An approach to predict the failure of water mains under climatic variations

Zainab Almheiri[1], Mohamed Meguid[2] and Tarek Zayed[3]

[1] PhD Candidate, Dept. of Civil Engineering and Applied Mechanics, McGill University, 817 Sherbrooke Street West, Montréal, QC H3A 0C3, Canada. E-mail: zainab.almheiri@mail.mcgill.ca.
ORCID: 0000-0002-5346-279X

[2] Professor and Chair, Dept. of Civil Engineering and Applied Mechanics, McGill University, 817 Sherbrooke Street West, Montréal, QC H3A 0C3, Canada. E-mail: mohamed.meguid@mcgill.ca.
ORCID: 0000-0002-5559-194X

[3] Professor, Dept. of Building and Real Estate (BRE), The Hong Kong Polytechnic University, M1504, Li Ka Shing Tower. E-mail: tarek.zayed@polyu.edu.hk.

ORCID: 0000-0003-3249-7712

Abstract: Urban water distribution systems are critical infrastructures, and their failure can lead to significant economic, environmental, and social losses including flood streets and loss of treated drinking water. Identifying the failure patterns of water mains over time under various conditions is an inexpensive approach for estimating the structural deterioration of water distribution systems. It is also an alternative method for direct inspection that requires intensive efforts and budget. Time-dependent factors such as temperature and precipitation variations can lead to changes in frost depths and ground movements resulting in stresses that exceed design values and increasing the potential of water main failures. A few studies have addressed the impact of climatic variations on the failure prediction of water mains. To fill this gap, a temporal approach for the failure prediction of water mains under climatic variations is presented. The proposed approach can predict the failure of water mains at selected locations (not only one location) and allow to not only predict the failure by a one time-step ahead but also obtain accurate failure predictions up to nine months ahead. Another purpose of the proposed model is to accommodate additional variables to predict the failure of water mains at selected locations. To achieve this objective, a vector autoregression model with exogenous variables that incorporates the impact of climatic variations was developed. Spatiotemporal data of water mains failure events and climate data are collected for this study from Quebec and Ontario, Canada. Monte Carlo method was applied to validate the reliability of the predictive model. In other words, the failure prediction of water mains uncertainties were generated using Monte Carlo simulation. Results show that climatic variations can provide valuable information for the failure prediction of water mains. Results also prove that the proposed model can accurately predict the temporal failure patterns of water mains at two water distribution systems simultaneously.

***Keywords*:** Climatic variations, multivariate time series, failure prediction of water mains, spatiotemporal data, Monte Carlo simulation

1. Introduction

Water mains are critical infrastructures (CIs) and more than 300 failures of water utilities (100-300 mm) could occur in a year and thus require urgent repair or replacement in North America [1]. Failure of water mains can cause long-term failure of water distribution systems, thus losing a huge amount of drinking water [2]. Over two trillion gallons of treated drinking water is wasted owing to an estimated 240,000 annual water main breaks in the United States [3]. According to the Water Infrastructure Report Card published by the American Society of Civil Engineering (ASCE) in 2017, the water systems such as water mains in the US have attained an overall grade of "D." This means that the water mains infrastructures are in poor to fair condition and a large portion of the infrastructure at risk of failure [3] . Similarly, the Canadian Infrastructure Report Card revealed that 23% of water infrastructures are in poor to fair condition [4] . Additionally, $1 trillion is needed for the maintenance of water infrastructure to meet the future demand for water over the next 25 years [3].

Climatic variations may cause damage to critical infrastructures such as water mains [5]. Temperature variations can lead to changes in frost depths and ground movements, resulting in stresses that exceed design values and increasing the potential of failure [6]. The failure of water mains may result from dynamic factors and static factors. Dynamic factors are varied over time, including environmental and operational factors such as soil moisture, temperature, landslide, external loads, and water pressure. Whereas, static factors are fixed over time, such as pipe diameter, pipe age, and pipe material [2]. Environmental factors, including temperature and climate change can lead to the deterioration and failure of water mains. Damage to buried infrastructures such as water mains may result from climate events (e.g., flood) and climate variations in temperature and rainfall patterns [7].

Affecting of Climate Variability on Water main Failures

As pointed out by Vreeburg et al. [8], the number of breaks of water mains that made of cast iron increases during the winter season, which suggests the association between temperature and failure occurrences. On the one hand, Bruaset and Saegrov [9] found that the failure rate of water mains generally increases with the decrease in temperature. On the other hand, Makar [10] concluded that frost loading is an essential factor in evaluating the failure of water mains. Frost heaves can impose significant strains on pipe walls and thus lead to distresses and possibly failure. Soil freezing and thawing cycles due to seasonal changes in temperature can lead to water leakage and eventually failure [9]. The association between the failure of water mains and influential factors such as pipe age, pipe diameter, water pressure, soil corrosivity, external loads, and temperature have been studied by previous studies [11-13]. In particular, results showed that the drop in temperature enhances the number of water main breaks. In other words, temperature fluctuation tend to increase the number of water main breaks in winter. Goodchild et al. [14] developed a linear model to predict the water main breaks as a function of environmental and climate variables such as solar radiation, minimum grass temperature, daily rainfall, soil moisture deficit, evapotranspiration, water content on the topsoil, water wind, and actual evaporation. Results proved that minimum grass temperature, daily rainfall, soil moisture deficit, and actual evaporation can significantly contributed to the prediction of pipe breaks for cast iron and asbestos cement pipe materials buried in clay and loam. In another study, Rajani et el. [11] developed a non-homogenous Poisson model to study the impact of temperature variations on water mains breaks that made of cast iron (CI), ductile iron (DI), and galvanized steel. They found that temperature variations may impact the failure of water mains. One of their study limitations is that the developed model addressed only one climate variable that may contribute to the failure of water mains. Other studies [15, 16] conducted a statistical analysis to prove the association climate variations (e.g., temperature and precipitation) and pipe failure in cold regions. Results proved that the freezing index is one of the most significant climate covariates that may contribute to the failure of water mains. Kabir et al. [17] found that water mains made of metallic material are at high risk of failure in summer and winter seasons. Kleiner and Rajani [18, 19] studied three climate variables to predict the annual breaks of water mains including cumulative annual rain deficit (RD_c), annual freezing index (FI), and snapshot rain deficit (RD_s). FI is a surrogate measure for winter severity in a given year, RD_c is a surrogate measure for average annual soil moisture, and RD_s is a surrogate measure for locked-in winter soil moisture. They applied deterministic and probabilistic models to predict the annual breaks of water mains using a time step of one year. Additionally, the developed models are able to predict the failure of water mains under the defined climate variables. Gould et al. [20] found that soil shrinkage in nonfreezing regions that result from lower soil moisture enhances the pipe failure rate. Statistical methods [5, 21] performed to study the impact of climate variations on the integrity of water distribution systems.

Failure Prediction Model of Water Mains

Numerous studies have addressed the failure prediction of water mains based on static and dynamic factors, such as pipe diameter, materials, length , pipe age, and temperature. Kleiner and Rajani [22] presented a method to forecast the failure rate of water mains by taking into account pipe aging, climate (including freezing index and rain deficit), and operation conditions (e.g., the cumulative length of replaced mains). Freezing index and rain deficit were calculated, given climate data, and then used to predict the failure of water mains. The results proved that the proposed method can be used to predict the failure rate based on climate and operational factors.

Al-Barqawi and Zayed [23] applied ANN to study the impact of temperature, rainfall, number of breaks, and operating pressure on the condition of water mains. Yamijala et al. [24] applied multilinear, multivariate exponential regression and logistic generalized linear model (GLM) to estimate the likelihood of pipe breaks, considering various variables, such as pipe diameter, materials, length, land use, temperature, soil moisture, and soil type. The results demonstrated that GLM performed well compared with other models. Jafar et al. [25] applied artificial neural networks (ANN) to predict the failure rate of water mains based on operational and static variables. The ANN was trained on data collected from a city in Northern France. The results showed that the ANN can be effectively utilized in the failure prediction of water mains. Nishiyama [26] applied ANN to forecast the failure of water mains based on pipe age, length, and soil type. Francis et al. [27] used Bayesian belief networks (BBNs) to predict the failure of water mains, considering several variables, such as pipe materials, diameter, age, demographic variables, and temperature. Shirzad et al. [28] suggested the use of support vector machine (SVM) over ANN to predict the failure rate of water mains. Both models were developed on the basis of several factors, including hydraulic pressure, pipe diameter, length, age, and depth. In another work, Kabir et al. [29] developed a model to predict the failure of water mains using Bayesian and multiple linear regressions based on soil resistivity, land use, freezing index, rain deficit, average temperature, pipe age, and length. The results showed the significant contribution of these factors on the failure prediction of water mains. Furthermore, Bayesian linear regression performed better than multiple linear regression. Demissie et al. [30] developed dynamic Bayesian network (DBN) to predict the number of pipe failures on the basis of static (e.g., pipe material and diameter) and environmental (e.g., freezing index and thawing index, rainfall deficit, and soil corrosion) variables. Farmani et al. [31] studied the impact of static (e.g., pipe diameter and length) and dynamic (e.g., pipe age and temperature) variables on the failure of water mains. K-mean clustering approach was applied to divide the dataset into homogenous groups according to the similarity in water main features. Afterward, the Evolutionary Polynomial Regression (EPR) model was developed to predict the number of failures based on soil type, diameter, and age. Winkler et al. [32] developed ensemble decision tree model using historical data from Austria to predict the failure of water mains on the basis of static and operational variables. Sattar et al. [33] developed an extreme learning machine (ELM) to predict the failure of water mains based on the pipe length, material, pipe protection method, and diameter. The model was trained using 9,500 instances of historical failure records from the Greater Toronto Area, Canada. Shirzad and Safari [34] predicted water main failures using Bayesian model applied multivariate adaptive regression splines and random forest to predict pipe failure in two water distribution systems based on pipe age, pipe length, depth, and average hydraulic pressure. Vališ et al. [35] proposed a dynamic linear approach based on Kalman recursion to forecast the failure rate of water mains using incomplete time series data. Almheiri et al. [2] developed models to predict the time of failure of water mains using intelligent approaches, including ANN, ridge regression, and ensemble decision tree (EDT). The developed models were trained using data collected from Quebec City water mains, considering variables, such as the materials, length, and diameter of pipes.

Motivation

The deterioration of water mains is a complex process that may results from critical factors can be climate variations, dynamic traffic load, internal pressure and corrosion. The impact of these factors may lead to different failure modes, holes due to corrosion, joint failure, circumferential break, longitudinal cracking or split due to high water pressure, and blow out. Partial and incomplete information regarding the failure of water mains can lead to high uncertainties in the failure prediction of water mains [36, 37]. The integrity of pipelines and the prediction of the failure of water mains is essential to maintain the sustainability of water main networks and keep water safe for human consumption. In addition, it helps owners and decision makers to: (1) develop strategies that mitigate the risk of failure; (2) and avoid early replacement of a pipeline before the end of its economic life safe .

Although the impact of climate changes on our infrastructures (e.g., water mains) has become a concern, only a few studies have addressed the failure prediction of water mains studying environmental and climate variables, such as temperature, freezing index, rainfall deficit, soil moisture, and soil type, for the failure prediction of water mains [22, 24, 27, 30, 31], Table 1. Besides, a few studies [22, 26, 35] have forecasted the estimated yearly failure of water mains in the future.

Unlike previous studies, we propose a new temporal framework to predict the failure of water mains at selected locations that share similar monthly failure patterns. The proposed framework can predict the failure of any water mains and at two locations simultaneously. Besides, we examine the contribution of other climate variables in the failure prediction of water mains, such as frost days, frost depth, dry index, and heavy precipitation. The main objectives of this study are as follows: (1) is to propose a temporal framework that can capture the spatial and temporal pattern of failure frequency, and accurately predict the monthly failure of water mains up to nine-time step ahead at selected locations, (2) is to validate the model reliability and performance via Monte Carlo simulations. To achieve these objectives, we collect spatiotemporal data of failure frequency of water mains and climate for selected locations, calculate and define climate variables from given climate data, and apply Granger causality test to define climate variables that provide useful information regarding the failure prediction of water mains in multiple cities. The definition of failure frequency here is the number of breaks per water mains that occurred at a specific time and location within a water distribution system (WDS). The failure modes of water mains at selected locations can be circumferential break, holes, joint failure, blow out, or longitudinal cracking. Pipe corrosion is the predominant reason for pipe failure modes [2].

Spatiotemporal data specifying where and when the failure of water mains occurred are collected with respect to climate records. In other words, spatiotemporal data are an extension of spatial databases that manage time and space information [38, 39]. Climate change can substantially affect the condition and service life of buried utilities, including water infrastructures. Canada's climate varies spatially (across regions) and temporally (from one season to another). Therefore, In this study, two Canadian cities (London and Quebec) are chosen as the selected locations for the failure prediction of water mains. As stated by Boyle et al. [7], these cities are prone to climate hazards (e.g., floods) and climate variables (e.g., changing in temperature and precipitations) that can cause damage to water distribution systems (e.g., water mains). Furthermore, Quebec and Ontario are the highest populated territories in Canada. Quebec and Ontario represent 23% and 39% of the total Canadian population, respectively (see Fig. 1), [40].

<Fig. 1>

<Table 1>

In this study, we develop a model that predict the failure of water mains at two water distribution systems simultaneously. Through data exploration, we find that the multivariate series failure of water mains in both cities (a) Quebec (b) and London follows similar temporal patterns (as shown in Figure 2). The proposed model also can be extended to predict the failure of water mains at more than two locations under climate variations. Besides, we propose two scenarios to predict the failure of water mains to prove that contribution of climate variations in predicting the failure of water mains (see section 2).

<Fig. 2>

2. The Proposed Framework

In the real world, a time series is behaviorally dependent on other processes; therefore, it is essential to simultaneously consider several time series, which are referred to as multivariate time series. Besides, a time series at different spatial locations can be considered a multivariate time series.

The linear interdependence among multivariate time series can be captured using different time series models, such as VAR. VAR is a stochastic process model that relates the values of the multivariate time series at a time *t* to some linear combinations of the multiple time series at previous times.

Suppose a K-variate time series VAR(p) model is given by

$$\boldsymbol{Y_t} = \boldsymbol{\Phi_1 Y_{t-1}} + \boldsymbol{\Phi_2 Y_{t-2}} + \cdots + \boldsymbol{\Phi_p Y_{t-p}} + \boldsymbol{W_t}\,, \qquad (1)$$

where $\boldsymbol{Y_t} \equiv (Y_t^{(1)}, \dots, Y_t^{(K)}), \{\boldsymbol{\Phi}_l : l = 1, \dots, p\}$ are $K \times K$ state-transition matrices and $\{\boldsymbol{W_t}\}$ is a white-noise process for each K-variate with a mean of zero and a covariance matrix Q. This white noise is assumed to be Gaussian, which can be expressed as $\boldsymbol{W_t} \sim \boldsymbol{iid\ Gau}\,(\boldsymbol{0},\ \boldsymbol{Q})$, where $\boldsymbol{iid}$ refers to independent and identically distributed. Moreover, the transition matrix $\boldsymbol{\Phi}$ contains weights that describe how the multivariate time series are linearly combined to influence the time series at a certain time. For instance, the weights in the transition matrix $\mathbf{\Phi_1}$ define how the multivariate time series at $t-1$ are linearly combined to influence the time series at time t.

Suppose VAR (1) process

$$\boldsymbol{Y_t} = \boldsymbol{\Phi}\, \boldsymbol{Y_{t-1}} + \boldsymbol{W_t}, (2)$$

Assume that $\mathbf{Y_t}$ is a stationary process and has a mean of zero means; therefore, all the autocovariance and cross-covariance functions depend on the lag difference between temporal and spatial indices. The lagged covariance matrices for observations $\mathbf{Y_1}, \ldots, \mathbf{Y_T}$ can be obtained by

$$\hat{\boldsymbol{c}}_{\boldsymbol{Y}}^{(\boldsymbol{\tau})} \equiv \frac{\mathbf{1}}{\boldsymbol{T}} \sum_{\boldsymbol{t}=\mathbf{1}}^{\boldsymbol{T}-\boldsymbol{\tau}} (\boldsymbol{Y_{t+\tau}} - \hat{\boldsymbol{\mu}})\, (\boldsymbol{Y_t} - \hat{\boldsymbol{\mu}})\, , (3)$$

where $\mathbf{C_Y^{(\tau)}}$ is to be the τth lagged covariance matrices of the process $\{\boldsymbol{Y_t}\}$ and $\hat{\boldsymbol{\mu}}$ is the *K*-dimensional empirical mean. The parameter matrices $\mathbf{\Phi}$ and $\mathbf{Q}$ can be estimated via the maximum likelihood or method of moments . The $\mathbf{\Phi}$ and $\mathbf{Q}$ estimators of the method of moments are given by Eqs. (4) and (5) [38].

In this study, the parameter estimation is based on the maximum likelihood. More details can be found in the study of [41].

$$\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{C}}_{\boldsymbol{Y}}^{(\mathbf{1})}\, (\hat{\boldsymbol{C}}_{\boldsymbol{Y}}^{(\mathbf{0})})^{-\mathbf{1}}.\ (4)$$

$$\hat{\boldsymbol{Q}} = \hat{\boldsymbol{C}}_{\boldsymbol{Y}}^{(\mathbf{0})} - \hat{\boldsymbol{\Phi}}\, (\hat{\boldsymbol{C}}_{\boldsymbol{Y}}^{(\mathbf{1})})\ . (5)$$

In $\mathrm{VARX}_{k,m}(p, s)$, $\{\mathbf{Y_t}\}$ is modeled in terms of its own previous p and s values of a stationary *m*-dimensional vector series $\{\mathbf{X_t}\}$:

$$\boldsymbol{Y_t} = \sum_{l=1}^{p} \boldsymbol{\Phi_l}\,\boldsymbol{Y_{t-l}} + \sum_{j=1}^{s} \boldsymbol{B_j}\,\boldsymbol{X_{t-j}} + \boldsymbol{W_t}, (\boldsymbol{6})$$

where p refers to the number of preceding values, $\{\boldsymbol{\Phi}_l \in R^{k\times k}\}_{l=1}^{p}$ are the estimated endogenous parameter matrices, and $\{\mathbf{B}_j \in R_j^{k\times k}\}_{j=1}^{s}$ are the estimated exogenous parameter matrices.

In this study, $Y \in R^{t\times d}$ and $X \in R^{t\times d}$ refer to the multivariate time series (e.g., spatiotemporal failure records), and climate variables collected from sensors (weather stations) at a time step t (t refers to a time-of-month index). The main aim of this study is the failure frequency prediction of water mains $\hat{y} \in R^{n}$ (number of breaks per water mains that occurred within a WDS) under climatic variations X. Climate variables, as exogenous variables, were incorporated into the model. Two scenarios are proposed in this study. In scenario 1, the failure of water mains at selected locations relies on its previous values. In scenario 2, the failure also relies on climate variations. The proposed framework is a multivariate time series (VAR) model where the future prediction of a time series Y_t relies on its previous values, whereas in the (VARX) model, Y_t relies on the previous values of Y_{t-1} and X_t. The proposed framework for the failure prediction of water mains that incorporates the contribution of climatic variations, Fig. 3. Intensive reviews of the literature were conducted to identify critical climate variables that contribute to the failure of water mains. Failure records of water mains at selected locations (Quebec and London, Ontario) were collected from municipalities, and climate data were obtained from weather stations located at selected locations. The climate and failure events were connected through the date of the failure of the water mains. The Granger causality test was then applied to investigate the climate variables that can provide statistically significant information that can be useful in predicting the failure of water mains. Two models were developed to predict the failure of water mains: vector autoregression

(VAR) and vector autoregression with exogenous variables (VARX). Both models can be applied to predict the failure of water mains at selected locations. However, the VARX model incorporates the impact of climatic variations on the failure prediction of water mains. Model calibration was implemented to define the number and type of the parameters in the model. Several candidate models of VAR and VARX were developed to select the best model in terms of the performance based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC) criteria (see Section 2). In addition, the models were designed to predict the failure of water mains up to nine months ahead. The developed models were evaluated using prediction mean square error, root mean square error, and accuracy. Finally, Monte Carlo (MC) simulation was applied to verify the reliability of the developed model.

<Fig. 3>

Granger Causality Test

In this study, the prediction of future water main failures relies on their own previous values and climate variables that "Granger-cause" the failure. The Granger causality test was applied to ensure that climate variables were helpful in the failure prediction of water mains at both selected locations. More specifically, suppose that X_t is a time series (or an "exogenous variable" in the present study) that can be used to predict the model target (Y_t). X_t can then be said to "Granger-cause" Y_t if the prediction of Y_t in terms of the previous values of Y_t and X_t is statistically significant more than doing so with the previous values of Y_t. An F-test was applied to obtain the p-value and find whether X_t provides statistically significant information that can be useful in predicting the failure of water mains (Y_t). In F-test, the null hypothesis (H_0) assums that all estimated coefficients of climate variables are equal to zero. In other words, these climate variables are not contributing to the failure of water mains. Whereas, alternative hypothesis (H_a) assumes that at one estimated coefficient of climate variable is nonzero and thus contributing to the failure of WDS. We reject null hypothesis if p-value is less than the confidence interval ($\alpha = 0.05$).

Let $\{X_t\}_{t=1}^{T}$ be the lagged variables of X and $\{Y_t\}_{t=1}^{T}$ for Y. Furthermore, suppose the vector $\{X_t\}_{t=1}^{T}$ can be donated as $\overrightarrow{X_t}$, and the Granger casualty test was performed by conducting the following regressions (Eqs. (7) and (8)):

$$Y_t \approx \boldsymbol{\Phi}.\overrightarrow{Y_{t-1}} + \boldsymbol{B}.\overrightarrow{X_t}. \quad (7)$$

$$Y_{t,k} \approx \boldsymbol{\Phi}.\overrightarrow{Y_{t-1}}. \quad (8)$$

Model Calibration and Selection Criteria

Considering the completeness of the model in terms of the number and type of the parameters that were to be incorporated into the predictive model is crucial. Time lags were considered during the model calibration. To achieve this goal, a greater number of estimation parameters were specified at first, and then the number of parameters was reduced to avoid model complexity and overfitting. The goodness of fit was used to test the performance of the statistical model. The best model can be selected among several candidate models. In a time series analysis, the AIC and BIC are common criteria used in model selection when the model parameters are estimated via the maximum likelihood method [42]. The candidate model with a minimum value of AIC and BIC can be chosen as the best model for predicting future values. AIC and BIC can be calculated using a set of model parameters (θ), the likelihood of the candidate model, the data evaluated at the maximum likelihood estimation of (θ), and the number of estimated parameters of the candidate model (k) as expressed in Eqs. (9) and (10) [43, 44].

$$AIC = -2 \log L(\hat{\theta}) + 2k. \quad (9)$$

$$BIC = -2\, log\, L(\hat{\theta}) + 2\, log\, k. \ (10)$$

Predictive Model Performance

The last two years of time series data are held out to validate the performance of the predictive model, and the remaining data were used to fit the model. A set of mathematical equations prediction mean square error (PMSE), root mean square error (RMSE), and accuracy was used for the model evaluation (Eqs. (11) and (12)). The model accuracy was measured by Eq. (13), where T refers to selected locations and time steps in the holdout sample.

$$PMSE = \frac{1}{T} \sum_{i=1}^{T} (FF(i) - forecast.FF\ (i))^2. \ (11)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^{T} (FF(i) - forecast.FF\ (i))}. \ (12)$$

$$Accuracy = 1 - \frac{1}{T} \sum_{i=1}^{T} \left| \frac{forecast.FF\ (i) - FF(i)}{FF(i)} \right|. \ (13)$$

3. Reliability of the Proposed Model

Monte Carlo (MC) simulation is applied to test for uncertainties in the developed predictive model. It is a class of computational algorithms that relies on drawing repeated S samples from the population distribution, referred as X_1, X_2,.. X_S. Given the samples, the distribution of $f(X)$ can be estimated using the empirical distribution of $\{f(X_S)\}_{S=1}^{S}$ which called MC approximation. MC simulation can be used to compute the expected/predictive value of any function of a random variable. Basically, after the random samples have been drawn, the arithmetic mean of the function applied to the samples can be calculated according to Eq. (14), [45].

$$E[f(X)] = \int f(X)p(x)dx \approx \frac{1}{S}\sum_{s=1}^{S} f(x_s), \quad (14)$$

where $x_s \sim p(X)$ the MC integration that is based on evaluating the estimated function over a range of fixed points. In this study, Monte Carlo was used to approximate the predicted value of failure frequency of water mains. In other words, the model reliability of the developed model was computed using MC approximation. Accuracy of MC approximation was measured under different sample iterations, from 100 to 1000, to validate the reliability of the developed model (using Eq. 13). The multivariate time series (e.g., spatiotemporal failure records) are assumed to be jointly Gaussian distributed with a mean of zero and covariance matrix **Q.** In this study, the developed temporal model was used to generate a set of random process samples that reflected the statistical properties of the actual data. The conditional simulation was performed using MC that required the cumulative distribution function of sample values (iterations) and the basic parameters input to be constant at each time *t* in the prediction horizon, and from iteration to another. The system was simulated from 400 to 1000 iterations to represent the possible distribution of the estimated of $f(X)$ using MC simulation. In other words, iterations of failure frequency were generated using MC simulation. Sample statistics was then estimated from the generated iterations at each time *t* in the prediction horizon. The predicted values of failure frequency of water main are the mean across the iterations.

4. Spatiotemporal Data and Collection Method

The proposed approach was developed to predict the failure of water mains at selected locations under climatic variations. Climate data were collected from a weather station located in selected locations, whereas the failure records of water mains were collected from municipalities. The climate and failure records were connected through the date of the failure of water main. Data connection and collection procedures of the climate and failure of water mains are depicted in detail in Fig. 4. The subsequent sections explain the data aggregation and collection of climate data and failure records of water mains in detail.

<Fig. 4>

Failure Data of Water Mains

The total monthly failure frequency of water mains at each spatial location was calculated from 1987 to 2001 using historical failure records. As prior mentioned, failure frequency is the number of breaks per water mains that occurred at a specific time and location within water distribution systems (WDS). The failure data in London and Quebec were obtained from the City of London and Sainte-Foy, respectively. The failure data consist of recorded information, such as pipe material, diameter, and length, year of installations, and breaks for each pipe within WDS. According to the failure data obtained from Quebec, the study period was defined to be spanned 15 years (1987-2001). Water mains in Quebec consist of different types of pipe material, namely, gray cast iron (CI), ductile iron with lining (DIL), ductile iron without lining (DIN), polyvinyl chloride (PVC), and concrete pipes. Similarly, pipe materials used in London are CI, PVC, concrete, copper, galvanized, polyethylene (PE), high-density polyethylene (PEX), steel, and cement asbestos pipes. The size of the water main is between 300mm and 600mm, and from 300 to 450 mm in Quebec and London, respectively. Based on our analyzing the data, we decided to study the monthly failure patterns (the number of breaks per water distribution system per month). Additionally, we assume that it will be possible to observe the impact of climate (seasonality) variations upon failure occurrence during single months (as previously shown in Figure 2).

Climate Data and Climate Variables

The climate data for a 15-year period (1987–2001) were collected from the Environment and Climate Change Canada (ECCC). These climate records were documented by weather stations that are located in different regions to monitor the climatic variations all over Canada. In this study, the climate records were obtained from Montréal–Mirabel International Airport station for Quebec and St. Thomas Water Pollution Control Plant for London. Each weather station has recorded climate information, such as air temperature (2 m from the surface) in Celsius (ºC), relative humidity (%), 10 m wind speed (WS) in (km/h), and P (mm). The weather station located in Quebec (named Montreal- Mirabel International Airport) is the closet weather station to Sainte-Foy city without any missing records of climate data. Additionally, we compare multiple climate records collected from weather stations located in Quebec to prove that the climate records over Quebec almost the same. Thus, the validity of using climate data from Montreal- Mirabel International Airport weather station. Inn this study, T and P are used to calculate the climate variables in Table 2. Climate variables that provide significant information on the failure prediction of water mains are considered in this study (see Section 7.1). The daily climate data obtained from ECCC were aggregated into average monthly records per year. The characteristics of the climate of the two selected locations are presented in Table 2.

Dry index (DI) was used as a surrogate measure for soil moisture over a defined period, according to Eq. (14) [16], where $\overline{\mathrm{T}}$ and P are the average temperature (°C) and cumulative P (mm) over a defined period, respectively. In this study, DI was calculated for the months from April to October. T and P were defined in the proposed model as they contributed to the failure of water mains [22].

$$DI = \frac{3 \times \bar{T}}{P}\,. \quad (14)$$

Freezing index (FI) and thawing index (TI), as also defined in the proposed model, are expressed in units of degree-days (°Cd). The FI is the total average daily temperature below zero and can be used as a measure of winter severity for the period from December to March (Eq. (15)). Whereas, TI is defined as the total average daily temperature above zero within a defined period (March to May) (Eq. (16)). The authors defined other climate variables that have not been studied in the failure prediction of water mains, including frost days and frost depth. In this study, frost days (F) is proposed by the authors and defined as a measure of the winter severity during winter where $\mathrm{T} < 0$. Whereas, frost depth (FD) can be calculated given FI_c, according to Eq. (17) [46], where FI_c is the cumulative freezing index. The name, unit, and description of each variable are summarized in Table 2. The climatic characteristics and failure frequency range of water mains of the studied spatial locations are presented in Table 3.

$$FI_{(i)} = \left|\sum_{i=1}^{n} \bar{T}\right|, \bar{T} < 0. \quad (15)$$

$$TI_{(i)} = \left|\sum_{i=1}^{n} \bar{T}\right|, \bar{T} > 0. \quad (16)$$

$$FD_{frost} = 0.0174 \times {FI_c}^{0.67}. \ (17)$$

<Table 3>

Data Preprocessing and Scaling

Data aggregation was applied to calculate the total monthly failure frequency of water mains over the defined period (1987-2000) at selected locations. The failure frequency records and climate data were combined through the date of the failure of the water mains (1987 and 2001). The last two years of data were held out for the validation of the performance of the predictive model. Failure and climate data were normalized to be in the range of [0, 1] using Eq. 18 to maintain the performance of the developed model.

$$X_{\text{normalized}} = \frac{X - X_{min}}{X_{max} - X_{min}}, \qquad (18)$$

where X represents the input and the output of the model (failure frequency and climate variables).

5. Results

The failure of water mains can occur in summer and winter. However, the chance of failure of water mains in winter was higher than that in summer (as previously shown in Figure 2). Moreover, the failure frequency in London was higher than that in Quebec over the defined period (1987–2001). This scenario may be caused by the diversity in geographical characteristics or other reasons. Additionally, the time series decomposition of water main failures in Quebec and London proves the seasonality trend in the failure pattern, Fig. 5. Seasonality means that the changes occurred at specific regular intervals, such as monthly, weekly, or quarterly. Thus, we aim to build a model that can capture seasonality trends.

<Fig. 5>

To reduce model complexity, climate variables that provide significant information regarding the failure of water mains were examined in this study. As previously mentioned, the Granger causality test was applied to consider climate variables that are useful in the failure prediction of water mains. The null hypothesis for the Granger causality test is that X_t does not Granger-cause Y_t, so an F-test was applied to obtain the p-value and find whether X_t provides statistically significant information that can be useful in predicting the failure of water mains (Y_t). The Granger causality test shows that the failure of water mains is significantly influenced by some climate variables (p-value < 0.05). Climate variables (X_t), such as T, F, FI, and FD, give significant information regarding the failure of water mains at selected locations. Thus, we use these climate variables in predicting the failure of water mains.

The VAR model was developed to predict the failure of water mains and capture the interdependencies among the failures of water mains with lagged endogenous variables at selected locations. The VARX model, on the other hand, was developed to predict the failure of water mains under climatic variations using MATLAB version R2018a software. The failure of water mains relies on its previous values and climate variables, including T, FI, F, and FD. The best model of VAR and VARX that was selected yields the minimal AIC and BIC values, given a set of candidate models. Table 4 presents a set of VAR candidate models, where VAR16 was selected as the best model. Table 5 illustrates a set of VARX candidate models, where VARX3 was selected as the best model. The complexity of VARX3 is less as it required a smaller number of parameters (30 parameters) to be estimated compare to the best candidate model of VAR (66 parameters). Besides, VARX3 incorporates the contribution of climate variations on water main failure.

<Table 4>

<Table 5>

The performance of VARX3 was tested using error matrices (Eqs. (11) to (13)). The average prediction error (PMSE, RMSE, and RMSE %) and percentage accuracy for the multivariate of the failure prediction of water mains at selected locations (Quebec and London), see Table 6. A nine-month prediction was issued every one month using the most recent actual failure dataset. The results prove that the model can predict up to nine months ahead with a good average accuracy (Table 6). Fig. 6 depicts the time series plots of three-month (a and b) and nine-month (c and d) predictions of water mains at selected locations (Quebec and London). A high failure frequency for the three-month prediction will be accrued in winter. The failure frequency in London is almost four times that in Quebec. Although the failure frequency is higher in winter, it is still high in London during summer. Additionally, Fig. 6 (c and d) shows that the proposed framework can predict the failure of water mains up to nine months ahead. The model can also capture the seasonality of the failure of water mains. Fig. 7 shows the performance of the failure prediction of water mains at selected locations. Overall, the accuracy of failure prediction up to nine months ahead is steady for Quebec and London (see Table 6 and Fig. 7).

<Table 6>

<Fig. 6>

<Fig. 7>

The system can be simulated thousands of times between 10^3 and 10^6 using MC simulation [47]. However, MC simulations were implemented at various numbers of iterations from 100 to 1000 to avoid expensive computational costs. MC simulations, on the other hand, were performed to verify the performance of the proposed model and measure model reliability. Results prove that the MC simulations converge to a good average accuracy from 100 to 1000 iterations (Fig. 8a). However, the MC simulation was performed at iterations of 400 to avoid computational cost.

Results prove that the MC simulations and proposed model have approximately the same accuracy (see Fig. 7 and 8b). Additionally, the actual and the mean predicted failure frequency of water mains at selected locations using different iterations of MC simulations is nearly the same, Fig. 9. The stacked values of actual and predicted failure of water mains for all iterations of MC simulation follow the same patterns over time, Fig. 9. Results of MC simulations prove the reliability of the proposed model. Therefore, the proposed model is reliable and can be applied to predict the failure of water mains nine months ahead at selected locations. In addition, the model incorporates the impact of climatic variations on the failure of water mains. Thus, municipalities can apply the developed model to predict the failure patterns of water mains given climate data and historical failure frequency of water mains at any water distribution system.

<Fig. 8>

<Fig. 9>

6. Conclusions

Failure prediction of water mains under climate variations is essential to avoid economic and environmental losses that may result from failure. The proposed approach can help decision-makers in municipalities to understand and predict failure of water mains under climate variations in any water distribution systems. Failure prediction of water mains, on the other hand, is an inexpensive approach compare to direct inspection for estimating the structural deterioration of a water distribution system. It may also help decision-makers come up with strategies and plans that will mitigate the failure of water mains under climatic variations. In future work, we will investigate the failure of water mains in other territories that do not share similar climate conditions. Additionally, we will study the characteristic of each induvial pipe within WDS and include other critical factors that may contribute to the failure of water mains. Although long-term failure prediction is possible, we believe that the failure of water mains relies on other critical factors. Hence, we will extend the model for long-term predictions (10 years ahead) with the incorporation of the most critical factors. Besides, we will utilize a powerful mathematical tool such as artificial neural networks (ANN) using big data to predict the failure of water mains.

In this study, we select Quebec and London that are the most populated territories in Canada. Based on data exploration and the proposed framework, the following conclusions can be made:

- Similar to previous studies [9, 17], the failure of water mains may occur during all seasons and reach its peak during winter seasons.
- The failure frequency of water mains in London is higher than that in Quebec. This condition may be due to other factors, such as soil types at each spatial location (the water mains in London may be buried inside active soil that experiences shrinkage when winter is over).
- A similarity in pipe failure patterns was found at both selected locations (Quebec and London).
- The failure pattern of failure frequency follows a monthly periodicity at both selected locations may be due to climate variations. Results of Granger causality analysis, on the other hand, prove that the fluctuations of climate factors such as FI, DI, F, and T may significantly (p-value < 0.05) contribute to the failure of water mains.
- The proposed temporal model can provide an accurate failure prediction of nine months ahead $(t + 9)$ of water mains at two locations simultaneously with data arriving at a frequency of one month. Although the failure events at two water distribution systems are

independent, yet the model shows a promising capability of predicting the failure of water mains under climate variations.

- Many critical factors other than climate variations may affect the failure of water mains that are hard to obtain most of the time. Thus, long-term failure perfection can be inaccurate. The proposed model, on the other hand, can be extended to predict the failure of water main at more than two locations and incorporate the impact of other factors.
- MC simulation was implemented to quantify the predicted mean of failure frequency of water mains. Results of MC simulations show how the failure of water mains at both selected locations changes over time at different iterations. Besides, the results of MC simulations prove the reliability of the developed model.

Acknowledgements

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Financial support provided by McGill-UAE fellowships in Science and Engineering to the first author is highly appreciated. The authors would like to thank Dr. Khalid Shahata (London Ontario) for their support in providing the failure data of water mains. The authors also would like to sincerely acknowledge Prof. Laxmi Sushama and Dr. Seok Geun Oh of the Civil Engineering Department at McGill University for their support in providing climate data.

Failure data of water mains of Quebec and London Ontario can be obtained from the City of London and Sainte-Foy municipalities, respectively. Climate data can be accessed through Environment and Climate Change Canada (ECCC), (https://www.canada.ca/en/environment-climate-change.html).

Author contributions

Z.A., M.M., and T.Z. conceived and designed the experiments; Z.A. conducted the experiments, analyzed and interpreted the results, and drafted the manuscript with significant support and comments from the rest of the co-authors. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

[1] J. Makar, R. Desnoyers, and S. McDonald, "Failure modes and mechanisms in gray cast iron pipe," *Underground Infrastructure Research*, pp. 1-10, 2001.

[2] Z. Almheiri, M. Meguid, and T. Zayed, "Intelligent Approaches for Predicting Failure of Water Mains," *Journal of Pipeline Systems Engineering and Practice*, vol. 11, no. 4, p. 04020044, 2020.

[3] ASCE. "2017 infrastructure report card." https://www.infrastructurereportcard.org/ (accessed July 10 2019).

[4] CIRC. "Canadian Infrastructure Report Card." http://canadianinfrastructure.ca/en/index.html (accessed July 22,2019

[5] B. Wols and P. Van Thienen, "Impact of weather conditions on pipe failure: a statistical analysis," *Journal of Water Supply: Research and Technology—AQUA*, vol. 63, no. 3, pp. 212-223, 2013.

[6] B. Wols, K. Van Daal, and P. Van Thienen, "Effects of climate change on drinking water distribution network integrity: predicting pipe failure resulting from differential soil settlement," *Procedia Engineering*, vol. 70, pp. 1726-1734, 2014.

[7] J. Boyle, M. Cunningham, and J. Dekens, "Climate Change Adaptation and Canadian Infrastructure," *International Institute for Sustainable Development (IISD): Winnipeg, MB, Canada*, 2013.

[8] J. Vreeburg, I. Vloerbergh, P. Van Thienen, and R. De Bont, "Shared failure data for strategic asset management," *Water Science and Technology: Water Supply*, vol. 13, no. 4, pp. 1154-1160, 2013.

[9] S. Bruaset and S. Sægrov, "An analysis of the potential impact of climate change on the structural reliability of drinking water pipes in cold climate regions," *Water*, vol. 10, no. 4, p. 411, 2018.

[10] J. Makar, "Failure analysis for grey cast iron water pipes," in *Proc., AWWA Distribution System Symp., American Water Work Association, Denver*, 1999.

[11] B. Rajani, Y. Kleiner, and J.-E. Sink, "Exploration of the relationship between water main breaks and temperature covariates," *Urban Water Journal*, vol. 9, no. 2, pp. 67-84, 2012.

[12] G. H. Bahmanyar and T. B. Edil, "Cold weather effects on underground pipeline failures," in *Pipelines in Adverse Environments II*, 1983: ASCE, pp. 579-593.

[13] A. Habibian, "Effect of temperature changes on water-main breaks," *Journal of transportation engineering*, vol. 120, no. 2, pp. 312-321, 1994.

[14] C. Goodchild, T. Rowson, and M. Engelhardt, "Making the earth move: modelling the impact of climate change on water pipeline serviceability," *Computing and Control in the Water Industry 2009: Integrating Water Systems*, pp. 807-811, 2009.

[15] D. Laucelli, B. Rajani, Y. Kleiner, and O. Giustolisi, "Study on relationships between climate-related covariates and pipe bursts using evolutionary-based modelling," *Journal of Hydroinformatics*, vol. 16, no. 4, pp. 743-757, 2014.

[16] D. Fuchs-Hanusch, F. Friedl, R. Scheucher, B. Kogseder, and D. Muschalla, "Effect of seasonal climatic variance on water main failure frequencies in moderate climate regions," *Water Science and Technology: Water Supply*, vol. 13, no. 2, pp. 435-446, 2013.

[17] G. Kabir, S. Tesfamariam, A. Francisque, and R. Sadiq, "Evaluating risk of water mains failure using a Bayesian belief network model," *European Journal of Operational Research,* vol. 240, no. 1, pp. 220-234, 2015.

[18] Y. Kleiner and B. Rajani, "Quantifying effectiveness of cathodic protection in water mains: theory," *Journal of infrastructure systems,* vol. 10, no. 2, pp. 43-51, 2004.

[19] Y. Kleiner and B. Rajani, "I-WARP: Individual water main renewal planner," *Drinking Water Engineering and Science,* vol. 3, no. 1, pp. 71-77, 2010.

[20] S. Gould, F. Boulaire, S. Burn, X.-L. Zhao, and J. Kodikara, "Seasonal factors influencing the failure of buried water reticulation pipes," *Water Science and Technology,* vol. 63, no. 11, pp. 2692-2699, 2011.

[21] S. Gould and J. Kodikara, "Exploratory statistical analysis of water reticulation main failures," *Research Rep. RR11, Monash Univ., Melbourne, Australia,* 2008.

[22] Y. Kleiner and B. Rajani, "Forecasting variations and trends in water-main breaks," *Journal of infrastructure systems,* vol. 8, no. 4, pp. 122-131, 2002.

[23] H. Al-Barqawi and T. Zayed, "Condition rating model for underground infrastructure sustainable water mains," *Journal of Performance of Constructed Facilities,* vol. 20, no. 2, pp. 126-135, 2006.

[24] S. Yamijala, S. D. Guikema, and K. Brumbelow, "Statistical models for the analysis of water distribution system pipe break data," *Reliability Engineering & System Safety,* vol. 94, no. 2, pp. 282-293, 2009.

[25] R. Jafar, I. Shahrour, and I. Juran, "Application of Artificial Neural Networks (ANN) to model the failure of urban water mains," *Mathematical and Computer Modelling,* vol. 51, no. 9-10, pp. 1170-1180, 2010.

[26] M. Nishiyama, "Forecasting water main failures in the city of Kingston using artificial neural networks," 2013.

[27] R. A. Francis, S. D. Guikema, and L. Henneman, "Bayesian belief networks for predicting drinking water distribution system pipe breaks," *Reliability Engineering & System Safety,* vol. 130, pp. 1-11, 2014.

[28] A. Shirzad, M. Tabesh, and R. Farmani, "A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks," *KSCE Journal of Civil Engineering,* vol. 18, no. 4, pp. 941-948, 2014.

[29] G. Kabir, S. Tesfamariam, J. Loeppky, and R. Sadiq, "Integrating Bayesian linear regression with ordered weighted averaging: Uncertainty analysis for predicting water main failures," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering,* vol. 1, no. 3, p. 04015007, 2015.

[30] G. Demissie, S. Tesfamariam, and R. Sadiq, "Prediction of pipe failure by considering time-dependent factors: Dynamic Bayesian belief network model," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering,* vol. 3, no. 4, p. 04017017, 2017.

[31] R. Farmani, K. Kakoudakis, K. Behzadian, and D. Butler, "Pipe Failure Prediction in Water Distribution Systems Considering Static and Dynamic Factors," *Procedia Engineering,* vol. 186, pp. 117-126, 2017.

[32] D. Winkler, M. Haltmeier, M. Kleidorfer, W. Rauch, and F. Tscheikner-Gratl, "Pipe failure modelling for water distribution networks using boosted decision trees," *Structure and Infrastructure Engineering,* vol. 14, no. 10, pp. 1402-1411, 2018.

[33] A. M. Sattar, Ö. F. Ertuğrul, B. Gharabaghi, E. McBean, and J. Cao, "Extreme learning machine model for water network management," *Neural Computing and Applications,* vol. 31, no. 1, pp. 157-169, 2019.
[34] A. Shirzad and M. J. S. Safari, "Pipe failure rate prediction in water distribution networks using multivariate adaptive regression splines and random forest techniques," *Urban Water Journal,* vol. 16, no. 9, pp. 653-661, 2019.
[35] D. Vališ, K. Hasilová, M. Forbelská, and Z. Vintr, "Reliability modelling and analysis of water distribution network based on backpropagation recursive processes with real field data," *Measurement,* vol. 149, p. 107026, 2020.
[36] H. Haider, R. Sadiq, and S. Tesfamariam, "Performance indicators for small-and medium-sized water supply systems: a review," *Environmental reviews,* vol. 22, no. 1, pp. 1-40, 2013.
[37] B. Rajani and S. Tesfamariam, "Uncoupled axial, flexural, and circumferential pipe soil interaction analyses of partially supported jointed water mains," *Canadian geotechnical journal,* vol. 41, no. 6, pp. 997-1010, 2004.
[38] N. Cressie and C. K. Wikle, *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.
[39] J. Han, M. Kamber, and J. Pei, "Data mining trends and research frontiers," *Data Min,* pp. 585-631, 2012.
[40] S. Canada. "Population estimates, quarterly." https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901 (accessed July 22,2019
[41] H. Lüktepohl, "Introduction to Multiple Time Series Analysis," ed: Springer-Verlag, New-York, 1993.
[42] H. Akaike, "A new look at the statistical model identification," in *Selected Papers of Hirotugu Akaike*: Springer, 1974, pp. 215-222.
[43] F. Fabozzi, S. Focardi, S. Rachev, and B. Arshanapalli, "Appendix E: Model selection criterion: AIC and BIC," *The Basics of Financial Econometrics; John Wiley & Sons, Inc.: Hoboken, NJ, USA,* pp. 399-403, 2014.
[44] M. J. Brewer, A. Butler, and S. L. Cooksley, "The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity," *Methods in Ecology and Evolution,* vol. 7, no. 6, pp. 679-692, 2016.
[45] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
[46] W. G. Brown, "Difficulties associated with predicting depth of freeze or thaw," *Canadian Geotechnical Journal,* vol. 1, no. 4, pp. 215-226, 1964.
[47] M. Mahmoodian and A. Alani, "Multi-failure mode assessment of buried concrete pipes subjected to time-dependent deterioration, using system reliability analysis," *Journal of failure analysis and prevention,* vol. 13, no. 5, pp. 634-642, 2013.
[48] G. Kabir, S. Tesfamariam, and R. Sadiq, "Bayesian model averaging for the prediction of water main failure for small to large Canadian municipalities," *Canadian Journal of Civil Engineering,* vol. 43, no. 3, pp. 233-240, 2015.
[49] D. W. Boyd, "Normal freezing and thawing degree-days from normal monthly temperatures," *Canadian Geotechnical Journal,* vol. 13, no. 2, pp. 176-180, 1976.

List of Tables

Table 1. Summary of failure prediction models of water mains based on climate and environmental variables 30
Table 2. Summary of climate variables, name, unit, and description..30
Table 3. Monthly record of Climatic Characteristics and failure frequency range of water mains of the selected locations (1987-2001)...30
Table 4. Set of VAR candidate model. 31
Table 5. Set of VARX candidate models. 31
Table 6. Average failure prediction error and accuracy in Quebec and London...31

List of Figures

Fig 1. Population weight of Canada by territory estimated in January 2020 [40]. ..... **Error! Bookmark not defined.**

Fig 2. Failure frequency of water mains at selected locations. .... **Error! Bookmark not defined.**

Fig 3. Proposed framework for the failure prediction of water mains under climatic variations. .... **Error! Bookmark not defined.**

Fig 4. Data collection method of water main failure and climate records. .... **Error! Bookmark not defined.**

Fig 5. Seasonal trend of failure frequency of water mains at (a) Quebec and (b) London (1987-2001). .... **Error! Bookmark not defined.**

Fig 6. Time series plot of the failure of water mains at selected locations (Quebec and London) (a), (b) three months and (c), (d) nine months ahead, respectively. FFQ: failure frequency in Quebec, FFL: failure frequency in London, blue line: actual data (Jan-87—Dec-00), green line: held-out data (Jan-00—Dec-01), red line: prediction (Jan-02—Sep-02). .... **Error! Bookmark not defined.**

Fig 7. Average prediction accuracy of nine months ahead. .... **Error! Bookmark not defined.**

Fig 8. (a) Average accuracy of MC simulations vs. number of iterations, and (b) average accuracy of MC simulations of nine months ahead at selected locations. .... **Error! Bookmark not defined.**

Fig 9. Stacked charts of actual and the mean predicted failure frequency of water mains in (a) Quebec and (b) London using different iterations of MC simulations. .... **Error! Bookmark not defined.**

Table 1. Summary of failure prediction models of water mains based on climate and environmental variables

Reference	Variables	Output
Kleiner and Rajani [22]	Operational and climate variables including freezing Index and rainfall deficit	Failure Rate
Yamijala [24]	Diameter, material, length, land use, soil type, soil moisture, temperature	Likelihood of break
Francis et al. [27]	material, diameter, age, demographic variables and temperature	Pipe breaks
Kabir et al. [48]	# of bursts, age, diameter, length, soil resistivity, soil corrosion	Failure rate
Demissie et al. [30]	Length, Diameter, number of previous failures, type of service connection, freezing index, thawing index, rainfall deficit and soil corrosion	Pipe breaks
Farmani et al.[31]	Length, diameter, age, temperature, freezing index	Pipe breaks

Climate Variable	Unit	Description
Temperature (T)	degree (°C)	Average Monthly T
Precipitation (P)	mm	Average Monthly Pre
Dry Index (DI); [16]	Degree(°C)/mm	Eq. (14)
Frost (F); Assumed by the authors	Days	T<0; from November to March (Winter period)
Freezing Index (FI) [22]	Degree-days (°Cd)	Eq. (15)
Thawing Index (TI) [49]	Degree-days (°Cd)	Eq. (16)
Frost Depth (FD) [46]	m	Eq. (17)

Table 2. Summary of climate variables, name, unit, and description.

Table 3. Monthly record of Climatic Characteristics and failure frequency range of water mains of the selected locations (1987-2001).

Variable	Average T	[a]DI	[a]FI	[a]TI	[a]FD	Failure frequency
Quebec	-18.04-21.52°C	1252 (°C)/mm	559.33 °Cd	504.01 °Cd	1.21 m	0-30
London	-9.98-23.11°C	3124 (°C)/mm	309.4 °Cd	543.7 °Cd	0.81 m	0-170

[a]maximum value of climate variables.

Table 4. Set of VAR candidate model.

Models	Lag Order	AIC	BIC
VAR16	**Lag-16**	**-334.53**	**-140.38**
VAR20	Lag-20	-301.82	-62.98
VAR24	Lag-24	-285.32	-2.811
VAR30	Lag-30	-254.49	91.53

*bold: refers to the best model.

Table 5. Set of VARX candidate models.

Models	Lag Order	AIC	BIC
VARX3	**Lag-3**	**-497.98**	**-406.48**
VARX9	Lag-9	-389.36	-227.88
VARX16	Lag-16	-399.23	-158.02
VARX20	Lag-20	-373.17	-87.73

*bold: refers to the best model.

Table 6. Average failure prediction error and accuracy in Quebec and London.

City	Prediction horizon						
Quebec	1-month	2-month	3-month	4-month	5-mnoth	6-month	9-month
PMSE	0.01	0.164	0.105	0.081	0.067	0.057	0.062
RMSE	0.085	0.125	0.261	0.216	0.165	0.174	0.374
RMSE%	0.122	0.181	0.363	0.423	0.464	0.524	0.65
Accuracy	0.83	0.81	0.66	0.66	0.65	0.66	0.67
London	1-month	2-month	3-month	4-mnoth	5-mnoth	6-mnoth	9-mnoth
PMSE	0.072	0.059	0.041	0.031	0.025	0.025	0.011
RMSE	0.270	0.240	0.184	0.149	0.124	0.124	0.068
RMSE%	0.270	0.337	0.318	0.352	0.329	0.329	0.237
Accuracy	0.72	0.66	0.64	0.62	0.65	0.68	0.69

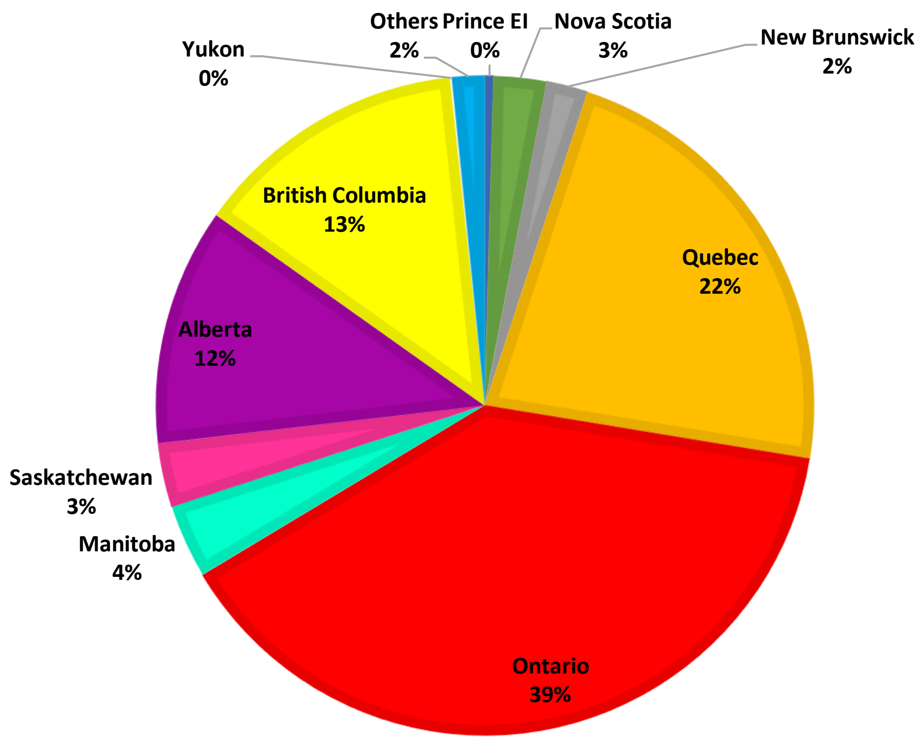
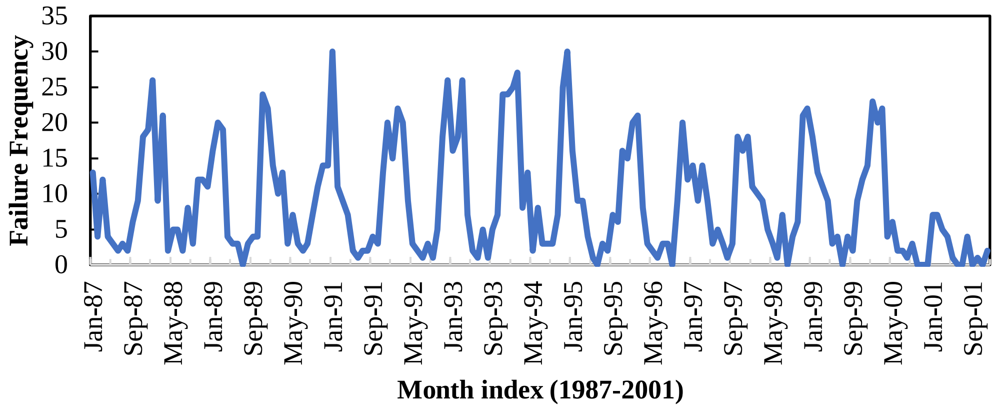
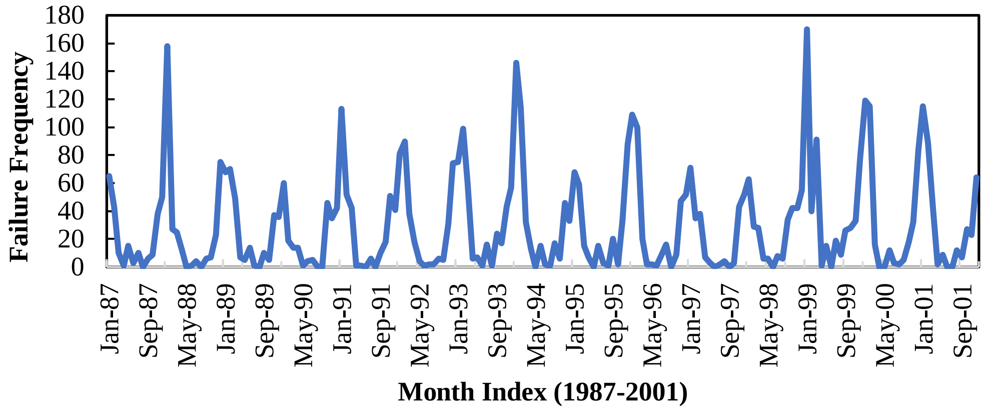


Fig 1. Population weight of Canada by territory estimated in January 2020 [40].



(a) Failure frequency of water mains at Quebec City.



(b) Failure frequency of water mains at London City.

Fig 2. Failure frequency of water mains at selected locations.

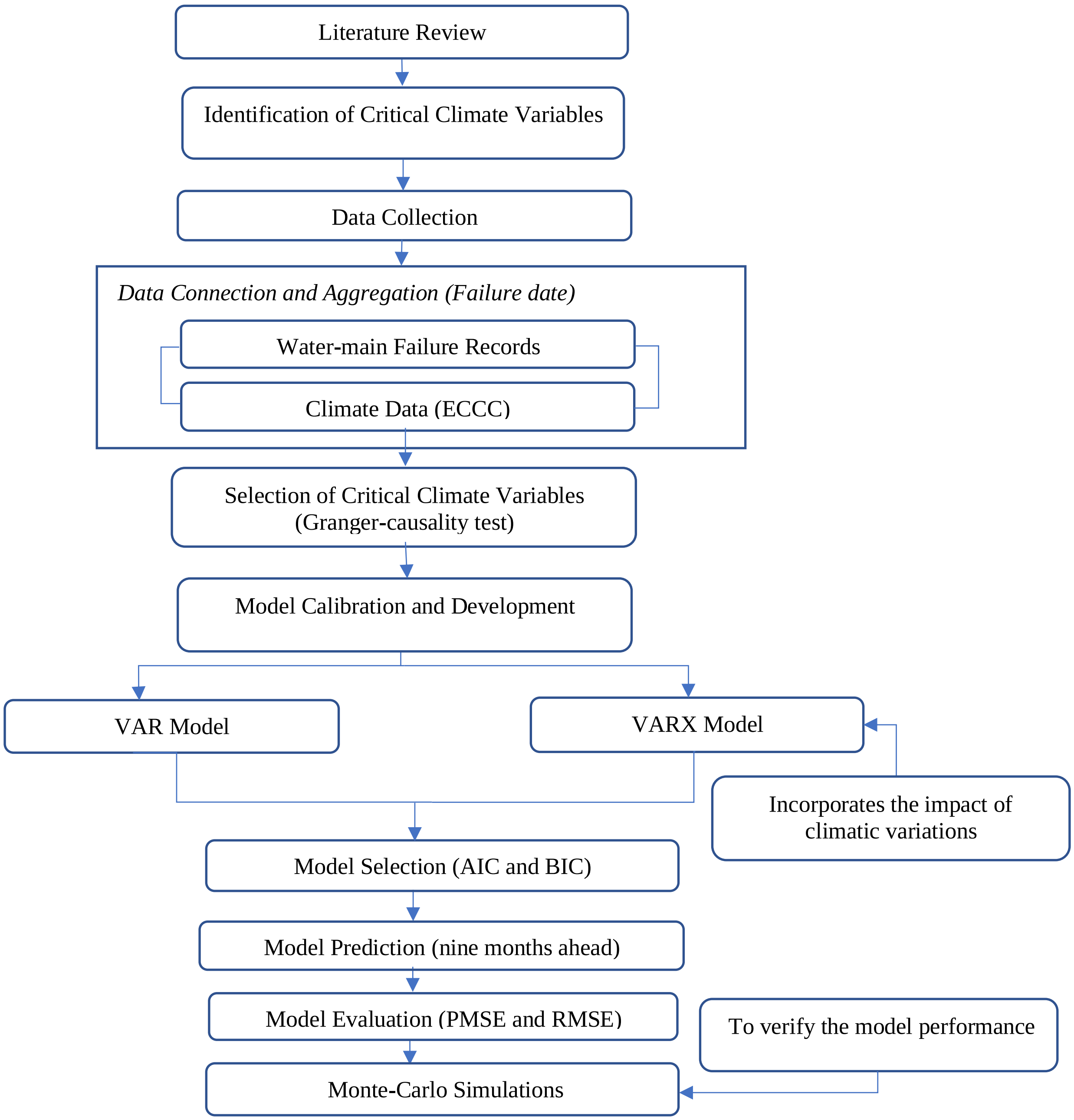


Fig 3. Proposed framework for the failure prediction of water mains under climatic variations.

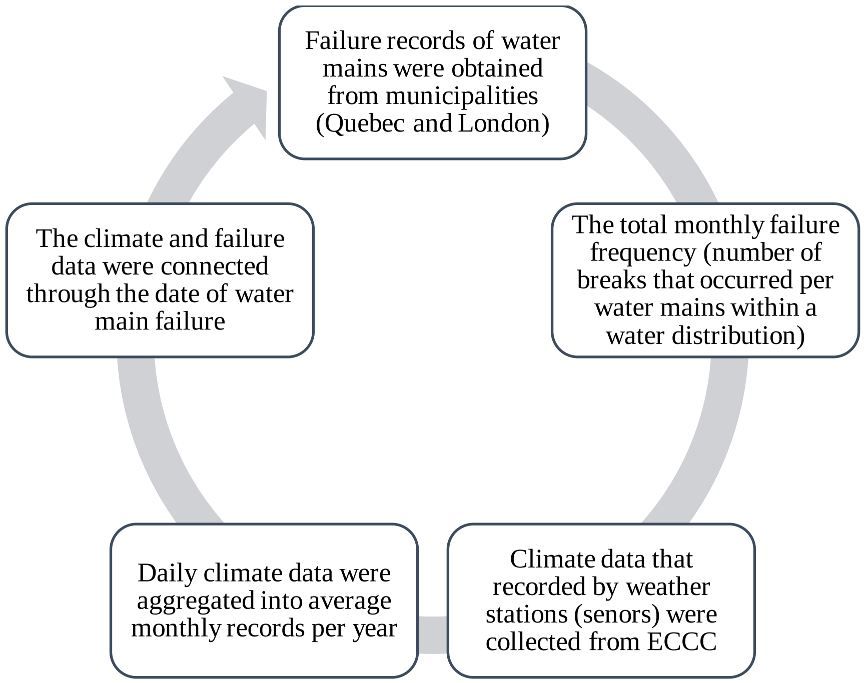
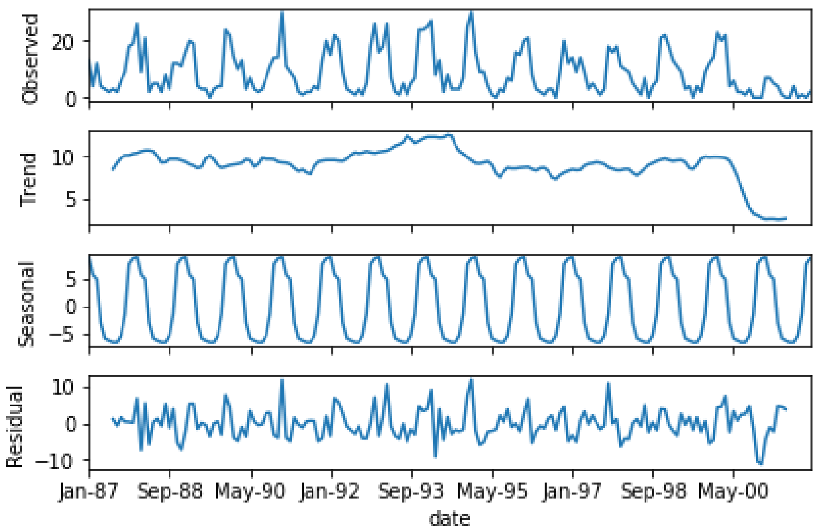
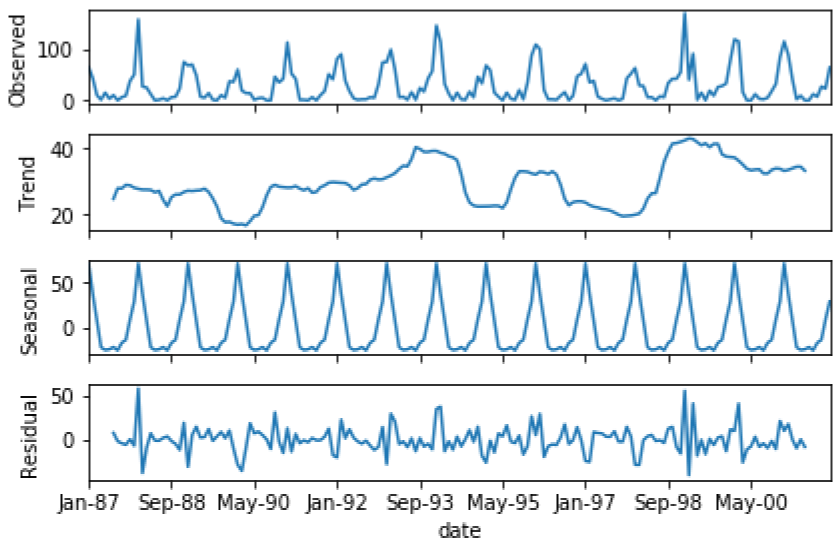


Fig 4. Data collection method of water main failure and climate records.

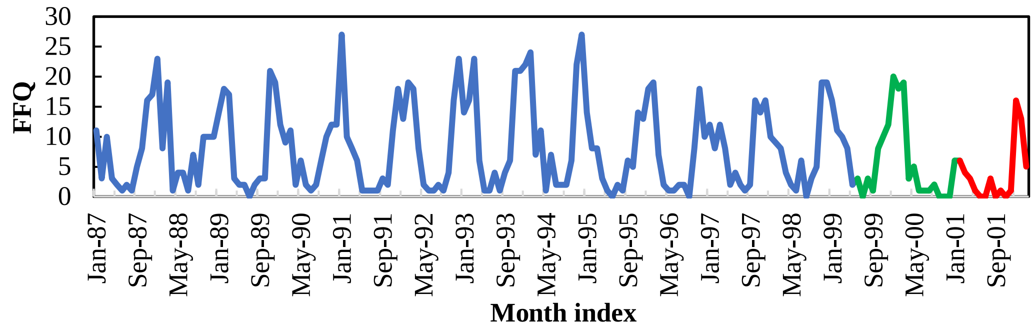

Observed
20
0
Trend
10
5
Seasonal
5
0
−5
Residual
10
0
−10
Jan-87
Sep-88
May-90
Jan-92
Sep-93
May-95
Jan-97
Sep-98
May-00
date

(a) Quebec

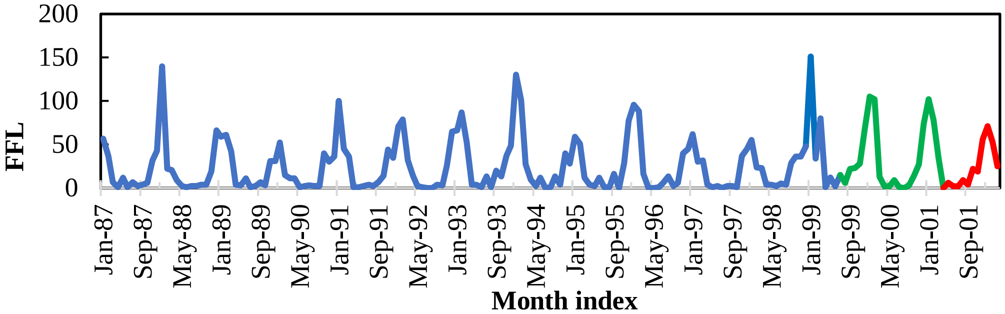


(b) London

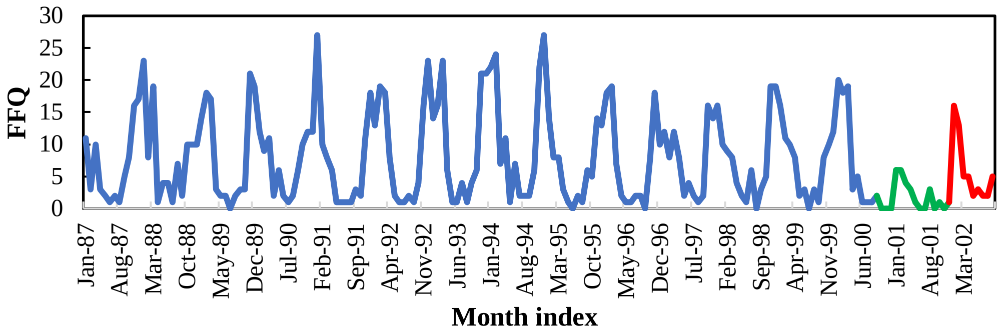
Fig 5. Seasonal trend of failure frequency of water mains at (a) Quebec and (b) London (1987-2001).



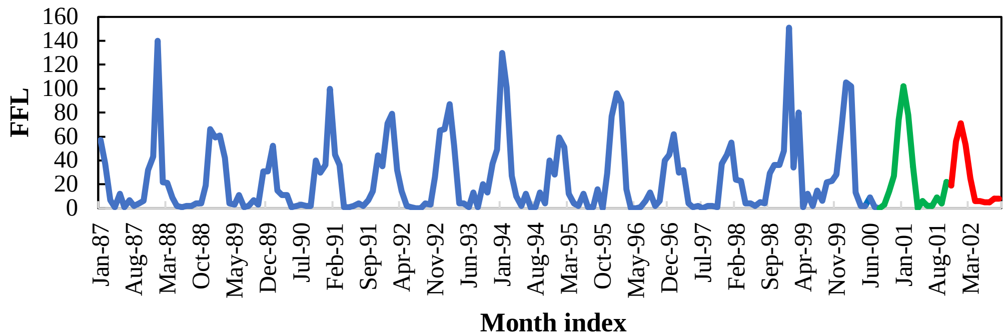
(a) Three months ahead of failure prediction of water mains in Quebec City, Quebec.



(b) Three months ahead of failure prediction of water mains in London City, Ontario.



(c) Nine months ahead of failure prediction of water mains in Quebec City, Quebec.



(d) Nine months ahead of failure prediction of water mains London City, Ontario.

Fig 6. Time series plot of the failure of water mains at selected locations (Quebec and London) (a), (b) three months and (c), (d) nine months ahead, respectively. FFQ: failure frequency in Quebec, FFL: failure frequency in London, blue line: actual data (Jan-87—Dec-00), green line: held-out data (Jan-00—Dec-01), red line: prediction (Jan-02—Sep-02).

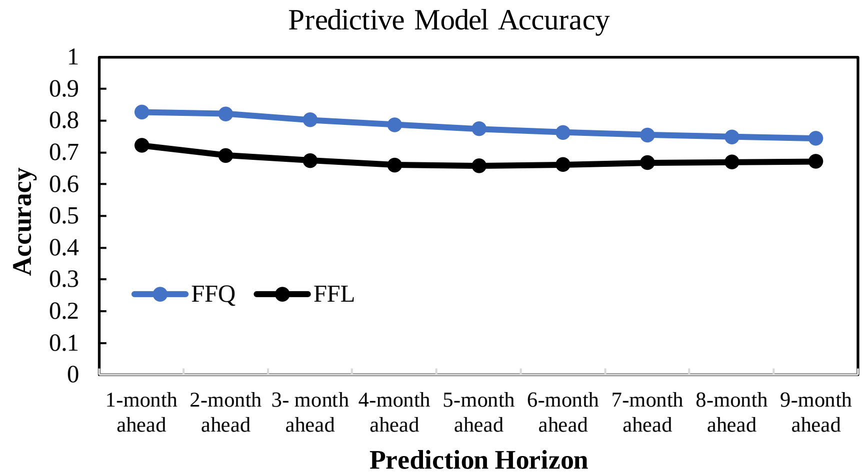
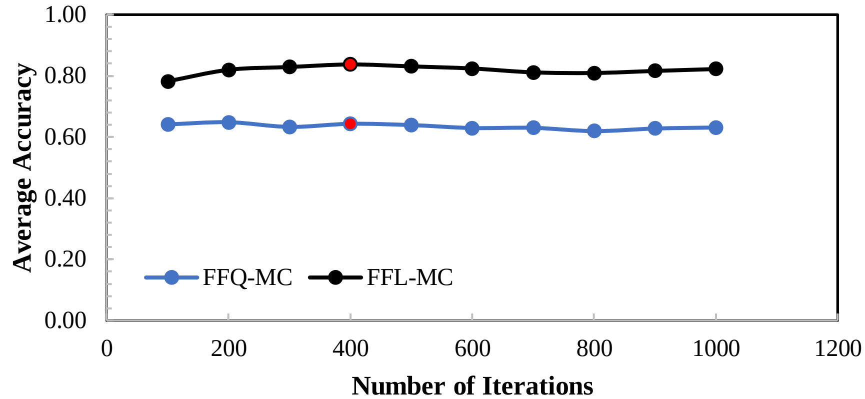
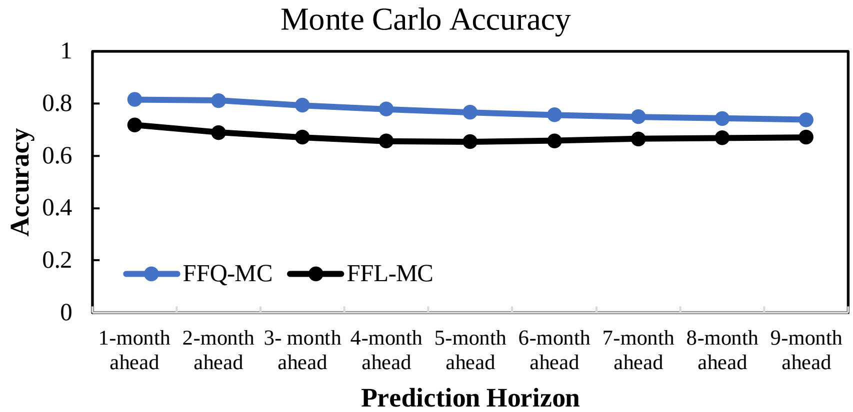


Fig 7. Average prediction accuracy of nine months ahead.

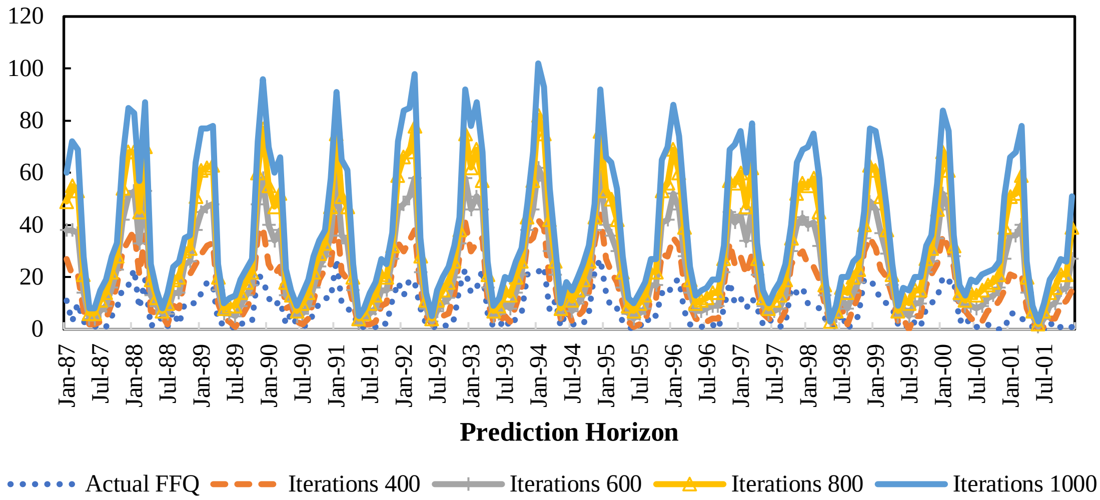


(a) MC simulations vs. number of iterations.

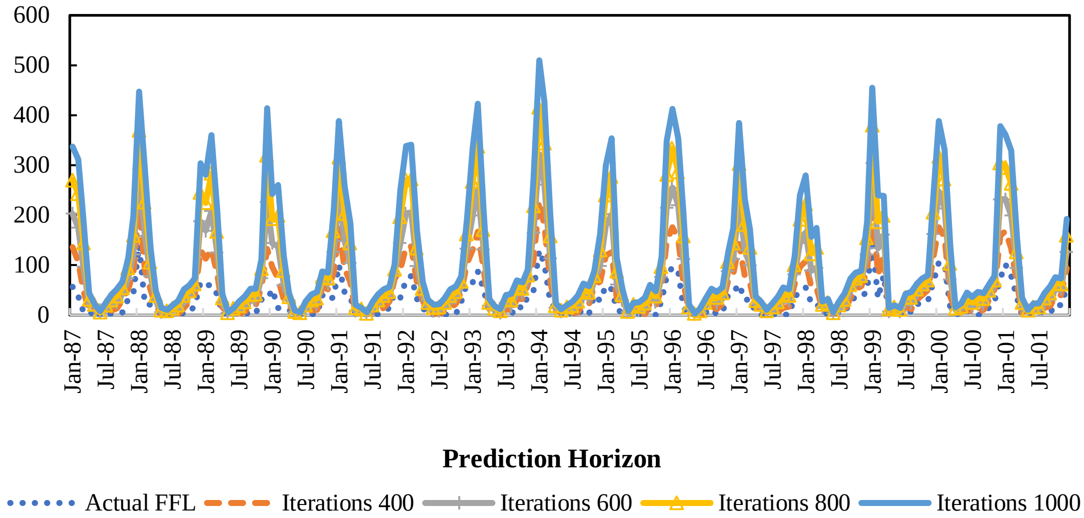


(b) Average accuracy of MC simulations (400 iterations).

Fig 8. (a) Average accuracy of MC simulations vs. number of iterations, and (b) average accuracy of MC simulations of nine months ahead at selected locations.



(a) Actual and predicted values of FFQ.



(b) Actual and predicted values of FFL.

Fig 9. Stacked charts of actual and the mean predicted failure frequency of water mains in (a) Quebec and (b) London using different iterations of MC simulations.