Modelling the impact of water temperature, pipe, and hydraulic conditions on water quality in water distribution networks

Hadi Mohammed* Hoese Michel Tornyeviadzi and Razak Seidu

Water and Environmental Engineering Group, Institute of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology (NTNU) in Ålesund, Larsgårdsvegen 2, 6009 Ålesund, Norway

*Corresponding author. E-mail: hadi.mohammed@ntnu.no

Abstract

Identifying and controlling the drivers of change in the quality of water within distribution systems requires a comprehensive understanding of the individual and interactive effects of relevant factors. This article examines the impact of water temperature, pipe characteristics, and hydraulic conditions on the microbiological, physical, and chemical parameters of water quality in the distribution network using Bayesian Dirichlet process mixture of linear models and random forest method. The study was based on a database of the distribution network for the city of Ålesund in Norway and records of water quality data measured at seven different locations in the network from 2013 to 2019. In both modelling approaches applied, temperature was identified as the main factor that controls the microbiological stability of water in the network. From the minimum to the maximum values of temperature in the pipes (3.35 °C–11.14 °C respectively), the probabilities of occurrence of bacteria in water increased from 0.36 to 0.95. Temperature was also shown to be an important factor that affects the chemical parameters of water quality (pH, alkalinity and electrical conductivity). Among the input parameters included in this study, concentration of residual chlorine was shown to have the strongest growth-inhibiting effect on Total Bacteria in the pipes. The results further showed that changes in the hydraulic conditions in the pipes (residence time and flow) were among the most important determinants of the physical, chemical and microbiological quality of water in the distribution network. The random forest models assigned minimal importance to the pipe characteristics and conditions on changes in the water quality parameters. However, the Bayesian models revealed that these parameters have significant impact on the quality of water in the pipes.

Key words: hydraulic conditions, pipe characteristics, temperature, Total Bacteria, water distribution network, water quality

Highlights

- Drivers of change in the quality of water in distribution pipes were evaluated.
- Stability of water in pipes is mainly controlled by temperature and water age.
- Residual chlorine has the strongest growth-inhibiting impact on bacteria in pipes.
- There is a higher chance of bacteria occurrence in metal than in plastic pipes.
- Age of pipes mainly affects turbidity, electrical conductivity and Total Bacteria in water.

INTRODUCTION

The production and distribution of microbiologically safe drinking water to the public is the fundamental objective of drinking water supply systems. Water that reaches the taps of consumers is expected to be of the same quality as when treated. However, whereas water is treated to ensure microbiological

safety, be aesthetically appealing and devoid of any chemicals that may alter the overall safety of water, the quality is not usually maintained after transmission through distribution networks.

Undesirable changes in the quality of treated water could occur during transmission in distribution networks often as a result of several factors. For instance, while disinfectant residuals are meant to inhibit the regrowth of bacteria in water distribution systems, the design of water distribution networks themselves could cause hydraulic instabilities, increased water age, and decay of residual chlorine, which may influence the proliferation of bacterial communities in water (Mays 2011; Mohamed & Ahmed 2011). As a result of sediment resuspension, scouring of biofilms from pipe walls, as well as changes in mixing conditions, fluctuations in hydraulic conditions in water distribution networks can influence not only the microbiological quality of water, but also important physical parameters such as turbidity and color (Aisopou et al. 2012). Further, studies have shown that the properties and conditions of the pipe materials are critical factors affecting failure in the pipes and associated water quality problems (Yu et al. 2010; Wang et al. 2012; Rezaei et al. 2015). For instance, pipe materials with rough inner surfaces have higher potential for regrowth of bacteria compared to pipes with smoother pipe materials (Chowdhury 2012). In addition, many old drinking water distribution systems are composed of metal pipes such as cast iron and steel, many of which are subject to corrosion. This can lead to the release of chemicals into water, affecting the quality integrity of the distribution system. Apart from corrosion in metal pipes, changes in the characteristics of treated water itself such as pH, alkalinity, hardness, temperature, and so on could affect the corrosivity of water in pipes (Salvato et al. 2003).

To date, research has investigated the occurrence of bacteria, mainly heterotrophic plate count (HPC) in water distribution pipes (LeChevallier 2003; Chen et al. 2013; Liu et al. 2014) and the effects of certain pipe materials, hydraulic conditions, and chlorine on the regrowth potential (Wang et al. 2012, 2014; Prest et al. 2016). Although maintaining the biological stability of water in distribution systems is the main priority for water supply systems, changes in chemical and/or physical parameters such as pH, total iron, calcium carbonate, electrical conductivity, turbidity, color, and the occurrence of corrosion in metal pipes may not only cause aesthetic or even health problems to consumers (Wang & Chen 2018), but also serve as important surrogates of the microbiological quality of water in pipes and could influence the regrowth of bacteria (Shi et al. 2018; Tchórzewska-Cieślak et al. 2019). Many of these parameters are routinely monitored in water distribution systems, yet the main causes of changes in these parameters are not widely investigated. Similarly, due to its diverse effect on the physical, chemical and microbiological quality of water, temperature is recognized as one of the most important controlling factors of water quality, including in distribution systems (Eck et al. 2016; Prest et al. 2016; Monteiro et al. 2017; Zlatanović et al. 2017). Water temperature can change during transmission due to exchange of heat between pipe materials and the surroundings. Yet, unlike at the raw water source and treatment stages, fluctuations in water temperature in distribution networks are not regularly monitored. According to a recent review, the link between water temperature and water quality issues in distribution networks is still being understood (Agudelo-Vera et al. 2020).

Given the complexities in the interaction of different factors on the overall stability of water in distribution pipes, a comprehensive understanding of the impact of various interacting factors is necessary for planning innovative control strategies that could ensure minimal changes to the quality of water reaching consumer taps. While many external factors including weather conditions and routine maintenance/rehabilitation works could affect the quality of water in pipe networks, it is evident from the literature that the internal conditions of the networks themselves are critical to ensuring the stability of water quality in the networks. Many researchers have studied the effects of different internal conditions in pipes on the quality of water in distribution systems mainly through laboratory-scale experiments (Codony et al. 2005; Lehtola et al. 2007; Hammes et al. 2008; Aisopou et al. 2012; Wasim et al. 2016; Zlatanović et al. 2017; Yoon et al. 2019; Cowle et al. 2020). Although

significant insights have been gained through these studies, the controlled conditions under which experiments are performed may be significantly different from real systems. Further, by relying on operational data, models have been applied in studying relations between controlling factors of instabilities in water distribution systems (Sadiq *et al.* 2008; Tabesh *et al.* 2009; Ramos-Martínez *et al.* 2014; Husband & Boxall 2015; Kutyłowska 2019). However, classical linear regression-based models that have been applied in many of these studies do not often consider the potentially stochastic nature of the dependence of water quality on pipe characteristics, operating conditions, and temperature. Owing to their probabilistic nature, Bayesian nonparametric (BNP) models can handle limitations in datasets and are more capable of revealing and interpreting highly complex relations that may not be identified by traditional statistical and/or machine learning models (Grzenda 2015; Wagenmakers *et al.* 2018). Rather than focusing only on the mean of a target parameter, BNP models draw inferences from how the distribution of the target parameter depends on predictors. Moreover, by enabling uncertainties to be accounted for and integrated in all stages of a model through Monte Carlo simulations with a given dataset, BNPs models could provide more credible interpretations of the impacts of different factors on a target parameter (Moe 2010).

The objective of this paper is to apply a BNP modelling framework to evaluate how temperature, characteristics and conditions of pipes, and hydraulic conditions, influence relevant parameters of water quality in distribution networks. The modelling framework was applied to a typical urban water distribution system to study how these input parameters could affect the distributions of relevant water quality parameters in the network. In addition, due to its ability to estimate predictor variable importance in a model, the random forest machine learning method was applied, and the results compared with those achieved with the BNP models. Rather than measuring water quality in distribution systems only to ensure regulatory requirements are met, data taken from these systems could be used to understand the behavior of the systems through data-driven modelling. This could provide vital information for ensuring the stability of water in distribution systems as well as improving customer perceptions and confidence in water supplies services.

MATERIALS AND METHODS

Data and source

The dataset used in this study was composed of seven-year (2013–2019) weekly records of water quality variables measured in the water distribution network of the city of Ålesund in Norway. The water quality and flow information in the pipes are measured using sensors and are used by the managers of the municipality's water and wastewater infrastructure for monitoring water quality at different zones in the network. The water quality parameters, which were used as targets in the models in this study were Total Bacteria, pH, turbidity, color, alkalinity, and electrical conductivity. To determine the microbial quality of water at each of the monitoring locations in the network, weekly water samples were analyzed for the presence of specific faecal indicator bacteria (FIB) such as *Escherichia coli* and coliform bacteria. However, no observation of these specific FIBs was made throughout the period for which data used in this study was collected. Therefore, the counts of Total Bacteria were used in this study. These parameters were measured at seven different locations in the network, as shown in Figure 1. The water treatment plant draws water from Brusdalsvatnet Lake, which is the only drinking water source for the city.

Using each of the six water quality parameters as a target, two sets of six different Bayesian regression models were constructed. In the first set of models, the input variables were pipe material type, length, diameter, age, water flow, water age, and water temperature. In the case of the model for Total Bacteria, residual chlorine was included as an additional input. In the second set of Bayesian regression models, interactions among the input parameters were used. This was used to evaluate how interactions among
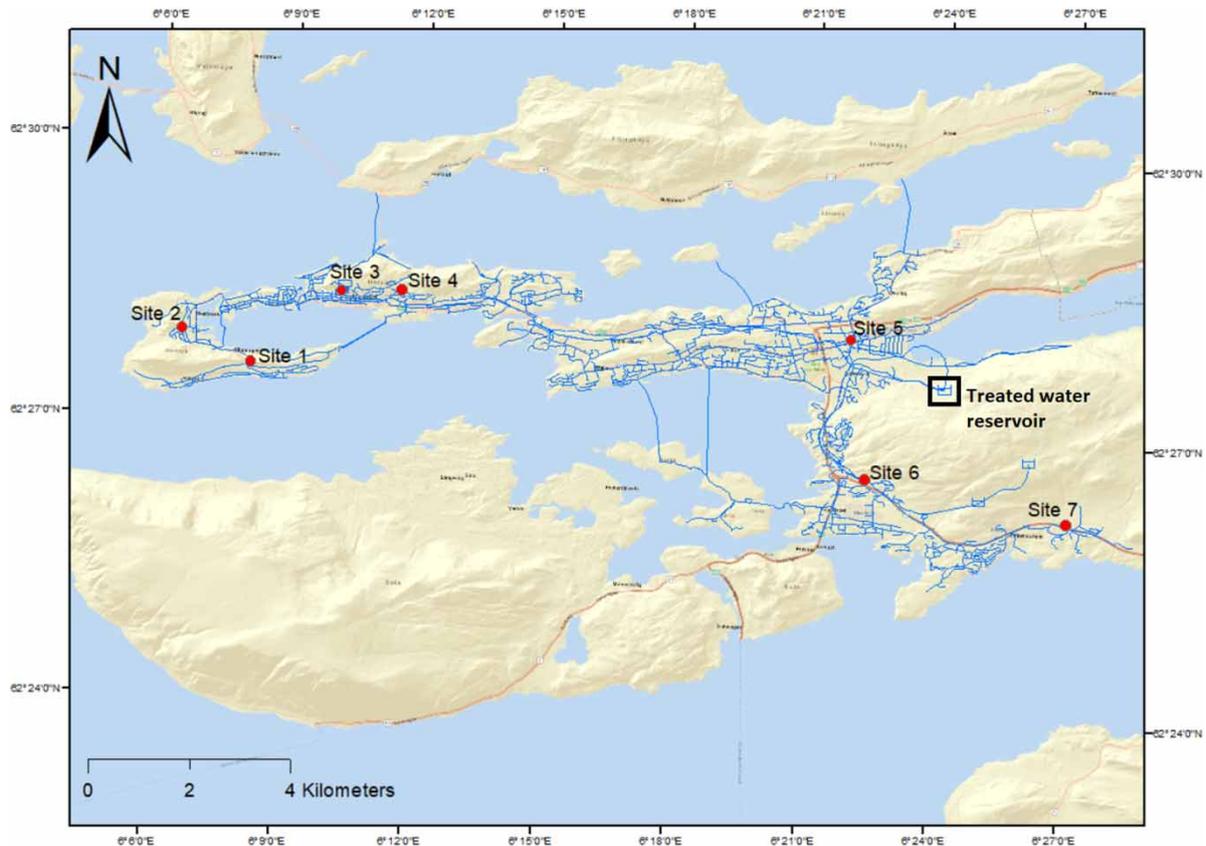
Figure 1 | The water distribution network for the municipality used in this study, showing the location of the treated water reservoir at the water treatment plant. The seven water quality measurement locations (S1 – S7) are shown in red spots.

the various predictors affect changes in the water quality parameters. Measurements of residual chlorine and temperature at each of the seven locations were not available at the time of this study; therefore, measurements taken from the treated water reservoir outlet for the same period as the other water quality parameters were used. While these may be different across the seven locations in the network, the measurements at the treated water reservoir outlet could be a useful representation of variations in the entire network. To match the water quality data taken from each measurement location with the information on the pipe characteristics within each measurement zone, the entire pipe network database was separated based on the polygons that define each zone using the selection by location tool in ArcGIS 10.6. For each zone, the length, diameter, and material type for individual pipes were recorded. Averages of the lengths and diameters of the pipes for each zone were then calculated to match the water quality data for each measurement location. The entire network is composed of 12 different pipe materials mainly, from cast iron of varying densities and Miscibility Gap Alloys (MCU) (45.4%); polyvinyl chloride (PVC) (17.3%); Polyethelyne (PE) and Polyethelyne low and high densities (PEL and PEH) (32.8%); Polypropylene (PPP) <1%, Low electrical resistivity pipes (LER) (<1%), Graphite pipes (Atomic Absorption Spectroscopy – AAS) (<1%); and BET (<1%). To include this information in the models, the two dominant material types for each zone were identified and assigned values from 1 to 3. The dominant pipe materials for each zone were identified based on the total length of each pipe material within each zone. Therefore, pipe materials with lengths greater than 2% of the total pipe length in the zone were used. This was used to match the water quality data at their respective locations. In the case of the hydraulic information (flow and water age) in the pipe networks, an existing EPANET model that has been calibrated for the network was used. Records of the water quality parameters were provided by the Water and Wastewater Department of the Ålesund municipality. It must be noted that owing to security issues, the authors of this study do not have permission to share the water quality data

from the distribution pipes used in this study. However, a summary of the datasets is presented in Table S1 in the supplementary materials.

Bayesian Dirichlet process mixture of linear models

The Dirichlet Process (DP) mixture of generalized linear model takes the form (Hannah *et al.* 2011; Fan & Bouguila 2012):

$$S_i \sim GLM(\cdot \,|X_i, \theta_i) \text{ for } i = 1, \ 2, \ 3, \ \ldots, n$$

$$\theta_i | G \sim G,$$

$$G | \alpha, G_0 \sim DP(\alpha, G_0)$$

$$N(\beta_0 + X_i^T \beta_j, \alpha_i^2)$$

$$G = \sum_{k=1}^{\infty} (w_k^* \delta_{\theta_k^*}) \sim DP(\alpha, G_0), \ with \ \sum_{k}^{\infty} (w_k^*) = 1$$

where S_i is the speed observation, *GLM* is a generalized linear model that is conditioned on the covariate vector (X_i), θ_i is the bundle of parameters over X_i and $Y|X_i$, G is the random distribution drawn from the DP, G_0 is the base distribution over the same space as G, and α is a positive scalar (concentration parameter), N is the Gaussian distribution, β_0 and β_j are regression parameters, α_i^2 is the variance of the predictors, $\#w_k^*$ is the mixing proportion, k is the number of mixture components, and $\delta_{\theta_k^*}$ is a Dirac delta function.

The model is implemented using a stick-breaking process (SBP), which involves repeated breaking of a unit length stick into disjoint categories (Fan & Bouguila 2013). The stick-breaking construction process is performed using the following conditions:

$$w_k^* = v_k \prod_{i=1}^{k-1} 1 - v_i$$

$$\theta_k^* \sim G_0$$

$$v_k \sim Beta(1, \alpha)$$

In this study, the DP mixture of linear regression model was built using the BayesRegression tool (Karabatsos 2015), which is a matlab-based tool that implements Bayesian nonparametric and parametric regression models. The characteristics of the water distribution pipes, namely, pipe material, length, diameter, and age; the hydraulic conditions in the pipes (water flow and water age), as well as water temperature were used as predictors for each of the six water quality parameters (Total Bacteria, pH, turbidity, color, alkalinity, and electrical conductivity). In the case of the Total Bacteria model, residual chlorine concentrations were included in the predictors.

To estimate the posterior distribution of the model parameters, Markov Chain Monte Carlo (MCMC) simulation was applied with 50,000 iterations. The convergence of the MCMC posterior analysis was assessed using trace plots and 95% Monte Carlo credible intervals (MCCI) for the model parameters. Posterior inferences from the MCMC results were made from 3,600 samples. The predictive fits of the models were further evaluated using the standardized residuals (z_i) of the response variables and the mean square predictive error criterion $(D(m))$ defined as follows:

$$z_i = \{y_i - E_n(Y|X_i, m)\}/\sqrt{Var_n(Y|X_i, m)}$$

where y_i is observed response, E is estimated response, and m is the number of models compared. Generally, a response value y_i is judged as an outlier when z_i is greater than 2.

$$D(m) = \sum_{i-1}^{n} \{y_i - E_n(Y|X_i, m)\}^2 + \sum_{i-1}^{n} Var_n(Y|X_i, m)$$

where the first term measures the goodness-of-fit to the data, and the second term is a penalty for model complexity. For a set of $m = 1, \ldots, M$ models compared, the model with the smallest value of $D(m)$ is selected as the one with the highest predictive accuracy.

Moreover, the proportion of variance in the response variables that were explained by the regression models were determined using the R-squared statistic (R^2), calculated using:

$$R^2 = 1 - \left(\frac{\sum_{i-1}^{n} \{y_i - E_n(Y|X_i)\}^2}{\sum_{i-1}^{n} (y_i - \bar{y})^2} \right)$$

Random forest models

The random forest (RF) introduced by Breiman (2001) is a machine learning algorithm that is made up of ensembles of decision trees. Each tree learns from a random sample of the training data points, which are drawn through bootstrapping. Generally, splitting of each node or tree is performed using random subsets of the input features in model training. Therefore, predictions are made by averaging the prediction from each tree in the forest with the aim of improving robustness and precision in the prediction. Typically, two thirds of the model training dataset are randomly sampled to formulate each decision tree, while the remaining set ('out-of-bag' samples) are used to test the accuracy of each decision tree. Due to its simplicity and diversity, the RF algorithm is widely used in both classification and regression problems. A major advantage of the RF algorithm is the ability to quantify the relative importance of each predictor on the prediction. The algorithm automatically computes this score after model training.

In this study, the RF classification algorithm was used to predict each of the six water quality parameters measured in the distribution pipes using the characteristics of the pipes, hydraulic conditions and water temperature as inputs. Before the model training, the time series data for each water quality parameter (target) was organized into four groups based on percentiles; $\leq$ 5th, $\leq$ 50th, $\leq$ 90th, and $\geq$ 90th percentile respectively. The entire dataset was randomly divided into two sections, 70% for model training and 30% for testing. For both training and testing sets, overall classification accuracies were evaluated together with their 95% confidence intervals. In addition, the Kohen's kappa metric, which measures the degree of agreement between the model predictions and actual observations for a classification model (Ben-David 2008) was evaluated. For each model, the relative importance of the predictor variables was used to assess their degrees of impact on each of the water quality variables predicted.

RESULTS AND DISCUSSION

Model accuracy measures

Convergence of the posterior estimates of the Dirichlet process mixture models were verified by examining the trace plots generated for each model. The results generally indicated good mixing and convergence of the MCMC samples. In addition, analysis of the 95% MCCI of the model parameters showed that the MCMC samples provided sufficiently small half-widths (ranged between

0.001–0.178 for Total Bacteria model, 0.01–0.213 for the pH model, 0.002–0.288 for color model, 0.008–0.201 for turbidity model, 0.001–0.323 for alkalinity model, and 0.024–0.138 for electrical conductivity model). This implied that the samples from the 50,000 MCMC iterations used in this study adequately converged to samples from the posterior distributions of the individual models. The goodness-of-fit statistics of the posterior estimates for each water quality parameter in the two sets of Bayesian models are presented in Table 1 (together with their 95% MCCI values). For each water quality parameter, the least D(m) statistic obtained from the MCMC iterations is shown in the table. Observation of the standardized residuals indicated that the residuals were all within – 2 and 2, indicating the absence of outliers. Further, the R^2 values show that between 51 and 94% of variance in the water quality parameters were explained by the first set of Bayesian regression models without parameter interactions. Marginal improvements in the model fit statistics were achieved in the second set of Bayesian models in which only the interaction terms were used as predictors. For instance, the R^2 values in these sets of models ranged from 0.57 in the model for pH to 0.96 in the model for Total Bacteria.

Table 1 | Posterior predictive model fit statistics of the Bayesian regression models for the various water quality parameters

Model	Posterior predictive SSE – D(m)		SSE fit to data – Gof(m)		Coefficient of determination – (R^2)	
	Value	**95% MCCI**	**Value**	**95% MCCI**	**Value**	**95% MCCI**
Model 1 (with individual predictor)						
Total Bacteria	28.32	19.02	8.86	1.25	0.94	0.08
pH	46.54	25.21	11.39	3.40	0.51	0.27
Turbidity	14.64	12.47	12.45	3.07	0.59	0.32
Color	45.33	26.52	7.23	2.21	0.68	0.03
Alkalinity	32.81	21.31	6.83	1.90	0.63	0.13
Electrical conductivity	24.38	16.07	5.76	1.11	0.78	0.24
Model 2 (with only interactions)						
Total Bacteria	18.41	12.34	6.37	1.14	0.96	0. 32
pH	38.48	23.21	9.83	3.15	0.57	0.23
Turbidity	11.62	8.56	7.96	2.47	0.63	0.28
Color	39.52	22.71	6.34	1.87	0.72	0.16
Alkalinity	27.93	18.44	5.03	1.21	0.68	0.22
Electrical conductivity	21.48	14.53	4.14	1.06	0.83	0.17

SSE = Sum of Squares due to Error.

The performance indices of the random forest models are shown in Table 2. In the testing datasets for the various water quality parameters, the classification accuracy ranged between 0.63 in the Total

Table 2 | Performance indices of the random forest models for each of the water qaulity parameters

Model	Accuracy (95% CI)		Kappa	
	Training (95% CI)	**Testing (95% CI)**	**Training**	**Testing**
Total Bacteria	0.75 (0.03)	0.63 (0.02)	0.77	0.61
pH	0.83 (0.02)	0.77 (0.13)	0.62	0.49
Turbidity	0.96 (0.03)	0.74 (0.22)	0.94	0.71
Color	0.83 (0.11)	0.65 (0.04)	0.88	0.68
Alkalinity	0.87 (0.04)	0.63 (0.30)	0.71	0.64
Electrical conductivity	0.77 (0.12)	0.67 (0.11)	0.66	0.55

Bacteria model and 0.77 in the pH model, indicating an averagely good performance. The performances of the random forest models were also evident from the Kappa statistics, which ranged between 0.49 in the model for pH and 0.71 in the model for turbidity. This statistic essentially measures how close the classification results matched the actual datasets and ranges from –1 (total disagreement) to 1 (total agreement).

Predictor significance: Bayesian regression models

Significance based on marginal posterior distributions

In typical Bayesian regression modelling, a predictor variable can be interpreted as significant when its 50% posterior credible interval excludes zero. This interval is bounded by the 25 and 75% of the posterior estimates. The 50% posterior intervals for the various predictors of the water quality parameters are shown in Figure 2.
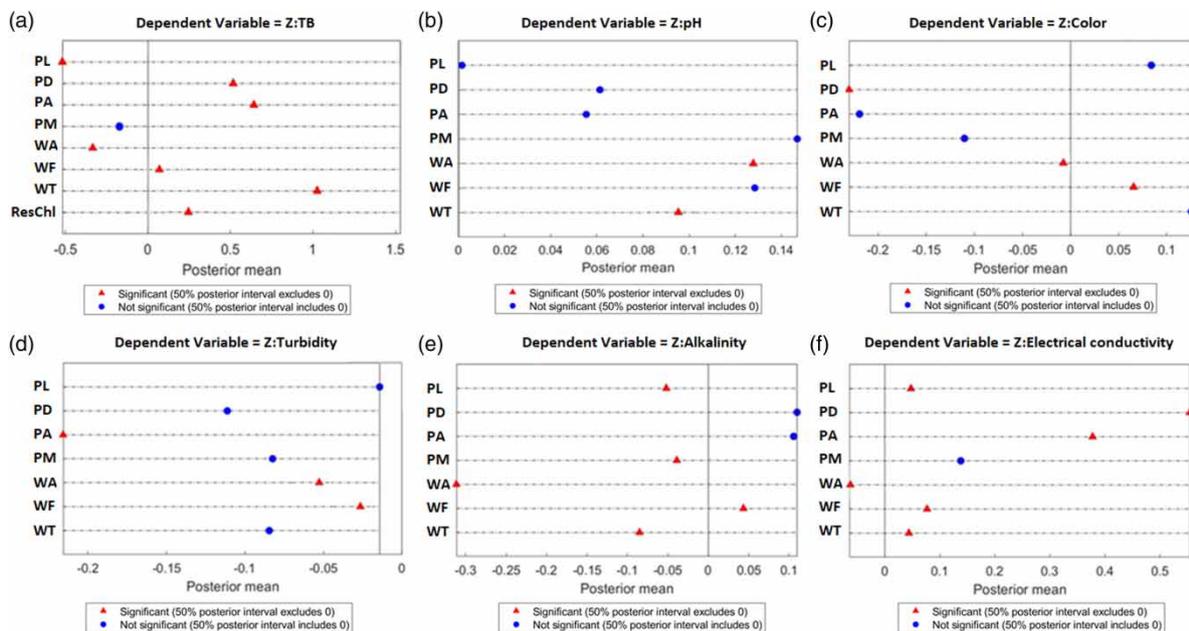

Figure 2 | Love plots of marginal posterior distributions models showing the posterior means and significance of the predictors for each water quality parameter. PL = pipe length, PD = pipe diameter, PA = pipe age, PM = pipe material, WA = water age, WF = water flow, WT = water temperature, ResChl = residual chlorine.

The results of the first group of models without interactions among the predictor variables (Figure 2) suggest that temperature, residual chlorine, water age, water flow, pipe age, diameter and length are significant predictors for Total Bacteria in the water distribution pipes. It must be noted that the values of temperature and residual chlorine were not directly taken from the seven measurement locations. Nonetheless, as temperature is one of the most critical factors that directly or indirectly control the microbiological quality of water (Ndiongue *et al.* 2005; Prest *et al.* 2016), maintaining disinfectant residual in water distribution pipes is a primary disinfection goal in many water distribution systems since it inhibits the regrowth of some microbial organisms. Studies have shown that the decay of residual disinfectant can be affected by several factors, including the type of pipe materials and age, water age, biofilm characteristics, constituents of treated water, as well as temperature (Lehtola *et al.* 2005; Berry *et al.* 2006; Al-Jasser 2007; Fisher *et al.* 2011; Chowdhury 2012; Zhang & Andrews 2012; Eck *et al.* 2016). In the case of pH, only temperature and water age were shown to be important predictors. Flow, water age, and pipe diameter were shown to be important predictors of

color while turbidity of water was mainly associated with changes in flow conditions as well as ages of water and the pipes. Except for pipe diameter and age (in the alkalinity model) and pipe material (in the model for electrical conductivity), all the input variables were shown to be significant predictors of these water quality parameters.

The results of the second set of Bayesian regression models generally revealed that the effects of pipe characteristics, hydraulic conditions, and water temperature on changes in the quality of water in distribution systems could be better understood by evaluating the effects of their interactions rather than considering each parameter in isolation. For instance, in the model for Total Bacteria (Figure 3(a)), many of the interactions were shown to be significant predictors, except for the interactions of pipe diameter with almost all the other parameters as well as pipe length with pipe age, pipe material with water age, and flow with temperature. Although parameters such as pipe characteristics and flow were individually shown to be insignificant predictors of pH (Figure 2(b)), interactions of these parameters were shown to be significant in the second set of models. Similar interactive effects were noted in the models for the other water quality parameters. It can also be observed in Figure 3(c) and 3(d) that most of the interactions that were found to be significant predictors of color were also shown to be significant for turbidity. Except for the effects of pipe diameter and age, similar parameter significance was found for these two water quality variables in the model without interactions (Figure 2). This demonstrates the link between these water quality variables. Similar interactive effects of the predictor variables could also be observed in the models for alkalinity and electrical conductivity in Figure 2.
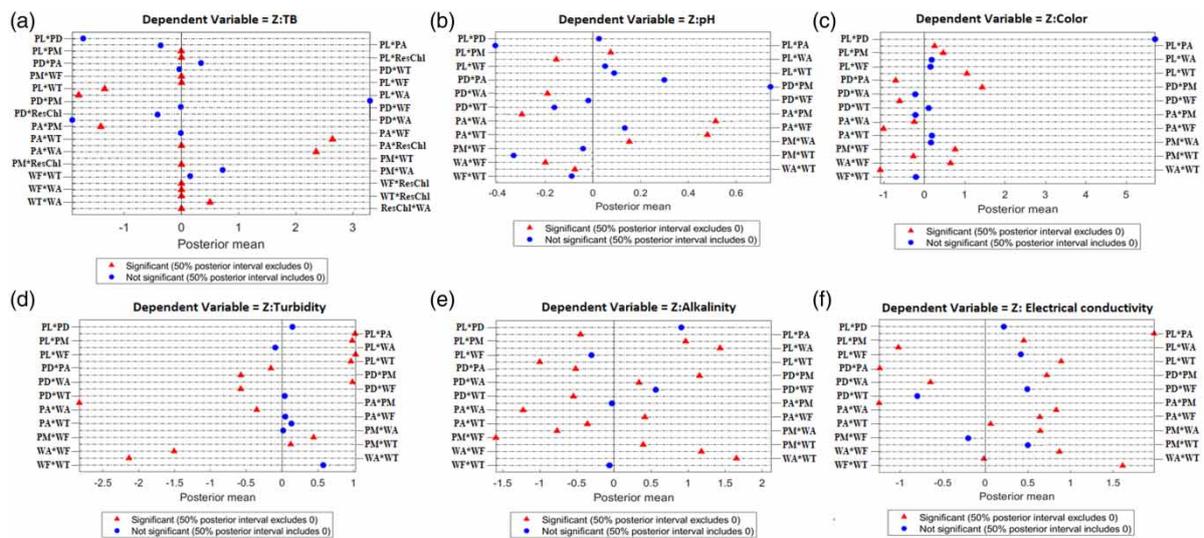

Figure 3 | Love plots of marginal posterior distributions models showing the posterior means and significance of the interactions among the predictors for the six water quality parameters.

Conditional probabilities of posterior predictions

A summary of the conditional probabilities used to evaluate the impacts of the various predictors on the water quality parameters is further presented in Table 3. The probabilities that the value of a water quality parameter being greater than zero at minimum and at maximum values of each predictor and conditional upon mean values of all other predictors were estimated. The conditional probabilities of the posterior predictions were calculated only for the first set of Bayesian models (without interactions). The results presented in Table 3 are based on 3,600 Monte Carlo samples drawn from a total of 50,000 samples that were generated in each model. In the table, values shown in grey color represent increasing probabilities from minimum values of the predictors to their maximums.

Table 3 | Posterior predictive summary statistics showing the conditional probabilities

Water quality parameter (Y)	P(Y > 0) (2.5%, 97.5%) MCCI – at minimum parameter value			
	Pipe length	Pipe diameter	Pipe age	Pipe material[a]
pH	0.52 (0.01, 0.91)	0.48 (0.09, 0.93)	0.52 (0.01, 0.99)	0.41 (0.01, 0.91)
	0.51 (0.01, 0.96)	0.53 (0.01, 0.97)	0.50 (0.01, 0.99)	0.56 (0.01, 0.99)
Turbidity	0.52 (0.04, 0.98)	0.57 (0.09, 0.98)	0.43 (0.05, 0.91)	0.57 (0.02, 0.98)
	0.57 (0.03, 0.93)	0.52 (0.02, 0.91)	0.62 (0.02, 0.98)	0.52 (0.08, 0.91)
Color	0.46 (0, 0.97)	0.62 (0.21, 0.99)	0.41 (0, 0.78)	0.53 (0.04, 0.95)
	0.53 (0.04, 0.98)	0.38 (0.01, 0.82)	0.58 (0.17, 0.97)	0.44 (0.07, 0.84)
Alkalinity	0.69 (0.28, 0.87)	0.29 (0.05, 0.74)	0.29 (0.16, 0.68)	0.45 (0.37, 0.68)
	0.23 (0.10, 0.85)	0.59 (0.07, 0.82)	0.58 (0.19, 0.75)	0.33 (0.13, 0.67)
Electrical conductivity	0.43 (0.12, 0.98)	0.14 (0, 0.51)	0.32 (0.04, 0.85)	0.37 (0.13, 0.94)
	0.52 (0.06, 0.99)	0.78 (0.46, 1.00)	0.76 (0.44, 1.00)	0.58 (0.20, 1.00)
Total Bacteria	0.86 (0.36, 1.00)	0.67 (0.15, 1.00)	0.68 (0.22, 0.90)	0.82 (0.38, 1.00)
	0.66 (0.14, 0.90)	0.87 (0.37, 1.00)	0.85 (0.35, 1.00)	0.70 (0.18, 0.90)
	Water Flow	Water Age	Water Temperature	Residual Chlorine
pH	0.51 (0.01, 0.99)	0.45 (0.01, 0.94)	0.48 (0, 0.91)	
	0.59 (0, 1.00)	0.55 (0.01, 0.98)	0.56 (0.01, 1.00)	
Turbidity	0.57 (0.05, 0.93)	0.46 (0.06, 0.95)	0.55 (0.26, 0.94)	
	0.46 (0, 0.95)	0.57 (0.01, 0.91)	0.45 (0.01, 0.85)	
Color	0.42 (0.09, 0.87)	0.41 (0.03, 0.95)	0.42 (0.02, 0.93)	
	0.56 (0.01, 1.00)	0.52 (0.27, 0.73)	0.59 (0.01, 1.00)	
Alkalinity	0.34 (0.28, 0.69)	0.63 (0.36, 0.98)	0.42 (0.13, 0.81)	
	0.67 (0.15, 0.92)	0.31 (0.19, 0.50)	0.18 (0.01, 0.90)	
Electrical conductivity	0.46 (0.21, 0.94)	0.48 (0.01, 1.00)	0.44 (0.06, 0.98)	
	0.69 (0.02, 0.99)	0.51 (0.13, 0.83)	0.63 (0.04, 0.99)	
Total Bacteria	0.81 (0.40, 1.00)	0.77 (0.36, 0.93)	0.36 (0.08, 0.97)	0.80 (0.38, 1.00)
	0.70 (0.12, 0.94)	0.83 (0.25, 1.00)	0.95 (0.41, 1.00)	0.50 (0.08, 0.73)

Probability increases with increasing value of parameter are shown in grey color.
[a]Grouped based on two dominant pipe materials at the location. 1 = (PVC, PE), 2 = (PVC, BET), and 3 = (PVC, MCU).

The characteristics of the pipes, hydraulic conditions and water temperature, which were used as inputs in the models, showed varying effects on the water quality parameters. Although the increases were marginal, the results indicate that the probability that the pH of water in the pipes increases were dependent on the type of pipe material and increases in diameter. For instance, at minimum and maximum pipe diameters respectively, the probabilities of increase in the pH were estimated as 0.48 (0.09, 0.93) and 0.53 (0.01, 0.97). In the case of the pipe materials, the model indicates that water pH is more likely to increase in areas with metal pipes (such as MCU) than in plastic pipes (PE). It is worth noting that PVC pipes were dominant in all the three groups used as inputs in this study; therefore, the main differences in the material groups were based on either MCU, BET and PE. In addition, the longer the water age, the higher the likelihood of increases in the water pH. Similar effects were predicted by the model for flow and water temperature.

The posterior predictive summary for the model also indicates that older and wider pipes may be more prone to higher levels of alkalinity and electrical conductivity, the effects being larger on the latter. Old metal pipe materials may be more prone to the release of ions into water as a result of corrosion, potentially affecting the levels of alkalinity and electrical conductivity. While the size of water distribution pipes does not have direct relations with the ionic concentrations in water, wider pipes could provide larger surface area where corrosion may occur, thus increasing the probability of release of chemicals into water. In addition, low pressure conditions in wider pipes may reduce flow velocity, increase residence time, and potentially higher ion concentrations. Moreover, the probability of increases in electrical conductivity in the pipes was predicted to be higher in areas with metal pipes

(including MCU) than those with plastics. In addition, both the hydraulic conditions and water temperature were positively associated with this water quality parameter. The pipe characteristics that were predicted to increase the probability of increases in the turbidity and color of water were the length of the pipe sections and their ages. While disparities in pipe materials and increases in their diameter were shown to have marginal effect on changes in water turbidity, these changes were predicted to cause considerable reductions in color of water. In relation to the hydraulic conditions, the results indicate that higher turbidity and color are more likely to occur in areas with longer water age. Increases in color were also found to be affected by increases in flow and water temperature.

With respect to Total Bacteria in the distribution network, the results indicate that metal pipe materials may be associated with higher counts than plastic pipes. In areas with metal pipes, the probability of occurrence of positive counts was predicted as 0.82, whereas a value of 0.7 was predicted for areas with plastic pipes. Longer pipes, higher flow rates, and higher concentrations of residual chlorine in the pipes were predicted to reduce the likelihood of the occurrence of Total Bacteria in water. However, increases in the counts of the bacteria were more likely to occur in larger and older pipes as well as longer residence time (water age) and at high water temperature. A graphical overview of the impacts of the various input parameters on the counts of Total Bacteria in the water distribution pipes is further shown in Figure 3. Among the input parameters that showed positive relations with the bacteria counts, temperature showed the greatest impact. This is supported by the results of the marginal posterior distributions presented in Section 3.2.1 of this study. At the minimum and maximum values of temperature in the pipes, the probabilities of occurrence of bacteria in water were 0.36 and 0.95 respectively. Similarly, increases in the concentration of residual chlorine in the pipes were shown to have the strongest negative effect on the occurrence of bacteria.

Predictor significance: random forest models

A summary of the impacts of the pipe characteristics, hydraulic conditions, and temperature on the various water quality parameters, as predicted by the random forest model, is presented in Figure 4. In this figure, the various predictors in the models are ranked in terms of the magnitude of impact in the prediction of each water quality parameter.

Like the results of the Bayesian regression model, temperature was ranked as the most influential parameter in the prediction of changes in Total Bacteria in the distribution pipes. In addition, the hydraulic conditions in the pipes as well as the concentration of residual chlorine similarly showed strong impact on the occurrence of the bacteria. Generally, temperature is known to be a major determinant of water quality due to its influence on physical, chemical and biological processes. In relation to the impact on the microbial quality of water in distribution systems, several studies have highlighted the strength of its influence on their regrowth and survival (Bagh *et al.* 2004; Francisque *et al.* 2009; Prest *et al.* 2016; Monteiro *et al.* 2017). Similarly, water age and disinfectant concentrations in pipes, which can both be affected by flow conditions, have been shown to be strong factors in shaping bacterial populations in water distribution pipes (Williams *et al.* 2005; Gomez-Alvarez *et al.* 2012; Wang *et al.* 2014). Therefore, since it may be difficult to control the temperature and water flow in the distribution network, ensuring low residence time and effective residual chlorine concentrations in the pipes are highly important in controlling the proliferation of bacteria in distribution systems.

Although the results obtained in the Bayesian regression showed a higher chance of bacteria occurring in areas with metal pipes than in plastic ones, the random forest model suggested that the type of pipe material and its age have little impact on variations in the bacteria counts. This may be due to the few categories of pipe materials used as inputs in training the model. Unlike the Bayesian regression, in which probabilistic inferences were drawn using Monte Carlo simulations with the dataset, the random forest model assigns each observation of Total Bacteria in the dataset to its corresponding material type (either 1, 2 or 3 in this study). This probably made it difficult for the model to establish relationships
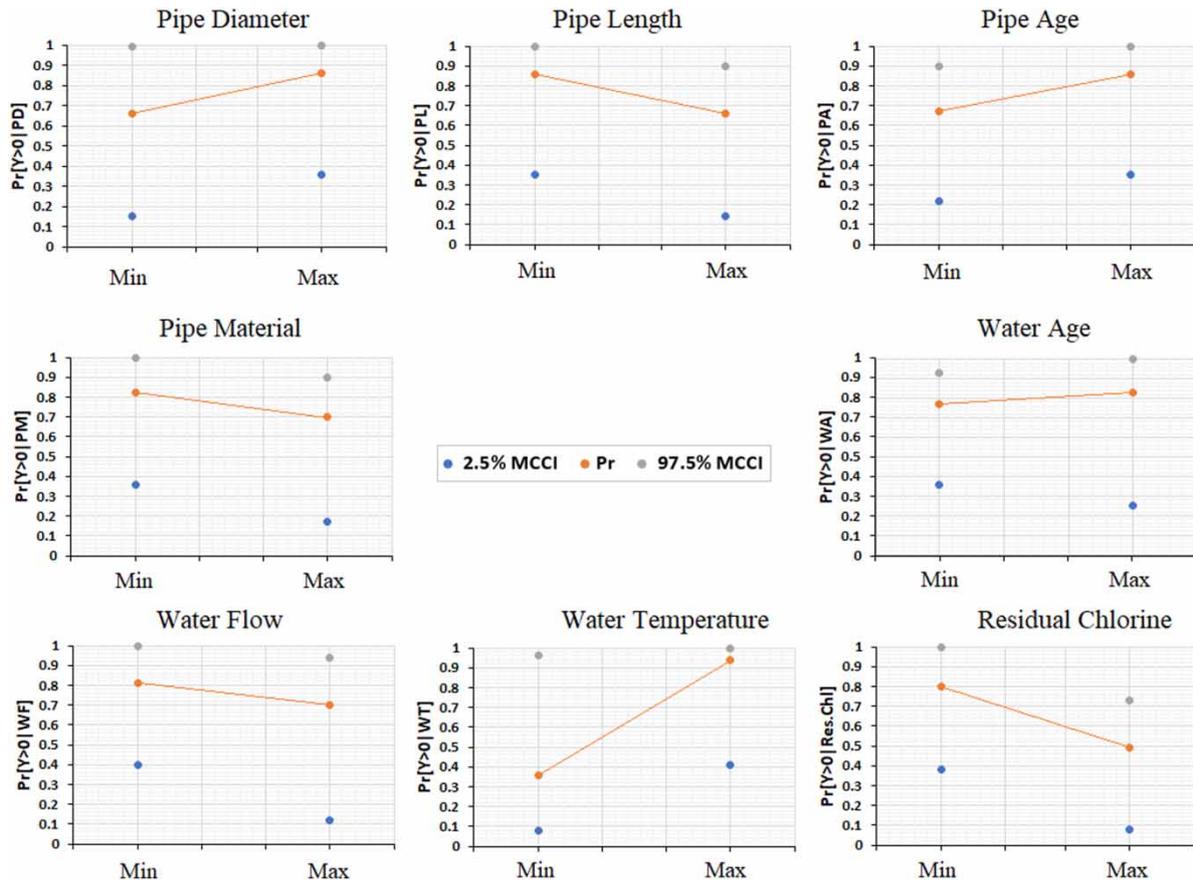
Figure 4 | Posterior predictive summary estimates for the Total Bacteria model showing the probalility of observing counts of bacteria in water as a function of each predictor variable.

between the types of pipe materials and the water quality parameters. Moreover, the Bayesian regression computes credible intervals for the posterior predictions, whereas the random forest model provides point estimates without probability-based uncertainty intervals. Not only can the type of pipe material affect the formation of biofilms in the distribution network, but also the proliferation and diversity of microbial populations (Yu *et al.* 2010; Chowdhury 2012; Wang *et al.* 2012). Results of a study that measured bacterial biomass in different pipe materials showed that bacterial densities in metal pipes (mainly iron) were between 10 and 45 times higher than in plastic-based materials such as PVC and PE (Niquette *et al.* 2000). Similarly, heterotopic plate count (HPC) in biofilms inside a model water distribution system were reported to be about 100 times higher on steel pipes than other pipe materials including PVC (Jang *et al.* 2011). While the results achieved in this study agree with those of these studies, diverse relationships between the type of pipe material and bacterial populations have also been previously reported in the literature. For instance, copper ions in copper pipes have been reported to decrease bacterial concentrations in water (Lehtola *et al.* 2004). Moreover, higher bacterial concentrations have previously been observed in plastic pipes such as PEX (cross-linked polyethylene) and PVC compared to copper and stainless-steel pipes (Schwartz *et al.* 1998; Van der Kooij *et al.* 2005). Similarly, Zacheus *et al.* (2000) reported no significant differences in bacterial concentrations with different pipe materials including stainless steel, PE, and PVC. This suggests that other site-specific conditions such as the quality of water could influence the relationships that are established between pipe materials and changes in the microbial quality of water in pipes.

For the remaining water quality parameters, the results suggest that the hydraulic conditions in the pipes as well as temperature are more important determinants of their variations. Increased retention times in water distribution systems can cause problems such as sediment deposition, ineffective control

of corrosion, and regrowth of microbial organisms (American Water Works Association 2002), which could all affect the physical, chemical, and biological quality of water in the network. In addition, sudden changes in hydraulic conditions in pipes could enhance the mobilization of accumulated materials at the walls of pipes, leading to increases in turbidity and color (Boxall & Saul 2005). After water treatment, the presence of turbidity and color may be due to internal contamination from release of ions into bulk water. For instance, changes in color of water in distribution networks could occur due to the release of iron in iron-based pipes (Vreeburg & Boxall 2007), implying that the type of pipe material in the network could influence water discoloration and associated chemical parameters. Moreover, studies have shown that changes in the concentrations of iron in water could induce changes in pH, alkalinity, and electrical conductivity (Bergel et al. 2013; Jachimowski 2017). However, the results of the random forest models suggest that pipe material has the least impact on the physical-chemical parameters of water quality included in this study (Figure 5). Similar to the impact on the microbial quality of water predicted by the random forest model in this study, this could be due to the limitations of the pipe material dataset, as previously explained.
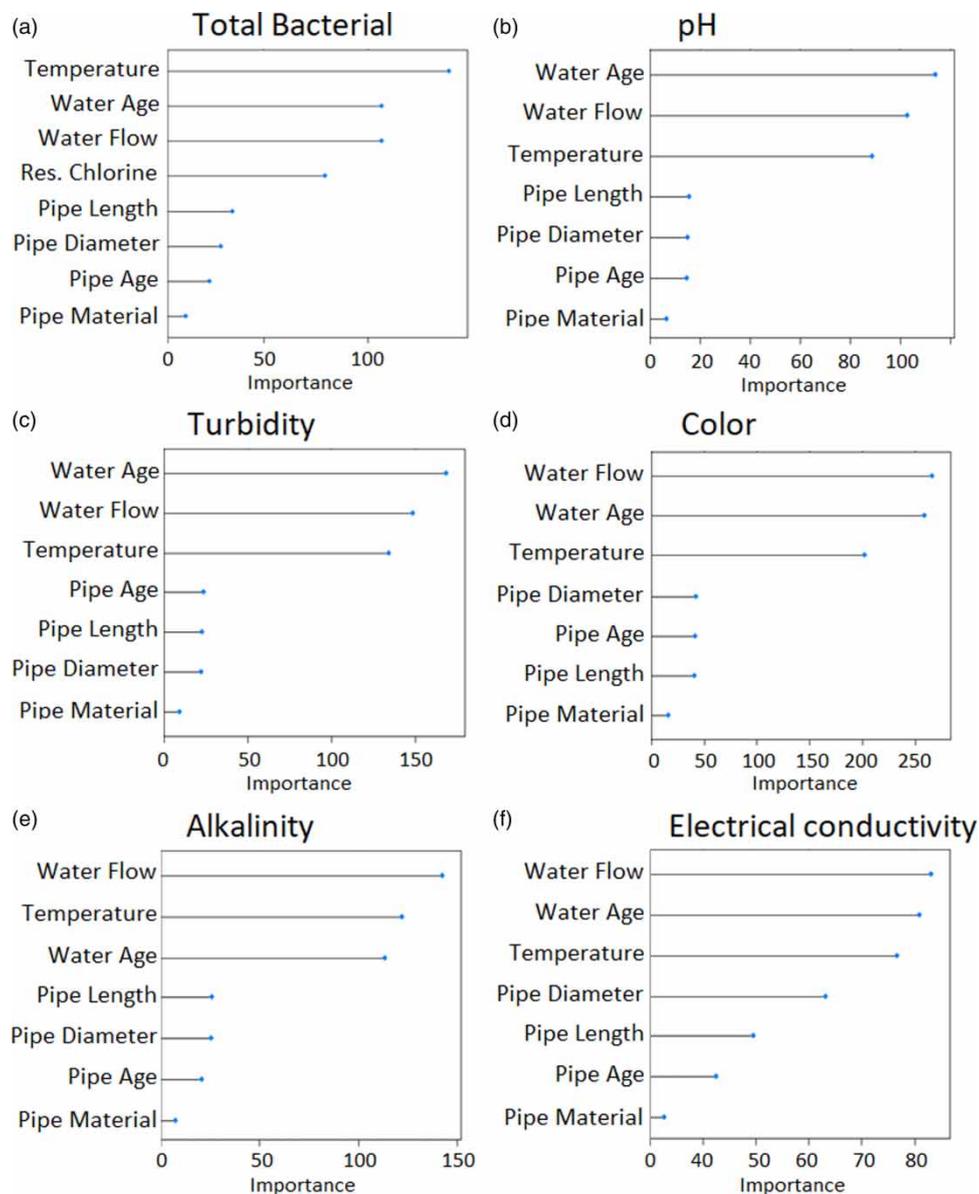

Figure 5 | Summary of the random forest model showing the relative importance of pipe characteristics, hydraulic conditions, and temperature on the variations in the water quality parameters.

Owing to complex physical, chemical, and biological processes that occur in typical water distribution systems, the quality of treated water may change before reaching consumers' taps. Several internal and external factors could influence the stability of treated water in distribution systems, resulting in violations in the standards of water quality and public health risks. As with many other components of water supply systems, different types of data are routinely collected in water distribution systems. The main contribution of this paper is to evaluate the degree of effects of relevant factors on water quality at different locations in a typical water distribution network. By relying on such datasets, the methodology applied in this study could be used to identify factors of greatest relevance that can be prioritized in order to minimize changes in the quality of water in the network and protect public health. Although this study focused primarily on the internal factors, the modelling framework can provide reliable information necessary for maintaining the physical, hydraulic, and water quality integrity of distribution systems when different types of information about distribution systems are available. Moreover, the methodology can be used to assess the impacts of different variables on different parameters of water quality at other stages of water supply systems such as raw water sources.

CONCLUSIONS

This paper applies a Bayesian nonparametric modelling framework and the random forest classification algorithm to evaluate the impacts of temperature, pipe characteristics and conditions, as well as hydraulic conditions on the quality of water in a typical urban distribution network. The study was based on weekly records of six water quality parameters measured across the city of Ålesund in Norway from 2013 to 2019.

In both modelling approaches, water temperature was identified as a major factor that controls the microbiological quality of water in the distribution network studied. From the minimum to the maximum values of temperature in the pipes (3.35 °C and 11.14 °C respectively), the probabilities of occurrence of bacteria in water increased from 0.36 to 0.95. Temperature was also shown to be an important factor that affects the chemical parameters of water quality (pH, alkalinity and electrical conductivity). The Bayesian regression model for Total Bacteria also confirmed the important role of residual chlorine in the control of the microbial quality of water in the distribution pipes, as it was shown to have the strongest negative influence on the bacteria levels. Similarly, the results from both modelling approaches showed that changes in the hydraulic conditions in the pipes (residence time and flow) are among the most important determinants of the physical, chemical and microbiological quality of water in the distribution network. Although the results of the random forest models indicated potentially minimal effect of pipe characteristics and conditions on the water quality parameters, the Bayesian regression revealed that these parameters are improtant factors. The results of this study suggests that metal pipe materials may be more associated with the occurrence of microbial organisms than plastic pipes. Different accounts on the association between metal pipe materials and bacteria have been previously reported in the literature. Although the results of this study agree with some of these accounts, the proportion of metal pipes relative to plastic pipes analysed in this work could be a contributory factor. Moreover, compared to plastic pipes, metal pipes have other advantages such as strength, durability, and temperature tolerance.

Information about which factors and to what extent they control the quality of water in distribution networks is essential for maintaining the safety of water supplied to the public. While it may be difficult to control some important factors such as temperature, improving the hydraulic conditions in a way that reduces residence times is key to ensuring the safety of water in distribution networks. In addition, the replacement of some metal pipe materials such as in water distribution networks with

plastic pipes may not only improve the biological stability of treated water, but also minimize other processes that could directly or indirectly alter the physical and chemical parameters of water quality.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available readers should contact the corresponding author for details.

REFERENCES

Agudelo-Vera, C., Avvedimento, S., Boxall, J., Creaco, E., de Kater, H., Di Nardo, A. & Jacimovic, N. 2020 Drinking water temperature around the globe: understanding, policies, challenges and opportunities. *Water* **12**(4), 1049.

Aisopou, A., Stoianov, I. & Graham, N. J. 2012 In-pipe water quality monitoring in water supply systems under steady and unsteady state flow conditions: a quantitative assessment. *Water Research* **46**(1), 235–246.

Al-Jasser, A. O. 2007 Chlorine decay in drinking-water transmission and distribution systems: pipe service age effect. *Water Research* **41**(2), 387–396.

American Water Works Association 2002 *Effects of Water age on Distribution System Water Quality*. American Water Works Association, Denver, CO, USA, p. 19.

Bagh, L. K., Albrechtsen, H. J., Arvin, E. & Ovesen, K. 2004 Distribution of bacteria in a domestic hot water system in a Danish apartment building. *Water Research* **38**(1), 225–235.

Ben-David, A. 2008 Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Systems with Applications* **34**, 825–832.

Bergel, T., Kaczor, G. & Bugajski, P. 2013 Technical conditions of water supply networks in small waterworks of the Małopolska and Podkarpackie Voivodeships. *Infrastructure and Ecology of Rural Areas* **3**, 291–304.

Berry, D., Xi, C. & Raskin, L. 2006 Microbial ecology of drinking water distribution systems. *Current Opinion in Biotechnology* **17**(3), 297–302.

Boxall, J. B. & Saul, A. J. 2005 Modelling discolouration in potable water distribution systems. *Journal of Environmental Engineering* **131**(5), 716–725.

Breiman, L. 2001 Random forests. *Mach Learn* **45**, 5–32.

Chen, L., Jia, R. B. & Li, L. 2013 Bacterial community of iron tubercles from a drinking water distribution system and its occurrence in stagnant tap water. *Environmental Science: Processes & Impacts* **15**(7), 1332–1340.

Chowdhury, S. 2012 Heterotrophic bacteria in drinking water distribution system: a review. *Environmental Monitoring and Assessment* **184**(10), 6087–6137.

Codony, F., Morato, J. & Mas, J. 2005 Role of discontinuous chlorination on microbial production by drinking water biofilms. *Water Research* **39**, 1896–1906.

Cowle, M. W., Webster, G., Babatunde, A. O., Bockelmann-Evans, B. N. & Weightman, A. J. 2020 Impact of flow hydrodynamics and pipe material properties on biofilm development within drinking water systems. *Environmental Technology* **41**(28), 3732–3744.

Eck, B. J., Saito, H. & McKenna, S. A. 2016 Temperature dynamics and water quality in distribution systems. *IBM Journal of Research and Development* **60**(5/6), 7–1.

Fan, W. & Bouguila, N. 2012 Variational learning of Dirichlet process mixtures of generalized Dirichlet distributions and its applications. In *International Conference on Advanced Data Mining and Applications*. Springer, Berlin, Heidelberg, pp. 199–213.

Fan, W. & Bouguila, N. 2013 Variational learning of a dirichlet process of generalized dirichlet distributions for simultaneous clustering and feature selection. *Pattern Recognition* **46**, 2754–2769.

402

Fisher, I., Kastl, G. & Sathasivan, A. 2011 Evaluation of suitable chlorine bulk-decay models for water distribution systems. *Water Research* **45**(16), 4896–4908.

Francisque, A., Rodriguez, M. J., Miranda-Moreno, L. F., Sadiq, R. & Proulx, F. 2009 Modeling of heterotrophic bacteria counts in a water distribution system. *Water Research* **43**(4), 1075–1087.

Gomez-Alvarez, V., Revetta, R. P. & Domingo, J. W. S. 2012 Metagenomic analyses of drinking water receiving different disinfection treatments. *Applied and Environmental Microbiology* **78**(17), 6095–6102.

Grzenda, W. 2015 The advantages of Bayesian methods over classical methods in the context of credible intervals. *Information Systems in Management* **4**, 53–63.

Hammes, F. A., Berney, M., Wang, Y., Vital, M., Koster, O. & Egli, T. 2008 Flowcytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Research* **42**, 269–277.

Hannah, L. A., Blei, D. M. & Powell, W. B. 2011 Dirichlet process mixtures of generalized linear models. *Journal of Machine Learning Research* **12**(Jun), 1923–1953.

Husband, S. & Boxall, J. 2015 Predictive water quality modelling and resilience flow conditioning to manage discolouration risk in operational trunk mains. *Journal of Water Supply: Research and Technology – AQUA* **64**(5), 529–542.

Jachimowski, A. 2017 Factors affecting water quality in a water supply network. *Journal of Ecological Engineering* **18**(4), 519–529.

Jang, H. J., Choi, Y. J. & Ka, J. O. 2011 Effects of diverse water pipe materials on bacterial communities and water quality in the annular reactor. *Journal of Microbiology and Biotechnology* **21**(2), 115–123.

Karabatsos, G. 2015 Bayesian Nonparametric Regression for Educational Research. In 2015 Annual Meeting Professional Development Course, Swissotel, Event Center Second Level, Vevey (Vol. 1).

Kutyłowska, M. 2019 Forecasting failure rate of water pipes. *Water Supply* **19**(1), 264–273.

LeChevallier, M. W. 2003 Conditions favouring coliform and HPC bacterial growth in drinking water and on water contact surfaces. In: *Heterotrophic Plate Counts and Drinking-Water Safety* (Bartram, J., Cotruvo, J., Exner, M., Fricker, C. & Glasmacher, A., eds). World Health Organization, Geneva, Switzerland, pp. 177–199.

Lehtola, M. J., Miettinen, I. T., Keinänen, M. M., Kekki, T. K., Laine, O., Hirvonen, A., Vartiainen, T. & Martikainen, P. J. 2004 Microbiology, chemistry and biofilm development in a pilot drinking water distribution system with copper and plastic pipes. *Water Research* **38**(17), 3769–3779.

Lehtola, M. J., Miettinen, I. T., Lampola, T., Hirvonen, A., Vartiainen, T. & Martikainen, P. J. 2005 Pipeline materials modify the effectiveness of disinfectants in drinking water distribution systems. *Water Research* **39**(10), 1962–1971.

Lehtola, M. J., Miettinen, I. T., Hirvonen, A., Vartiainen, T. & Martikainen, P. J. 2007 Estimates of microbial quality and concentration of copper in distributed drinking water are highly dependent on sampling strategy. *International Journal of Hygiene and Environmental Health* **210**, 725–732.

Liu, G., Bakker, G. L., Li, S., Vreeburg, J. H. G., Verberk, J. Q. J. C., Medema, G. J. & Van Dijk, J. C. 2014 Pyrosequencing reveals bacterial communities in unchlorinated drinking water distribution system: an integral study of bulk water, suspended solids, loose deposits, and pipe wall biofilm. *Environmental Science & Technology* **48**(10), 5467–5476.

Mays, L. 2011 *Water Transmission and Distribution*. American Water Works Association, Washington, DC.

Moe, S. J. 2010 Bayesian models in assessment and management. In *Environmental Risk Assessment and Management From A Landscape Perspective*. pp. 97–120.

Mohamed, H. I. & Ahmed, S. S. 2011 Effect of simplifying the water supply pipe networks on water quality simulation. *International Conference for Water. Energy Environment*, 26–28 March 2011, Sharjah, UAE, 41–46.

Monteiro, L., Figueiredo, D., Covas, D. & Menaia, J. 2017 Integrating water temperature in chlorine decay modelling: a case study. *Urban Water J.* **2017**(14), 1097–1101.

Ndiongue, S., Huck, P. M. & Slawson, R. M. 2005 Effects of temperature and biodegradable organic matter on control of biofilms by free chlorine in a model drinking water distribution system. *Water Research* **39**(2005), 953–964.

Niquette, P., Servais, P. & Savoir, R. 2000 Impacts of pipe materials on densities of fixed bacterial biomass in a drinking water distribution system. *Water Research* **34**(6), 1952–1956.

Prest, E. I., Hammes, F., van Loosdrecht, M. C. M. & Vrouwenvelder, J. S. 2016 Biological stability of drinking water: controlling factors, methods, and challenges. *Frontiers in Microbiology* **2016**(7), 45.

Ramos-Martínez, E., Gutiérrez-Pérez, J. J., Herrera, M., Izquierdo, J. & Pérez-García, R. 2014 Pipe database analysis transduction to assess the spatial vulnerability to biofilm development in drinking water distribution systems. In: *Mathematical Modeling in Social Sciences and Engineering* (J. C. Cortés Lopéz, L. A. J. Sánchez & R. J. V. Micó, eds), Nova Science Publishers, Hauppage, NY, 71–80.

Rezaei, H., Ryan, B. & Stoianov, I. 2015 Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. *Procedia Engineering* **119**(1), 253–262.

Sadiq, R., Saint-Martin, E. & Kleiner, Y. 2008 Predicting risk of water quality failures in distribution networks under uncertainties using fault-tree analysis. *Urban Water Journal* **5**(4), 287–304.

Salvato, J. A., Nemerow, N. L. & Agardy, F. J. 2003 *Environmental Engineering*. John Wiley & Sons, Hoboken, NJ.

Schwartz, T., Hoffmann, S. & Obst, U. 1998 Formation and bacterial composition of young, natural biofilms obtained from public bank-filtered drinking water systems. *Water Research* **32**(9), 2787–2797.

Shi, B., Wan, Y., Yu, Y., Gu, J. & Wang, G. 2018 Evaluating the chemical stability in drinking water distribution system by corrosivity and precipitation potential. *Water Science and Technology: Water Supply* **18**(2), 383–390.

Tabesh, M., Soltani, J., Farmani, R. & Savic, D. 2009 Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling. *Journal of Hydroinformatics* **11**(1), 1–17.

Tchórzewska-Cieślak, B., Pietrucha-Urbanik, K. & Papciak, D. 2019 An approach to estimating water quality changes in water distribution systems using fault tree analysis. *Resources* **8**(4), 162.

Van der Kooij, D., Veenendaal, H. R. & Scheffer, W. J. 2005 Biofilm formation and multiplication of legionella in a model warm water system with pipes of copper, stainless steel and cross-linked polyethylene. *Water Research* **39**(13), 2789–2798.

Vreeburg, I. J. & Boxall, J. 2007 Discolouration in potable water distribution systems: a review. *Water Research* **41**(3), 519–529.

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J. & Meerhoff, F. 2018 Bayesian inference for psychology. Part II: example applications with JASP. *Psychonomic Bulletin & Review* **25**(1), 58–76.

Wang, Y. H. & Chen, K. C. 2018 Removal of disinfection by-products from contaminated water using a synthetic goethite catalyst via catalytic ozonation and a biofiltration system. *International Journal of Environmental Research and Public Health* **2018**(11), 9325.

Wang, H., Hu, X. & Hu, C. 2012 Effects of chlorine and pipe material on biofilm development and structure in a reclaimed water distribution system. *Water Science and Technology: Water Supply* **12**(3), 362–371.

Wang, H., Masters, S., Edwards, M. A., Falkinham III., J. O. & Pruden, A. 2014 Effect of disinfectant, water age, and pipe materials on bacterial and eukaryotic community structure in drinking water biofilm. *Environmental Science & Technology* **48**(3), 1426–1435.

Wasim, M., Li, C. Q., Robert, D. J., Mahmoodian, M. & Setunge, S. 2016 Experimental investigation of factors influencing external corrosion of buried pipes. In *Proceedings of the 4th International Conference on Sustainability Construction Materials and Technologies (SCMT'16)*.

Williams, M. M., Domingo, J. W. S. & Meckes, M. C. 2005 Population diversity in model potable water biofilms receiving chlorine or chloramine residual. *Biofouling* **21**(5–6), 279–288.

Yoon, S., Kim, S. S. & Park, N. S. 2019 Experimental investigation of water quality responses to inorganic pollutants for the establishment of a contamination warning system. *Water Supply* **19**(7), 1965–1977.

Yu, J., Kim, D. & Lee, T. 2010 Microbial diversity in biofilms on water distribution pipes of different materials. *Water Science and Technology* **61**(1), 163–171.

Zacheus, O. M., Iivanainen, E. K., Nissinen, T. K., Lehtola, M. J. & Martikainen, P. J. 2000 Bacterial biofilm formation on polyvinyl chloride, polyethylene and stainless steel exposed to ozonated water. *Water Research* **34**(1), 63–70.

Zhang, H. & Andrews, S. A. 2012 Catalysis of copper corrosion products on chlorine decay and HAA formation in simulated distribution systems. *Water Research* **46**(8), 2665–2673.

Zlatanović, L., Van Der Hoek, J. P. & Vreeburg, J. H. G. 2017 An experimental study on the influence of water stagnation and temperature change on water quality in a full-scale domestic drinking water system. *Water Research* **123**, 761–772.