.12	0.82	-0.30	-0.76	0.56
$a_6$	0.26	-0.13	0.93	0.09	0.25	0.86
$a_7$	0.24	-0.06	-0.82	0.10	0.36	-0.92
$a_8$	-0.14	0.13	-0.31	0.24	0.43	-0.19
$a_9$	-0.20	0.12	1.42	-0.31	-0.42	0.79

higher, it is likely that at least a portion of the water stays in the network by mixing with the new water at pipe junctions. As a result of the more complex structure of the medium-sized network it is not possible to identify changing peaks as an effect of the changing demand multiplier; however, a similar effect may be responsible for the skewed distributions in Figure 3. For the minimum and average water age the PCE is able to approximate the basic shape of the probability density function; however, small fluctuations were not resolved as a high expansion order would be needed in that case. A similar effect can be seen for the maximum water age. The basic shape of the PDF is well approximated by the PCE; however, the histogram from the Monte Carlo approach shows a bifurcation in the distribution with a second peak for a higher water age. These observations are backed by the development of the expansion coefficient in Table 5.

In contrast to the application in the strongly aggregated network, the coefficients have not yet converged to zero, implying that a higher order expansion is needed for a more accurate approximation.

Comparing the computational effort, it has to be taken into account that the PCE uses a tenth of the function evaluations used for the Monte Carlo simulation. In comparable applications it has been shown that the PCE converges far quicker in the mean and variance of the probability distribution.

## CONCLUSION

The PCE has been applied to a small and a realistic medium-sized water distribution system to evaluate the effect of demand uncertainties on the resident time of water in the network. It has been shown that the PCE gives comparable results to Monte Carlo simulations while reducing the number of evaluations of the full system by an order of magnitude. This means that for a fixed number of simulations the PCE performs considerably better. Even for the relatively small normally distributed perturbation of the demand used in this paper the probability distribution for the water age is far from normal. This means that perturbation methods are not suitable for this task. Further, it has been shown that the size of the network has a considerable influence on the maximum water age.

Additionally, uncertainty quantification was tested on three different kinds of quality indicators. For each of them probability density function may be multi-modal, which has an interest for the WDN management in terms of water safety. The effect of the different water ages is much more distinct in the application of the bigger network. The ensuing difference for the multimodal peaks often correspond to different water paths.

This presented work opens a new direction of research in water distribution networks. It is possible to extend it to applications in a higher dimensional parameter-space; however, the sampling base approach of the MC and non-intrusive PCE is limited to application with lower numbers of parameters, as the computational cost increases exponentially. This effect can, in part, be mitigated through the application of sparse-grid sampling methods or Latin hypercube sampling.

## ACKNOWLEDGEMENTS

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

- Blatman, G. & Sudret, B. 2011 Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics* **230** (6), 2345–2367.
- Cacuci, D. G., Ionescu-Bujor, M. & Navon, I. M. 2005 *Sensitivity and Uncertainty Analysis, Vol. II: Applications to Large-Scale Systems*, Vol. 2. CRC Press, Boca Raton, FL, USA.
- Cross, H. 1936 *Analysis of Flow in Networks of Conduits or Conductors*. University of Illinois at Urbana Champaign, College of Engineering. Engineering Experiment Station, IL, USA.
- Dandy, G., Mantopoulos, A., Qiu, M., Shizas, G. & Yao, Z. 2013 Advantages and limitations of the use of water age as a surrogate for water quality in water distribution systems. In: *EWRI 2013*, ASCE, Cincinnati, OH, USA.
- Der Kiureghian, A. & Ditlevsen, O. 2009 Aleatory or epistemic? Does it matter? *Structural Safety* **31** (2), 105–112.
- Di Cristo, C., Leopardi, A. & de Marinis, G. 2015 Assessing measurements uncertainty on trihalomethanes prediction through kinetic models in water supply systems. *Journal of Water Supply: Research and Technology-AQUA* **64** (5), 516–528.
- Fabrie, P., Gancel, G., Mortazavi, I. & Piller, O. 2010 Quality modeling of water distribution systems using sensitivity equations. *Journal of Hydraulic Engineering-ASCE* **136** (1), 34–44.
- Fu, G. & Kapelan, Z. 2011 Fuzzy probabilistic design of water distribution networks. *Water Resources Research* **47**, W05538. doi:10.1029/2010WR009739.
- Hart, D., Rodriguez, J. S., Burkhardt, J., Borchers, B., Laird, C., Murray, R., Klise, K. & Haxton, T. 2019 Quantifying hydraulic and water quality uncertainty to inform sampling of drinking water distribution systems. *Journal of Water Resources Planning and Management* **145** (1), 13.
- Helton, J. C. & Davis, F. J. 2003 Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety* **81** (1), 23–69.
- Jungnickel, D. 2005 *Graphs, Networks and Algorithms*. Springer, Berlin, Germany.
- Le Maître, O. & Knio, O. M. 2010 *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer Science & Business Media, Dordrecht, The Netherlands.
- Piller, O. & Tvard, L. 2014 Modeling the transport of physicochemical parameters for water network security. *Procedia Engineering* **70**, 1344–1352.
- Piller, O., Gilbert, D. & Van Zyl, J. E. 2010 Dual calibration for coupled flow and transport models of water distribution systems. In: ASCE (ed.), *Water Distribution Systems Analysis WDSA*, Tucson, AZ, pp. 722–731.
- Piller, O., Gilbert, D., Haddane, K. & Sabatie, S. 2011 Porteau: An object-oriented programming hydraulic toolkit for water distribution system analysis. In: *Eleventh International Conference on Computing and Control for the Water Industry (CCWI 2011)*, 5–7 September 2011, Exeter, UK.
- Raei, E., Shafiee, M. E., Nikoo, M. R. & Berglund, E. 2018 Placing an ensemble of pressure sensors for leak detection in water distribution networks under measurement uncertainty. *Journal of Hydroinformatics* **21** (2), 223–239.
- ResiWater 2019 ANR and BMBF Project, Innovative Secure Sensor Networks and Model-Based Assessment Tools for Increased Resilience of Water Infrastructures. <http://www.resiwater.eu/> (accessed 15 January 2019).
- Rossman, L. A. & Boulos, P. F. 1996 Numerical methods for modeling water quality in distribution systems: a comparison. *Journal of Water Resources Planning and Management* **122** (2), 137–146.
- Savic, D. A., Kapelan, Z. S. & Jonkergouw, P. M. 2009 Quo vadis water distribution model calibration? *Urban Water Journal* **6** (1), 3–22.
- Smith, R. C. 2013 *Uncertainty Quantification: Theory, Implementation, and Applications*, Vol. 12. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- Sumer, D. & Lansey, K. 2009 Effect of uncertainty on water distribution system model design decisions. *Journal of Water Resources Planning and Management* **135** (1), 38–47.
- Tsakiris, G. & Splilotis, M. 2017 Uncertainty in the analysis of urban water supply and distribution systems. *Journal of Hydroinformatics* **19** (6), 823–837.
- Xiong, F., Xiong, Y., Greene, S., Chen, W. & Yang, S. 2009 A new sparse grid based method for uncertainty propagation. In: *ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, San Diego, CA, USA. American Society of Mechanical Engineers, pp. 1205–1215.
- Xiu, D. 2010 *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, Princeton, NJ, USA.
- Xiu, D. & Karniadakis, G. E. 2002 The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing* **24** (2), 619–644.
- Xu, T. Y. & Qin, X. 2014 Integrating decision analysis with fuzzy programming application in urban water distribution system operation. *Journal of Water Resources Planning and Management* **140** (5), 638–648.

First received 15 January 2019; accepted in revised form 4 May 2019. Available online 3 June 2019