Environ Monit Assess (2021) 193:336
https://doi.org/10.1007/s10661-021-09068-0

Real-time contamination zoning in water distribution networks for contamination emergencies: a case study

Mohammad Reza Bazargan-Lari · Sharareh Taghipour · Mehdi Habibi

Received: 2 November 2020 / Accepted: 11 April 2021

Abstract Contamination of urban water distribution systems (WDS) is a critical issue due to the number of victims that might be impacted in a short period of time. Any effective rapid emergency response plan for reducing the number of sick people or deaths among those drinking the polluted water requires a reliable forecast of the water contamination zoning map (CZM). The water CZM is a visual representation of the spread of contamination at the time of contamination detection. This study presents a novel methodology based on the rough set theory (RST) for real-time forecasting of the CZM caused by simultaneous multi-point contamination injection in WDS. Our proposed methodology consists of (i) a Monte Carlo simulation model to capture the uncertainties in a multi-point deliberate contamination, (ii) a numerical simulation model for simulating pipe flow, and (iii) a rough set-based technique for real-time CZM for a WDS equipped with a set of monitoring stations. The proposed methodology can be used for any type of random contamination of WDSs as well as emergencies in deliberate contamination of water distribution networks.

Keywords Water distribution system · Rough set theory · Monitoring stations · EPANET · Contamination emergency management · Arsenic

M. R. Bazargan-Lari · S. Taghipour (✉)
Department of Mechanical and Industrial Engineering, Ryerson University, Toronto, Canada
e-mail: sharareh@Ryerson.ca

M. R. Bazargan-Lari
e-mail: bazargan@Ryerson.ca

M. Habibi
Department of Civil Engineering, East Tehran Branch, Islamic Azad University, Tehran, Iran
e-mail: m.habibi79eng@gmail.com

Introduction

Water distribution systems (WDS) are expensive lifeline infrastructures for providing drinking water to billions of people (Emamjomeh et al., 2020). Recent and historical events that are pointed in literature review section show that the WDS are vulnerable to deliberate contamination threats by criminals. Intentional contamination of WDS is becoming a critical issue due to the number of fatalities it causes in a short period of time, large economic consequences, and long-lasting society impacts. The WDS are in general complex infrastructures since they are usually large-scale, spatially extensive, and governed by nonlinear hydraulic equations. The complexity of the problem in deliberate contamination is considerable since the propagation of the released contaminant across WDS is flow dependent. Moreover, contamination's characteristics (e.g., contaminant injection time, concentration and location) in intentional contamination of WDS are highly stochastic which will further increase the complexity of contamination emergency management.

Most of WDS are equipped with a limited set of monitoring stations for detecting contaminant intrusion events. Since the collected information obtained from the monitoring stations at the detection time are limited to some specific points (monitoring stations), obtaining an overview of the layout of the contaminated zones, which is crucial for emergency managers and decision-makers, is a difficult task. So, it can be concluded that the current warning systems are necessary, but not sufficient for an effective response to accidental and intentional contaminations. This study particularly focuses on developing a powerful tool, namely water contamination zoning map (CZM), for determining the spread of contamination once a contaminated is detected. A water CZM is a type of emergency response decision aid tool for highlighting the areas, which are affected by contamination. The water CZM is assisting water utility operators in making better decisions for public health protection against contamination events. The main problem in determining real-time CZM is originating from the inherent uncertainty in the collected data once a contamination is detected. This uncertainty is mainly due to small sample size and incomplete information (inexact data). These are caused due to the limited number of monitoring stations, yet a fraction of which may report contamination at the detection time. This paper is an attempt to answer the question of how to overcome the vagueness and imprecisions encountered with in determining the real-time water contamination zoning map in a water distribution system at the time of contamination detection. In this paper, for the first time, we propose a rough set-based decision support tool for emergencies in deliberate contamination of water distribution networks. Our proposed tool is designed to provide reliable forecasts of water CZM using the data reported by a monitoring station at the detection time. The forecasts can be used to decide what to prioritize for rapid contaminant flushing from the contaminated drinking WDS. The developed decision support tool is applicable to any WDS equipped with monitoring stations.

The organization of the paper is as follows: in the "Literature review" section, the related works are studied and summarized, the methodology and its main components are discussed in detail in the "Methodology" section. The application of the proposed methodology in a real-world WDS is described in the "Case study" section, followed by the results and discussions in the "Results and discussions" section. Finally, the "Summary and conclusions" section describes the summary and conclusion of the study. Based on our findings and experience, some suggestions for future research are provided as well.

Literature review

Safety and quality of drinking water in WDS have been of particular interest to many researchers in the past few decades (Ahsan et al., 2017; Saravanan et al., 2018; Scheili et al., 2020; Sharif et al., 2017). Accidental or intentional chemical contamination in water distribution networks is an ongoing potential threat to WDS. Accidental contamination of WDS due to malfunction of chlorine stations, low quality of water source, pipe breaks, and leak repairs have been studied by different researchers (Cristo & Leopardi, 2008; Kansal et al., 2012; Kessler et al., 1998; Kumar et al., 1999; Rathi & Gupta, 2014; Xu et al., 2010). Moreover, WDS are vulnerable to deliberate contamination threats by criminals as well (Bazargan-Lari, 2014; de Winter et al., 2019; Khorshidi et al., 2019; Naserizade et al., 2018). Despite the frequency of accidental contamination is greater than the deliberate contamination, the consequence of the latter is usually more significant (de Winter et al., 2019). Recent and historical evens show that water infrastructures have been targeted since 3000 BC, and the number of attacks is significantly increasing over time (Birkett, 2017). Specifically, several intentional contaminations of drinking water have been reported (e.g., in Sri Lanka (2006), Kosovo (1998–1999), Turkey (1992), North Carolina (1977)), from which some occurred in the last decades (Forest et al., 2013). Nowadays, the WDS of most cities worldwide are vulnerable to the intentional contamination since the majority of municipal WDS contain unsupervised and accessible components such as fire hydrants. Deliberate contamination of WDS can be simply done through fire hydrants using mobile pressurized tanks. Propagation of the released contaminant across WDS continues through such uncertain behavior until it reaches to a monitoring station and is detected. The emergency response and disaster recovery are vital in contamination emergency management, especially for WDS serving water to a large number of people. Some investigators are focused on emergency warning systems. However, in a contaminant warning system, placing sensors at all locations in WDS is infeasible owing to its significant cost. Therefore, contamination detection using a limited set of sensors in such uncertain complex environment is the first challenge in responding to water contamination incidents. Several attempts have been carried out in recent years to overcome the problem of event detection by determining the optimal number and layout of the deployed sensors within WDS (Ciaponi et al., 2018, 2019; de Winter et al., 2019; Khorshidi et al., 2018; Yang & Boccelli, 2016, 2017). Hu et al. conducted a comprehensive literature survey on the sensor placement problem for contamination detection in WDS (Hu et al., 2018). Adendoja et al. (2019) presented a review of the existing methodologies for contamination identification in water distribution networks and outlined the complications of the existing sensor placement strategies (Adendoja et al., 2019). It is worth noting that previous investigations are mostly focused on WDS, which are exposed to single-point contamination to reduce the complexity of the developing surveillance, and monitoring programs.

Most of WDS are equipped with a limited set of monitoring stations for detecting contaminant intrusion events. While it is essential to determine the optimal locations of available monitoring stations, determining the source of contamination from the collected information obtained from the stations is another difficult task to be done right after contamination detection (Adedoja et al., 2018). Contaminant source location and its characteristics, required for controlling the contamination and cleanup operations, have been investigated (Guan et al., 2006; Khan & Banik, 2017; Liu et al., 2008, 2010; Perelman & Ostfeld, 2012; Rutkowski & Prokopiuk, 2018; Shang et al., 2007; Wagner & Neupauer, 2013; Wang & Harrison, 2012; Yan et al., 2016). Seth et al. (2016) and Adendoja et al. (2019) reviewed and compared the suggested source identification methods in the event of contamination in WDS and provided a set of recommendations on this matter. However, the proposed methods for source identification have limitations on the amount of data and information required and the processing time.

Very recently, Grbčić et al. (2020, 2021) and Lučin et al. (2021) proposed fast and reliable machine learning-based approaches for identifying the pollution source. Although identifying the pollution source is necessary, it is usually not enough for emergency managers and decision-makers who have to implement a rapid effective response. Real-time knowledge about the spread of the contamination and the layout of the contaminated zones at the detection time will lead to a more effective emergency response plan. In other words, it can be argued that having a reliable spread map of contamination at the detection instant is as important as identifying the contamination source.

Despite the importance of water CZM for an effective and rapid emergency response plan, the literature is surprisingly limited on determining a real-time reliable forecast of water CZM once contamination is detected. In this paper, we filled the gap of the literature in addressing the uncertainties in determining the CZM by presenting a rough set-based decision support tool for emergencies in deliberate contamination of water distribution networks. Rough set theory (RST) is an efficient tool for dealing with inexact and uncertain information in complex problems which has drawn the attention of numerous researchers from various fields (Dai et al., 2017; Szul & Kokoszka, 2020; Wei & Liang, 2019). The RST can be effectively used for deriving decision rules from data (Luo et al., 2020). Recently, the RST is utilized for evaluating the performance of supply chains (Rajesh, 2020), seasonal disaster predictions (Rajesh & Rajendran, 2019), forecasting energy consumptions in building (Szul & Kokoszka, 2020), and modeling uncertainties in multi-criteria decision making (Riaz et al., 2019). A survey of recent applications of RST has been provided by (Wei & Liang, 2019). However, there are very few applications of RST in water resource planning and management. In particular, rough set is used to evaluate water resource vulnerability (Chen et al., 2019), operation rules of reservoirs (Barbagallo et al., 2006), short-term rainfall forecasts (Sudha, 2017), and real-time flood forecasting (Wu et al., 2015).

Methodology

Water distribution simulation models are designed to simulate deterministic problems. However, contamination events of WDS are often accidental and highly stochastic. The shortcoming of a simulation model in incorporating the uncertainties in contamination can be addressed by including a linkage between a Monte Carlo simulation model and a flow and contaminant transport simulation model for simulating a finite set of possible events. Whereas, considering the required number of simulations, such approaches are in general time consuming and inefficient for real-time purposes. To overcome the problem of run time in incorporating the vagueness and imprecisions encountered with determining a real-time water CZM at the time of contamination detection, a combined methodology is considered in this paper. The flowchart of the proposed methodology for real-time contamination zoning in a WDS is presented in Fig. 1. The methodology consists of the following main steps: (a) gathering the required basic data and information; (b) generating feasible pollution scenarios using a Mont Carlo simulation model; (c) creating a water distribution network model for hydraulic and water quality analysis of WDS, and (d) applying the RST. The method is based on the analysis of water flow and contaminant transportation in WDS. In the first step, the required data for setting up, calibrating and verifying the numerical hydraulic simulation model of the water distribution network are collected. The required data and information can be categorized into two exclusive types: the information required to simulate the flow through the WDS (such as topological and physical characteristics of all components of a WDS as well as the users' demands), and the data required for simulating possible contamination threats. Examples of the data for simulating the flow include geometric characteristics of pipes network, pipes' roughness, lengths, diameters, nodal demand, and water level in reservoir. Examples of data for simulating contamination threats include the characteristics of substances that may be used in intentional contamination as well as the location of monitoring stations, check-points, and potential points for contamination injections. Checkpoints (CPs) are extra points distributed on the area under study for generating real-time contour lines of CZMs.

The proposed methodology requires a special step for including possible contamination events. Time, locations, and mass concentration of contamination intrusions are uncertain (random) variables that should be taken into account in intentional contamination studies. Therefore, any deterministic approach is not appropriate and can be misleading. Monte Carlo simulation (MCS) can be used as a classic method to incorporate the uncertainties of pollution injection. MCS is a well-established promising technique that has been widely used to quantify the inherent uncertainty of real-world engineering problems. MCS relies on repeated random sampling to mimic the randomness in a system. In order to simulate the mentioned randomness in deliberate contamination, in the second step, possible pollution scenarios are generated through a MCS for covering the whole range of contamination events. The inputs of this stage are stochastic characteristics of possible pollution injections, and the output is a set of possible pollution scenarios, which are not limited to single point contamination events. In the third stage, the analysis of WDS for different contamination scenarios is aided using a numerical hydraulic and water quality simulator.

In this stage, any of the contamination scenarios developed using the MCS model is numerically simulated by solving the governing flow and contaminant transport equations for determining the contamination concentration at monitoring stations (conditional attributes) and check points (decision attributes) once contamination is detected. The conditional and decision attributes providing the required finite set of observations (decision board) for learning the underlying relationship function between the observations in a form of decision table with attributes.

The Rough Set Theory (RST) is then used for real-time forecasting of the CZM caused by simultaneous multi-point contamination injection in WDS. The summarized decision table was randomly divided into train and test data sets. The train set is used for generating the set of RST-based If–Then rules. The contamination concentration of each checkpoint can be found using the trained RST model. The performance of the trained RST-based model for determining the contaminant concentration at check-points is then evaluated using classical statistical measures for the test set.

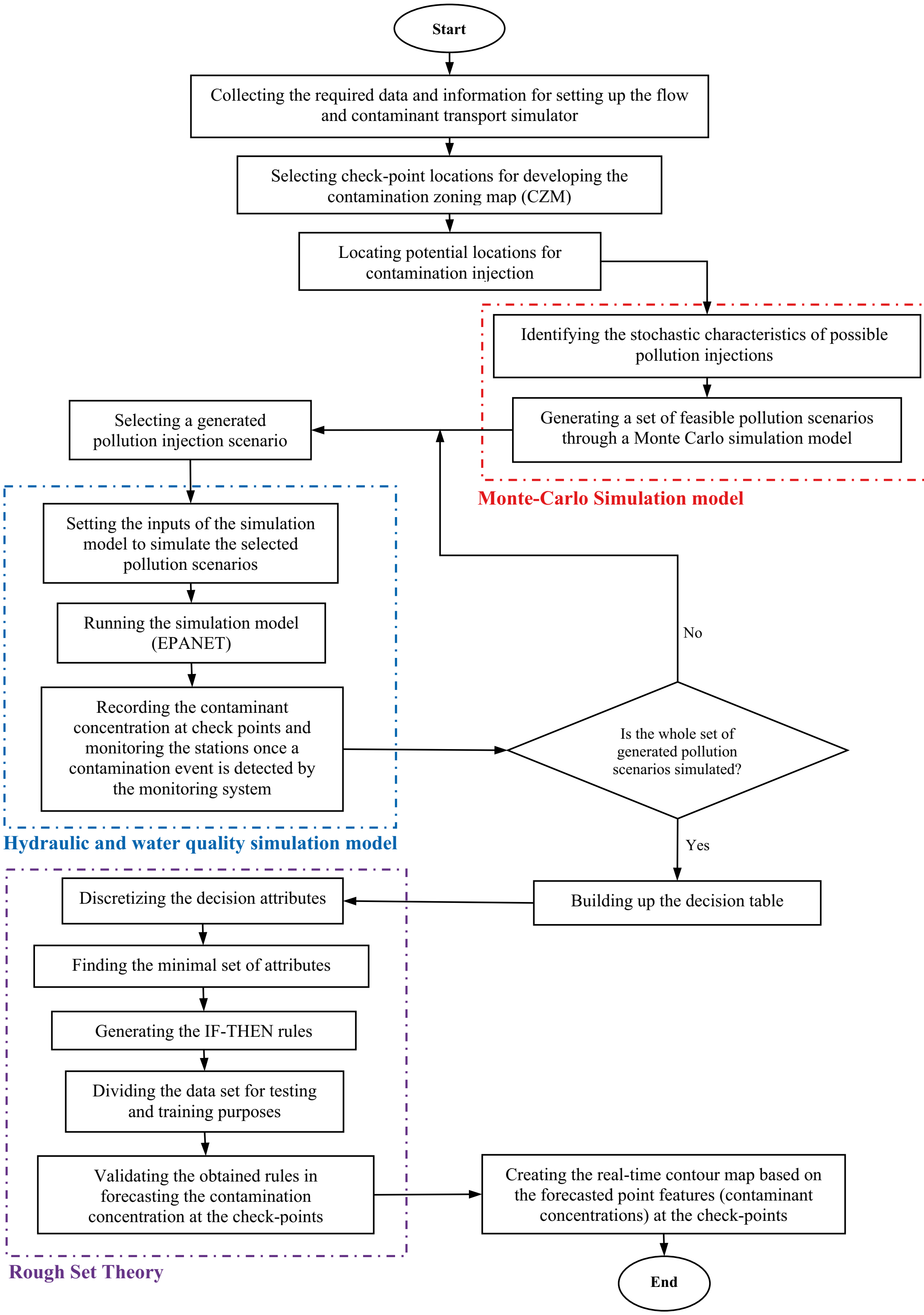

Fig. 1 Flowchart of the proposed methodology

The contaminant concentration amount at check-points provides the required values for contour interpolation for real-time CZM. Contour interpolation is a technique which uses the point with known contamination concentrations to estimate the values at other unknown points to forecast the CZM once a contamination event is detected.

In the following, the main components of the proposed framework are explained:

Hydraulic and water quality simulator

The forefront parts of the proposed methodology are forecasting the hydraulic response of WDS to contamination events, and the analysis of the hydraulic behavior of WDS for determining the nodal contaminant concentration over time. The simulation model predicts the flow directions and water quality contaminant concentration throughout the WDS for any generated pollution events. The required number of simulation runs depends on the number of contamination scenarios generated through the MCS. For simulating the contamination events, the total simulation period is divided into several time steps. Flow and contaminant transport in each time-step depend on the external conditions (external demand, contamination injection rate, etc.), current network parameters, concentration levels of contaminant and the flow rates calculated from the previous time step. The equations to be solved implicitly for calculating the fluid flow in WDS are the conservation of mass at each node, conservation of energy for each pipe, and pipe friction head loss with respect to the governing limitations on nodal pressure head and nodal water demands. Conservation of energy is a nonlinear relation between flow and head loss in each pipe that must be solved simultaneously for determining the nodal heads. A solver is required to simultaneously solve the governing nonlinear conservation equations. The solver must have the flexibility to be coupled with MCS model for performing all the required hydraulic and water quality simulations. EPANET version 2.0 (EPANET) which is an open-source public-domain simulation platform developed by the US EPA's drinking water division, is suitable to be used for modeling hydraulic and water quality behavior within WDS in the suggested framework. The governing equations and solution algorithms used in EPANET are fully described in (Rossman, 2000).

In this study, an R code is developed to call EPANET solver for any of the simulated contamination events to evaluate the contaminant concentration at monitoring and check-points once a contamination event is detected. Since at different contamination events, the physical characteristics (such as pipe diameter and roughness coefficients) of the WDS remain constant, the developed R code is used to adjust the inputs of the simulation model based on the spatial and temporal characteristics of any selected intentional contamination events (such as contaminant time and concentration, and the utilized injection node(s) for contaminating. EPANET then solves the hydraulic and water quality equations. Afterward, the developed R code is searching among the outputs of the quality simulation model to extract the concentration of the contaminant at the detection time at the considered index points (i.e., monitoring stations and check-points). The nodal contaminant concentration is recorded in a table called decision table (DT). The number of rows in DT is equal to the number of included scenarios in MCS, and the columns of DT are corresponding to the attributes (contamination concentration at monitoring and check-points).

RST

A set is a collection of elements that could be of two types: crisp and vague. In crisp sets, all elements can uniquely be determined, while in the vague set there is a kind of impreciseness (uncertainty) in distinguishing the elements with respect to the attributes. A rough set is a set of objects that cannot be uniquely represented by classes since there is uncertainty due to the vagueness of the classification. Drawing conclusion from vague data set is challenging. RST is a tool for dealing with problems with vagueness and imprecision. Using the RST, some additional information (knowledge/data) can be extracted about the elements of a vague data set.

In the proposed methodology, the output of the Monte Carlo-based simulations is an information system in the form of a table. This table is called the decision table which is a pair of nonempty finite set of objects, U, and nonempty finite set attributes, A. U is the universe and $A = \left\{a_1, a_2, \dots, a_m\right\}$ is an arbitrary set of attributes. The pair (U, A) form the required information system for real-time contamination zoning and the RST serves the process of discovering knowledge. In this study, the attributes are contamination concentrations (CCs) at monitoring stations, and the conditional attributes (decision attributes) are the CCs at the check-points. Table 1 shows a few rows of the obtained decision table for the case study in the following section. The table summarizes the outputs of 25 deliberate contamination events, randomly generated through a Monte Carlo simulation to present some basic concepts of RST. Rows of this table labeled 1 to 25 are called objects (entities). In this example, the universe $U = \{1,2,3,\ldots,25\}$ consists of 25 objects expressed by a set of attributes A. Each attribute $a \in A$ is the contamination concentration in mg/L at some specific nodes. For example, the value of CC at node number 1375 is denoted by CC_{N1375}, where $CC_{N1375} \in A$. As it can be seen in Table 1, the set of attributes A is divided into conditional attributes, $C = \{N_{1375}, N_{1459}, N_{1137}, N_{1072}, N_{1299}, N_{1878}, N_{1337}\}$, that are the CC at the monitoring nodes at the detection time. The decision attributes $D = \{N_{2070}, N_{2309}, N_{2343}, N_{2466}, N_{2376}, N_{2346}, N_{2243}, N_{2202}, N_{2408}, N_{2482}, N_{2071}, N_{2813}, N_{2414}, N_{2270}, N_{2370}, N_{2190}, N_{2247}, N_{2416}\}$ are the CC at the check-points for the case study described in the "Methodology" section. Obviously, for the attribute sets, $C \cup D = A$ and $C \cap D = \phi$. There is dependency between C and D that can be represented by a function: $f_a : U \rightarrow CC_a$ for any attribute $a \in A$. For instance, $f_{N1459}(1) = 0.02$ means that the CC at the first contamination scenario at monitoring station number 1459 is 0.02 mg/L. While, for this contamination event, the CC at the detection time at check-point number 2070 is 0.16 mg/L or $f_{N2070}(1) = 0.16$.

DTs may contain many objects having the same features. In order to reduce the table's size, among the objects having the same features, only one of them (namely indiscernible objects) is considered in RST. Indiscernibility (similar) relation, $IND(S)$, for any subset S of A is defined as (Ip et al., 2007; Zhang et al., 2020):

$$IND(S) = \{(i,j) \in U^2 | \forall a \in S, a(i) = a(j)\} \tag{1}$$

where $a(i)$ and $a(j)$ denote the values of attribute a for i^{th} and j^{th} elements, respectively. If $(i,j) \in IND(S)$, then i and j are indiscernible by attributes from S. For example, in Table 1, scenario numbers 1, 17, 18, and 20 are characterized by the same values of attribute N_{1375} and the value is 0.04. Similarly, scenario numbers 16, 23, and 24 are characterized by the same values of attribute N_{1375} and the value is 0.03. The equivalence relation induces a partition, namely a family of pairwise disjoint nonempty subsets of the universe whose union is U (Yao & She, 2016). Based on the equivalence relation for N_{1375}, we can define $IND(\{N_{1375}\}) = \{\{4,5,10,11,12,13,14\}, \{2,6,7,8,9,21\}, \{3,15,19,22,25\}, \{16,23,24\}, \{1,17,18,20\}\}$ as an indiscernibility of relation. Therefore, all the conditions reflected in the information system may not be necessary for making a decision. In this case, reducing some of the condition attributes is possible through the abovementioned approach. Indeed, the RST provides a sound mechanism in attribute reduction and has attracted the attention of different researchers worldwide (Abbas & Burney, 2016) mostly because of its potential for finding hidden patterns in data. Through the attribute reduction mechanism, the minimal set of attributes that preserves the information of interest can be found, and the decision rules can then be deduced. In this study, the contamination events revealing the same effects on the monitoring stations form the building blocks of elementary knowledge about the contamination events. These elementary building blocks are called elementary sets. Any union of elementary sets is called a crisp set. Other sets are called rough or imprecise set (Pawlak, 1998). For defining a rough set, we need to define upper and lower approximation sets. For a subset X of the universe that may be expressed as a union of some elementary sets of equivalence relation R, the upper approximation set, $\overline{R}X$, is a set of objects which possibly belong to the target set and the $\overline{R}X$ have nonempty intersection with X:

$$\overline{R}X = \cup\{Y \in U|R : Y \cap X \neq \phi\} \tag{2}$$

The lower approximation set, $\underline{R}X$, is the set of elementary sets completely included in X:

$$\underline{R}X = \cup\{Y \in U|R : Y \subseteq X\} \tag{3}$$

$(\overline{R}X, \underline{R}X)$ is called a rough set of Y with respect to R (Wei & Liang, 2019). The rules induced from its lower approximation are certainly valid and such rules are called certain since they are surely belonging to a given decision class. While the rules induced from the upper approximation are possibly valid

Table 1 A sample of decision-making table including contamination concentration data

Scenario	Conditional attributes							Decision-making attributes				
	Contamination concentration at monitoring stations(mg/L)							Contamination concentration at check-points (mg/L)				
ID	N1375	N1459	N1137	N1072	N1299	N1878	N1337	N2070	N2309	N2343	N2466	N2376
1	0.04	0.02	0	0	0	0	0	0.16	0.02	0.01	0.02	0.01
2	0.01	0.01	0	0	0	0	0	0.05	0.01	0.01	0.02	0.01
3	0.02	0.03	0	0	0	0	0	0.06	0.01	0.02	0.05	0
4	0	0.01	0	0	0	0	0	0.02	0	0	0.01	0
5	0	0.03	0.01	0	0	0	0	0	0	0.02	0.04	0.02
6	0.01	0.03	0	0	0	0	0	0.02	0	0.02	0.04	0.02
7	0.01	0.03	0.01	0	0	0	0	0.03	0	0.02	0.05	0.02
8	0.01	0.03	0.01	0	0	0	0	0.03	0	0.02	0.04	0.02
9	0.01	0.01	0	0	0	0	0	0.03	0	0	0.01	0
10	0	0.03	0.01	0	0.15	0	0	0	0	0.02	0.04	0.02
11	0	0.02	0.01	0	0	0	0.03	0	0	0.01	0.03	0.01
12	0	0.02	0.01	0	0	0	0.02	0	0	0.01	0.03	0.01
13	0	0.02	0.01	0	0	0	0.03	0	0	0.01	0.03	0.01
14	0	0.02	0.01	0	0	0	0.03	0	0	0.01	0.03	0.01
15	0.02	0	0	0	0	0	0	0.08	0.01	0	0.01	0
16	0.03	0.02	0.01	0	0	0	0	0.1	0.01	0.01	0.03	0.01
17	0.04	0.01	0	0	0	0	0	0.13	0.02	0.01	0.02	0
18	0.04	0.01	0	0	0	0	0	0.15	0.02	0.01	0.02	0.01
19	0.02	0.04	0	0	0	0	0	0.08	0.01	0.03	0	0
20	0.04	0.04	0	0	0	0	0	0.15	0.02	0.03	0	0
21	0.01	0	0	0.01	0	0	0	0.02	0	0	0	0
22	0.02	0	0	0.01	0.01	0	0	0.07	0.01	0	0	0
23	0.03	0	0	0.01	0.01	0	0	0.11	0.01	0.03	0	0
24	0.03	0	0	0	0	0	0	0.13	0.02	0	0	0
25	0.02	0	0.21	0	0	0.02	0.02	0.07	0.01	0	0	0

Table 1 (continued)

Scenario	Decision-making attributes												
	Contamination concentration at check-points (mg/L)												
ID	N2346	N2243	N2202	N2408	N2482	N2071	N2813	N2414	N2270	N2370	N2190	N2247	N2416
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0.01	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0.01	0	0	0	0	0	0	0	0	0	0	0	0
8	0.01	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0.01	0	0	0	0	0	0	0	0	0.11	0	0	0
11	0.01	0	0	0	0	0	0.01	0.17	0	0	0.11	0	0
12	0.01	0	0	0	0	0	0.01	0.08	0	0	0.05	0	0
13	0.01	0	0	0	0	0	0.05	0.08	0	0	0.01	0	0
14	0.01	0	0	0	0	0	0.01	0.08	0	0	0.05	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0.08	0.02
20	0	0	0	0	0	0	0	0	0	0	0	0.09	0.02
21	0	0.03	0.01	0.01	0	0	0	0	0	0	0	0	0
22	0	0.03	0.01	0.01	0.01	0	0	0	0	0	0	0	0
23	0	0	0.01	0.01	0.02	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0.17	0	0	0	0
25	0.19	0	0	0	0	0.01	0.01	0.01	0	0	0	0	0

(Gogoi et al., 2013). In the RST, the imperfect knowledge is expressed by a boundary region of a set, $BR(X)$, where the elements cannot be certainly categorized as belong or not belong to X. This special boundary, namely rough set, is the difference between the $\overline{R}X$ and $\underline{R}X$:

$$BR(X) = \overline{R}X - \underline{R}X \tag{4}$$

Obviously, for a crisp set, $BR(X)$ would be an empty set. As an instance, scenarios 7 and 8 in Table 1 are in boundary for $N2466$, because these two have the same conditional attributes. Therefore, the decision class of $N2466$ is rough, since the boundary region is already not empty. Classifying the data in terms of equivalence classes is challenging because the set of objects are in general indistinguishable in accordance to the attributes, and they cannot be uniquely represented by equivalent classes.

Degree of uncertainty in sample set X can be evaluated using a ratio index namely roughness coefficient (R_C):

$$R_C = \frac{N_{\underline{R}X}}{N_{\overline{R}X}} \tag{5}$$

where $N_{\underline{R}X}$ and $N_{\overline{R}X}$ are the number of elements in lower and upper approximation sets, respectively. R_C is a dimensionless index varying in the range [0,1]. This index plays an important role in extracting knowledge since if $R_C = 1$, X is exactly dependent on R. While if R and X are totally independent, $R_C = 0$. For $0 < R_C < 1$, where X is vague (rough dependence) with respect to R, a roughness membership function is used to quantify the degree of overlap. The roughness membership function is the conditional probability that an object belongs to the set X with respect to the attributes. Using these dependencies measures, a set of simplified decision rules in the form of If–Then will be developed.

Equations (5) and (6) are the decision rules for determining the relationships between a collection of equivalence classes (Rajesh & Rajendran, 2019).

$$Des(X_i) \rightarrow Des(\gamma)\ if\ P(\gamma|X_i) \geq \theta \tag{6}$$

$$Des(X_i) \rightarrow Des(\hat{\gamma})\ if\ P(\gamma|X_i) \leq 1 - \theta \tag{7}$$

where X_i is equivalence class of a reduced set of conditional attributes. The $Des(X_i)$ denotes the description of the equivalence class X_i, and γ is a partition of the finite set of objects, U. For a threshold value, θ, in the range (0.5,1], if $P(\gamma|X_i) \geq \theta$, the object belongs to γ with a specific certainty. However, for $\hat{\gamma}$, which is used for representing the elements that are in U but not in γ, the $P(\gamma|X_i)$ is less than or equal to $(1 - \theta)$. The rules that can be removed without causing any inconsistency will be removed through a procedure called rule generalization (Rajesh & Rajendran, 2019). In this paper, the obtained generalized rules are then utilized for forecasting contaminant concentration using the data recorded at the monitoring stations.

In summary, the RST is a promising tool to combat the situations where the objects having equal description are assigned to the same classes using rule induction (Abed-Elmdoust & Kerachian, 2012). Rule induction is a method of finding hidden patterns in data, even from incomplete data sets using the RST. Then, the tabular data could be condensed into a set of If–Then rules (decision attributes). Decision attributes constitutes the training set in RST through the training procedure. The obtained rules are representation of the underlying pattern in the data and are capable to be utilized for prediction. A decision rule is applicable to a specific object, if its description matches the conditional part of the rule.

Case study

The proposed methodology has been applied to a zone of district 19 in the southern part of Tehran's WDS with domestic demands. The location of the studied WDS in Tehran metropolitan area is shown in Fig. 2. This WDS serves a zone of about 3.68 km^2 with elevations ranging between 1093.63 and 1106.07 m above mean sea level. Figure 3 shows the layout and the components of the studied WDS. As shown in this figure, the network is fed by Mehrabad reservoir. The minimum pipe's diameter is 4 inches (100 mm), and the maximum is 28 inches (700 mm). It approximately provides water to a population of 112,400 inhabitants. The population is growing by 0.8% each year. The studied WDS is configured with nodes (elevation, demand and pressure), pipes (length, diameter and roughness), and physical data for tanks, and pressure reducing valves (PRVs).

Calibration of the WDS is performed by adjusting the pipe's roughness coefficients to decrease the differences between the predicted and observed values (nodal pressure, pipe flows, tank water level) for given demand loading conditions and operation policies. The pipes' roughness varies in the range of [95, 110]. The minimum pressure requirement for each node is 26 m. These constraints are applied to all nodes in the designing stage.

Table 2 summarizes the main input data used for simulating the WDS.

This WDS serves a mix of residential, commercial, and industrial users such as shopping centers, hospitals, schools, hotels, etc. According to the Iranian Guidelines (Design criteria of urban and rural water supply and distribution systems, 2013), the average daily demand in the community is 240 L per day, and the maximum and daily fluctuations are set to 2.4 and 0.6 times of the average daily demand, respectively. In the study area, it is required to provide 15 L/min of fire flow for a minimum of 4 h through the fire hydrants installed. They are mostly installed on pipes with diameter above 100 mm. Figures 4 and 5 show the nodal pressure and water velocity in average water demand and peak hourly demand, respectively. As shown in these figures, the developed simulation model confirms the reliability of the studied WDS in terms of sufficient nodal pressure required for supplying the demand and velocity for both normal and peak hourly demand conditions.

The location of the monitoring stations, check-points and considered possible injection nodes for the contamination injection in the studied WDS are shown in Fig. 6. The individual nodes are identified by 4-digit node numbers where the first digit indicates the type of nodes. Through a field survey, 16 possible injection locations are flagged. These locations are the exposed fire hydrants that are accessible for intentional contamination (for example by arsenic) through a simple pump (represented by 4-digit node numbers started with 3 in Fig. 6). Since arsenic does not make changes to the color and smell of drinking water, it has been considered in the previous studies for determining the optimal layout of the monitoring stations for detecting deliberate contamination events (Bazargan-Lari, 2014, 2018; Khorshidi et al., 2018; Naserizade et al., 2018). The considered monitoring stations are represented by 4-digit node numbers starting with 1 in Fig. 6. In this study, a set of possible contamination events by arsenic is generated through a MCS. The locations, time, and the duration of the contamination injection and arsenic concentration are considered as uncertain input variables. The initial contamination concentration of the potential injection nodes is zero, and during the contamination events, it increases in the range of [0.01 mg/L to 1 mg/L]. EPANET is then utilized for simulating the flow and arsenic transport for each of the generated contamination events. It may also be noted that EPANET is capable of simulating water flow and arsenic transport and has been employed by several researchers (Khorshidi et al., 2019; Klosterman et al., 2010; Naserizade et al., 2018). A set of check-points are selected for monitoring the contaminant concentrations, and for contour accuracy validation. The check-points are identified by 4-digit node numbers that starting with 2 in Fig. 6. As it can be seen in Fig. 6, the check-points are spread out over the study area. A set interface code is developed for extracting the pollution concentration at the check-points for each contamination event.

Results and discussions

The proposed methodology for real-time CZM using RST requires a numerical simulation model, EPANET, to predict the hydraulic behavior of WDS and a Monte Carlo-based simulation. The Monte Carlo simulation is used to tackle the uncertainties in the locations, time, and the duration of contamination injection and arsenic concentration. Deliberate contamination events are assumed to occur in the nighttime period (i.e., from 6:00 pm to 6:00 am) when the surveillance accuracy decreases due to darkness. During the procedure, a list of 340,000 Monte Carlo-based discrete water contamination events are randomly generated to cover possible single-point and multi-point injection of contaminations among the 16 potential locations shown in Fig. 6. In this study, the maximum injection duration is set to an hour owing to the frequency of police patrol during night.

Since the water distribution system must be simulated for each of the developed contamination events, a computer program is developed to call the generated contamination events, simulate the contamination movement, and store the contamination state (the concentration of the contaminant at all WDS nodes) of WDS once it is detected by the monitoring stations.

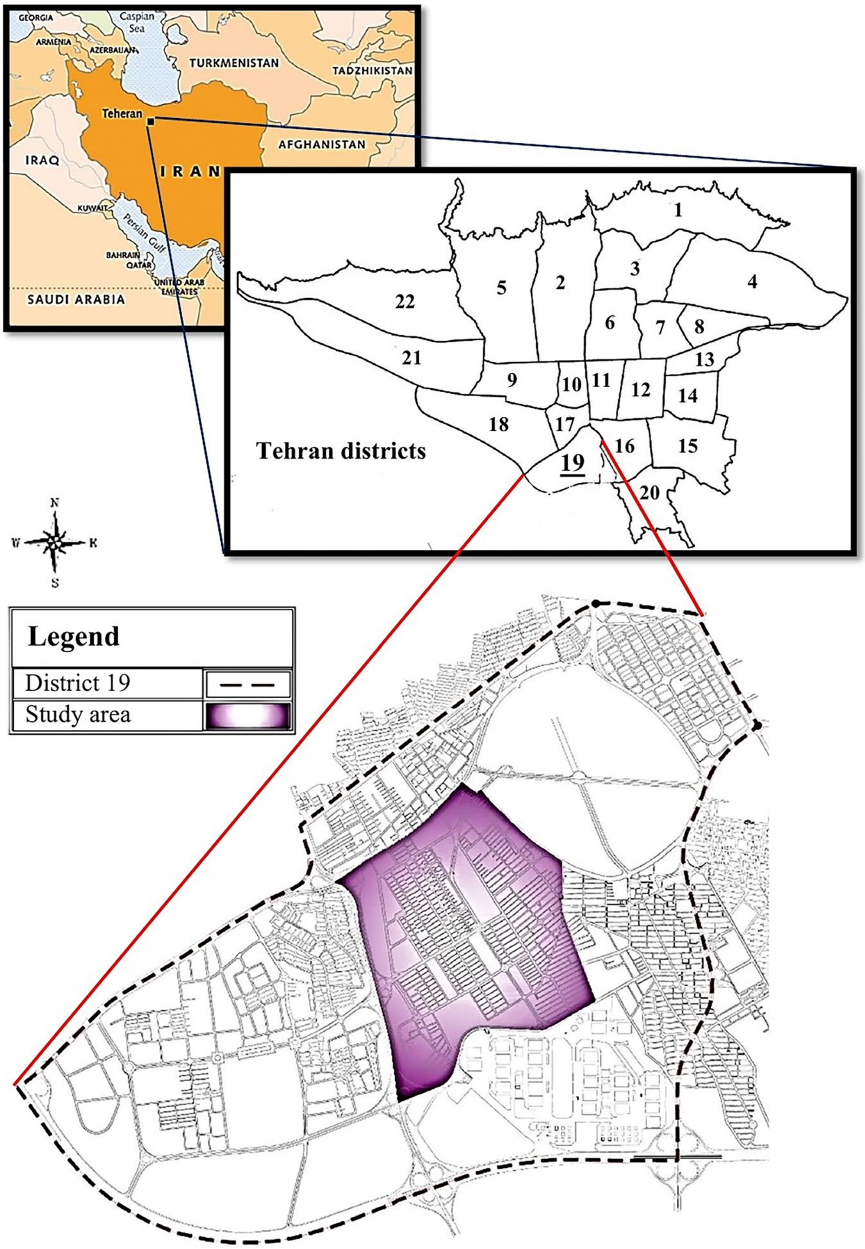

Fig. 2 Scheme of the studied WDS

For health concerns, the allowable arsenic level must be strictly regulated. There is a standard, for the total allowable arsenic level in drinking water. According to the guidelines of the WHO (World Health Organization) standards for drinking water, the maximum allowable arsenic level in drinking water is 0.01 mg/L (Hema & Sundararajan, 2019; Neppolian et al., 2010; USEPA, 2001). Therefore, in this study, the detection threshold is set to the maximum allowable arsenic level.

For the studied WDS, simulating a contaminant scenario requires 7 s on a 2.67 GHz Intel i7 dual-core CPU. We have simulated all the generated contamination events. The simulation of the whole scenarios takes around 28 days. The outputs of the simulations are then stored in a table, called decision table, for applying the RST.

A part (80%) of the simulated contamination events are utilized for training. In the context of

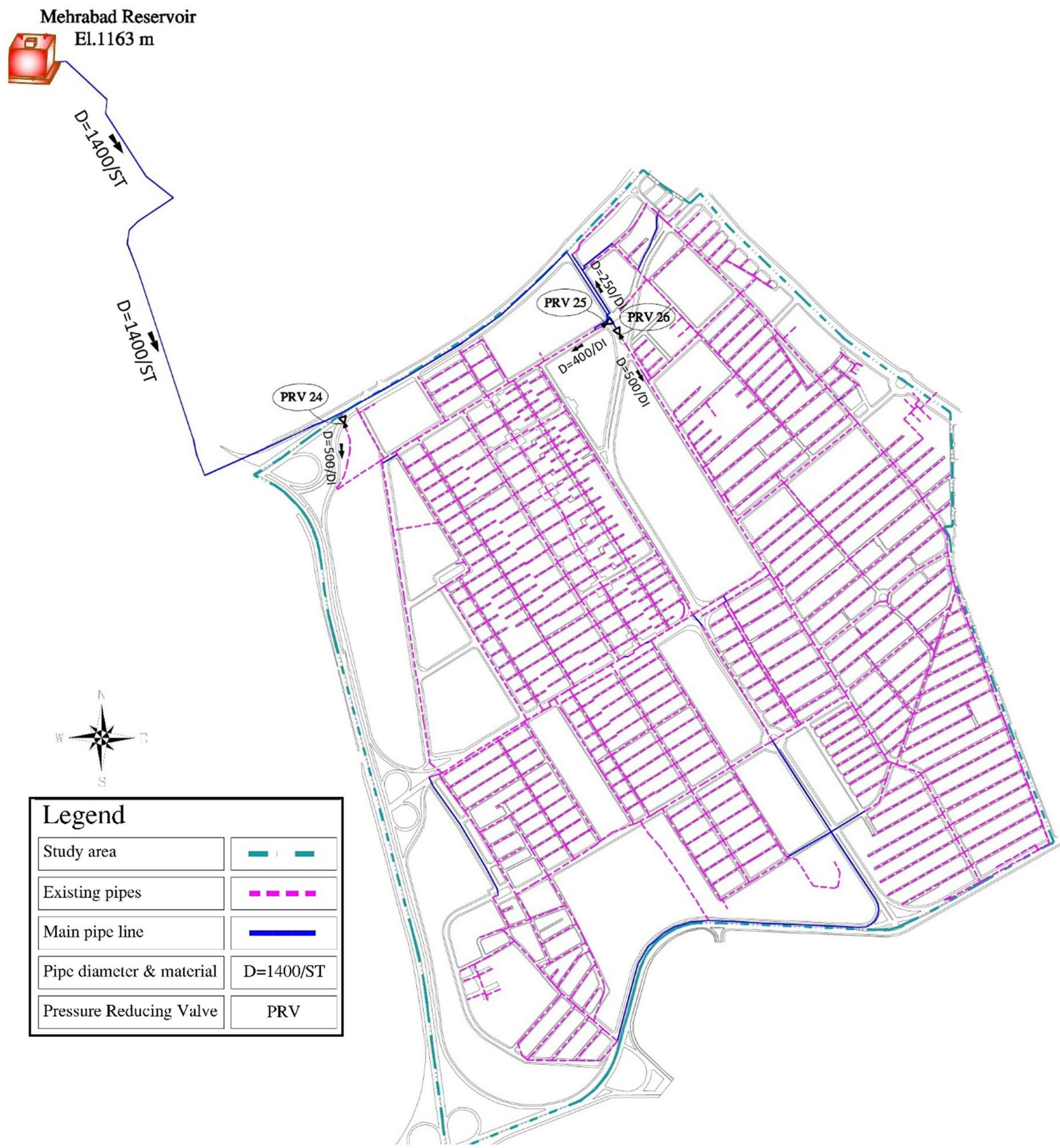

Fig. 3 Layout and components of the studied WDS

RST, training data set is in the form of decision-making table, partially shown in Table 1. In order to reduce the cardinality of the value sets in the obtained decision table, discretization technique is used to replace the conditional real-valued attributes with intervals. Removing redundant attributes is a crucial operation to reduce the information system, and determine a subset of the attributes that are sufficient to describe the decision attributes. In this study, a genetic algorithm suggested by (Li & Cercone, 2005) is utilized for dealing with the problem of attribute selection. The obtained data set (decision table) is then imported to ROSETTA (ROSETTA, 2017) for determining the minimal set

Table 2 Values of the main parameters used in the quantity and quality simulation model of the studied WDS

Parameter	Value
Pipe diameter (m)	0.1–0.7
Total pipe length (km)	74
Required water storage-tank volume (m3)	20,000
Average per capita water demand (liters per capita per day)	240
Daily peaking factor	1.5
Hourly peaking factor	1.6
Maximum allowable flow velocity (m/s)	2
Minimum pressure needed at each node (m)	26
Maximum allowable pressure at each node (m)	50
Minimum water level in tank (m)	0.2
Maximum water level in tank (m)	4.5
Number of nodes in the water distribution network	851
Number of storage tanks in the water distribution network	1
Number of PRVs in the water distribution network	3

of attributes, and generating the rules. Rough Set Toolkit for Analysis of Data (ROSETTA) is a rule-based modeling package developed by (Komorowski et al., 2002) for implementing the RST using a heuristic generic algorithm (Emam et al., 2017; Lei et al., 2019; Sulaiman et al., 2016; Tripathy, 2016). The suggested methodology extracts the governing rules. The discovered pattern in this paper consisted of 2811 rules which are reduced to a set of 149 rules through the rule reduction procedure. A

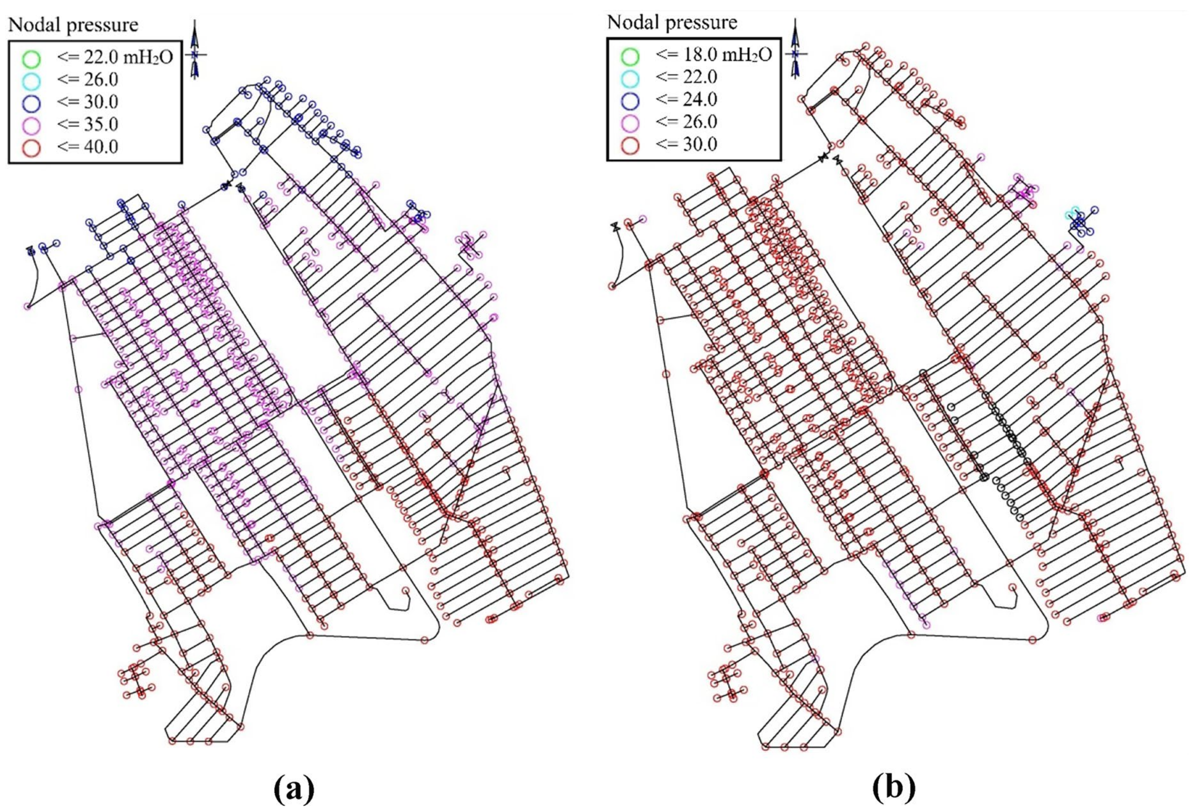

Fig. 4 Nodal pressure in average water demand **a** and peak hourly demand **b**

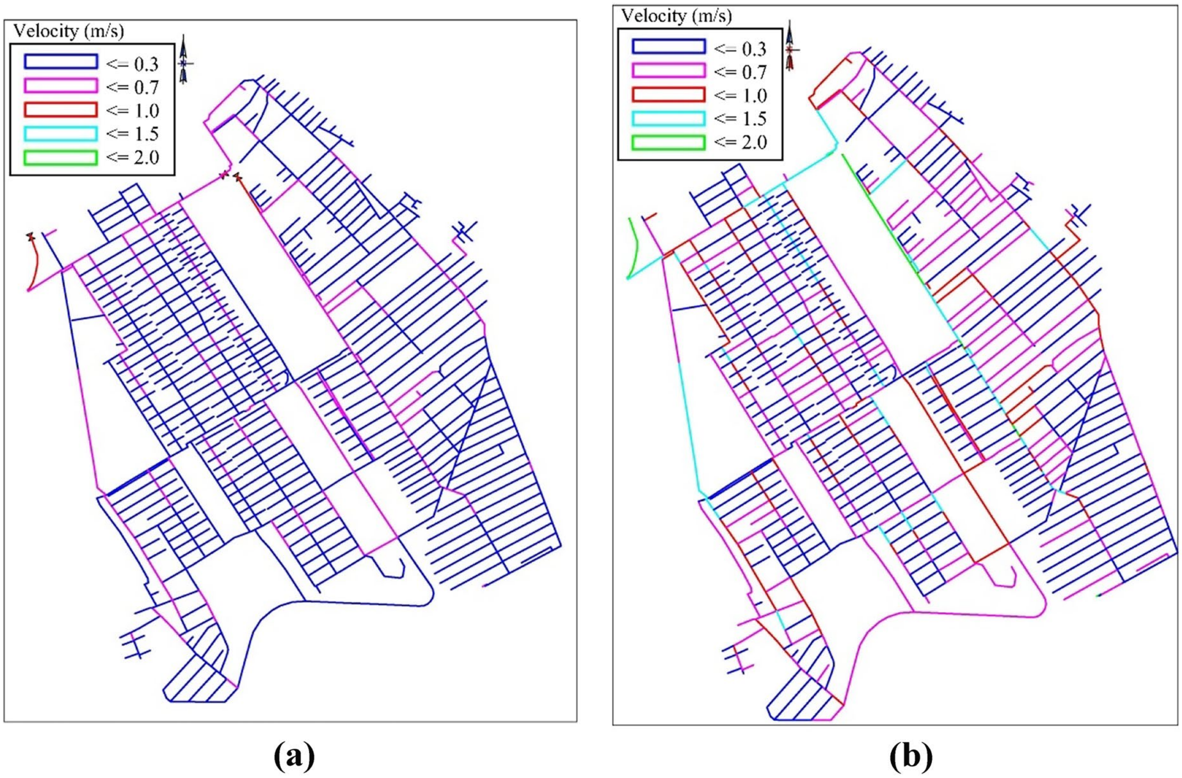

Fig. 5 Pipe flow velocity in average water demand **a** and peak hourly demand **b**

few examples of the rules discovered for the contamination concentration at check-points are listed in Table 3.

Influential parameters, obtained thorough the RST, are the contaminant concentration at monitoring stations once contamination is detected. As an instance, scenario number 125 is one of the considered scenarios, in which node numbers 3393 and 3497 are utilized for contaminating the WDS at 4:00 am and 5:00 am, respectively. In this contamination event, which was excluded in the training stage, the arsenic concentration in node 3393 and 3497 were 0.58 and 0.71 mg/L, respectively. This arsenic contamination event is simulated using EPANET. Two of the monitoring stations (node numbers 1299 and 1878) which were the closest monitoring stations to the contaminated points detected the event. For determining the real-time CZM, the operator should only be aware about the contamination concentrations at the monitoring nodes. The trained RST then provides the contaminant concentration at the check-points to form the CZM. Figure 7 shows the outputs of the simulation model (contaminant concentration at the check-points) and the corresponding rough set-based predictions for scenario number 125. For some check-points such as 2071 and 2370, the model forecasted the exact values. However, for some others, the predictions encountered some sort of errors that are clearly shown in Fig. 7. The calculated R-square value for this scenario is 0.947, which indicates the good performance of the trained rough set. The predicted values at check-points are then used for contour interpolation, and thereafter, the water CZM. The obtained rough set-based CZM corresponding to the specific contamination instance (scenario number 125) is presented in Fig. 8. This figure reveals that the proposed rough set-based model is able to accurately forecast the CZM in real-time. As it can be seen, the errors shown in Fig. 7 for scenario number 125 do not have a considerable influence on the resulted CZM.

The correlation coefficient, root mean square error (RMSE), mean absolute error (MAE), and Bias are well-known statistical measures, employed to

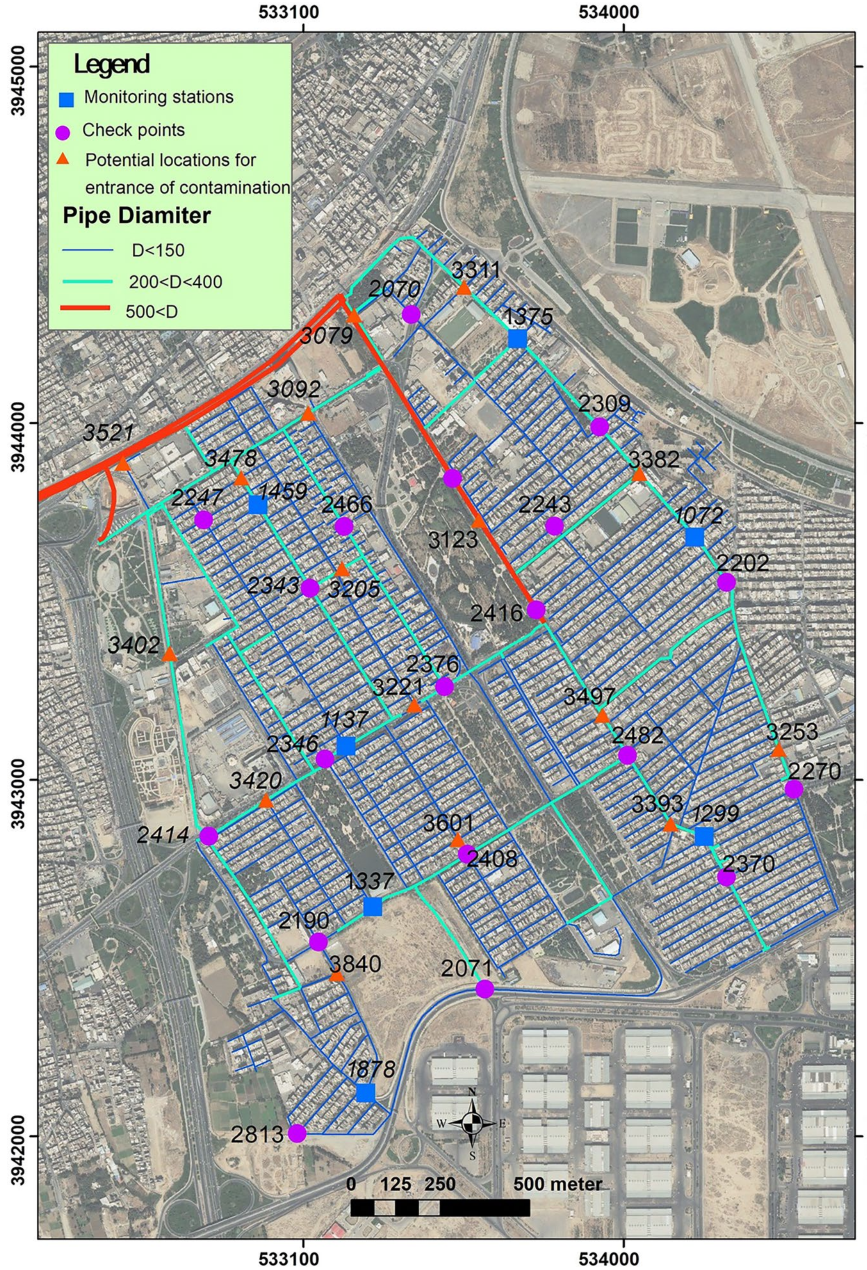

Fig. 6 Location of the monitoring stations, checkpoints and considered possible injection nodes for the contamination injection in the WDS

evaluate the performance of the trained RST in the validation process. The values of the statistical performance measures are presented in Table 4. The predictive performance was remarkable with the average correlation coefficient of 0.969 and RMSE of 0.0058.

Moreover, the visual representation of the RST based CZM provides an overview of the performance and applicability of the suggested methodology. To illustrate the capabilities of the trained RST in determining the CZM, 8 simulated contaminated events were randomly selected among the test set. The upper counter maps in Fig. 9, show the simulation-based CZMs, and the lower ones show the corresponding forecasted RST based CZMs. The comparison of the contour maps shown in Fig. 9 reveals that, in general, the proposed model is deserved to be utilized as a powerful tool for real-time application. More specifically, as it can be seen in Fig. 9, the RST

Table 3 Some of the rules discovered using the rough set theory for forecasting the contamination concentration at check-points

Rule ID	Obtained If–Then Rules for forecasting contamination concentrations at check-points
Rule 1	If $0.04 \leq CC_{N1375} < 0.07$ and $0.02 \leq CC_{N1459} < 0.04$ and $CC_{N1137} < 0.01$ and $CC_{N1072} < 0.01$ and $CC_{N1299} < 0.01$ and $CC_{N1878} < 0.01$ and $CC_{N1337} < 0.01$ then CC_{N2070}=0.16 mg/L, CC_{N2309}=0.02 mg/L, CC_{N2343}=0.01 mg/L, CC_{N2466}=0.02 mg/L, CC_{N2376}=0.01 mg/L, CC_{N2346}=0.00 mg/L, CC_{N2343}=0.00 mg/L, CC_{N2202}=0.00 mg/L, CC_{N2408}=0.00 mg/L, CC_{N2482}=0.00 mg/L, CC_{N2071}=0.00 mg/L, CC_{N2813}=0.00 mg/L, CC_{N2414}=0.00 mg/L, CC_{N2270}=0.00 mg/L, CC_{N2370}=0.00 mg/L, CC_{N2190}=0.00 mg/L, CC_{N2247}=0.00 mg/L, CC_{N2416}=0.00 mg/L
Rule 2	If $0.02 \leq CC_{N1375} < 0.03$ and $0.02 \leq CC_{N1459} < 0.04$ and $CC_{N1137} < 0.01$ and $CC_{N1072} < 0.01$ and $CC_{N1299} < 0.01$ and $CC_{N1878} < 0.01$ and $CC_{N1337} < 0.01$ then CC_{N2070}=0.06 mg/L, CC_{N2309}=0.01 mg/L, CC_{N2343}=0.02 mg/L, CC_{N2466}=0.05 mg/L, CC_{N2376}=0.00 mg/L, CC_{N2346}=0.00 mg/L, CC_{N2343}=0.00 mg/L, CC_{N2202}=0.00 mg/L, CC_{N2408}=0.00 mg/L, CC_{N2482}=0.00 mg/L, CC_{N2071}=0.00 mg/L, CC_{N2813}=0.00 mg/L, CC_{N2414}=0.00 mg/L, CC_{N2270}=0.00 mg/L, CC_{N2370}=0.00 mg/L, CC_{N2190}=0.00 mg/L, CC_{N2247}=0.05 mg/L, CC_{N2416}=0.01 mg/L
Rule 3	If $0.04 \leq CC_{N1375} < 0.05$ and $CC_{N1459} < 0.01$ and $CC_{N1137} < 0.01$ and $CC_{N1072} < 0.01$ and $CC_{N1299} < 0.01$ and $CC_{N1878} < 0.03$ and $CC_{N1337} < 0.03$ then CC_{N2070}=0.13 mg/L, CC_{N2309}=0.00 mg/L, CC_{N2343}=0.02 mg/L, CC_{N2466}=0.04 mg/L, CC_{N2376}=0.01 mg/L, CC_{N2346}=0.01 mg/L, CC_{N2343}=0.00 mg/L, CC_{N2202}=0.01 mg/L, CC_{N2408}=0.00 mg/L, CC_{N2482}=0.00 mg/L, CC_{N2071}=0.00 mg/L, CC_{N2813}=0.00 mg/L, CC_{N2414}=0.00 mg/L, CC_{N2270}=0.00 mg/L, CC_{N2370}=0.01 mg/L, CC_{N2190}=0.00 mg/L, CC_{N2247}=0.00 mg/L, CC_{N2416}=0.00 mg/L

base model is able to accurately highlight the critical zones very well. Determining critical zones plays an important role in any rapid effective response to the detected contamination events.

Summary and conclusions

In a water distribution system (WDS), once the monitoring stations detect a contamination event, rapid emergency response planning must be started as quickly as possible to control the number of sick people or deaths among those consumed the polluted water. Real-time water contamination zoning maps (CZMs) are crucial information required for any effective emergency response action. They are simple and easy-to-understand tools for assisting emergency managers and operators, especially in determining the catastrophic area, quickly identifying the affected people, making better decisions for crisis management, aiding/evacuating the affected people, and finally flushing the WDS. The main concern is to accurately determine the hazardous zones of a drinking WDS once an intentional contamination event is

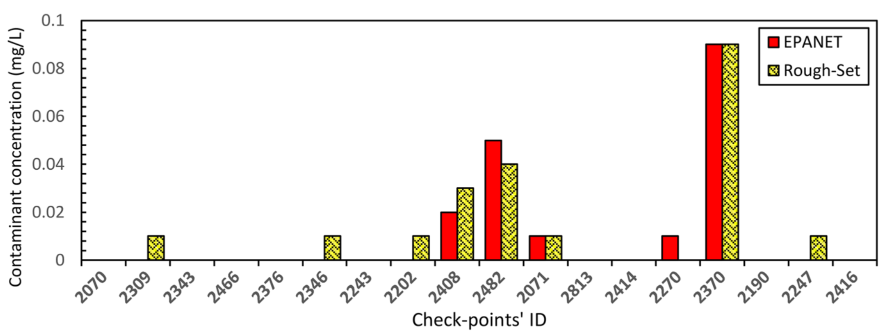

Fig. 7 The arsenic concentration at the check-points for a specific contamination event (scenario number 125)

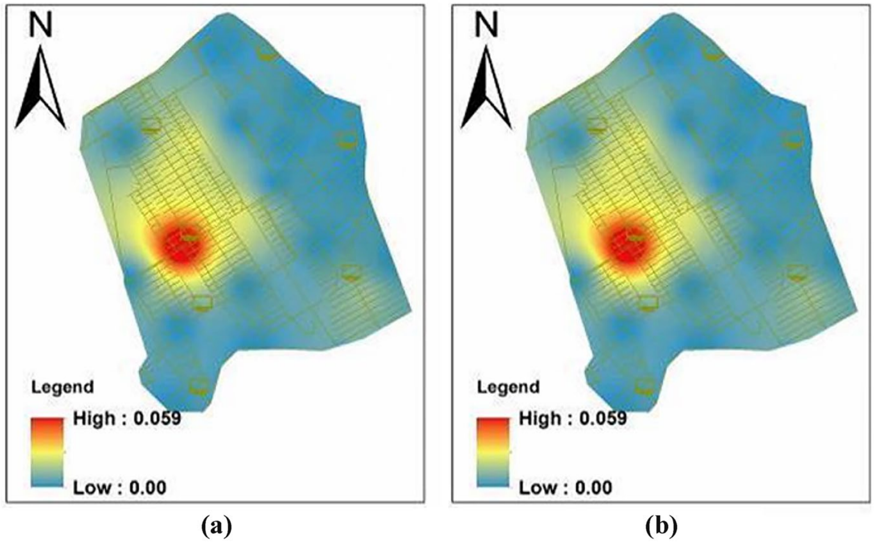

Fig. 8 The simulated (**a**) and forecasted (**b**) CZM corresponding to scenario number 125

detected. This is a complex task and the literature is surprisingly limited on suggesting a practical methodology. In this study, a new approach is developed to provide emergency managers with a real-time decision support tool. In contrast, the goal herein is to determine a real-time decision support algorithm that can be used for determining the CZM.

The methodology presented in this paper is based on a flow and contaminant transport simulation, EPANET, and a viable approach, namely rough set theory (RST). RST is used as an artificial intelligence-based technique for extracting meaningful knowledge and making predictions for determining the CZM. Unpredictable time-varying system behavior has been taken into account through a Monte Carlo simulation. The widely used water distribution simulator, EPANET, allows the simulation of the generated possible contamination events for estimating the hydraulic and water quality behavior within the WDS. The knowledge gained by experience (the simulation of the contamination events) is then analyzed by RST to deal with the inherent uncertainty and vagueness of the system. In this paper, it was demonstrated that RST can be used as a novel, reliable and powerful tool for predicting CZM in deliberate contamination events as types of random contamination of drinking water distribution systems. The suggested rough set rule-based model presented herein is capable of determining the required values for contour interpolation for real-time CZM. The results show that the predictive performance of the proposed methodology is remarkable, and RST is an efficient tool for handling and extracting useful information from a large amount of data obtained through a Monte Carlo-based simulation. Therefore, the proposed methodology is generic and can be applied to any WDS and any hazardous contaminant. This work has been applied to a real WDS in southern part of Tehran's metropolitan area in Iran, with 112,400 inhabitants. Overall, the study shows that RST performs well in determining the CZM and offers benefits in terms of providing an effective rapid emergency response plan. Such findings have following implications to theory and practice:

Table 4 Values of the statistical performance measures

Parameter	Value
Correlation coefficient	0.969
Mean absolute error (mg/L)	0.0034
Root mean square error (mg/L)	0.0058
Bias (mg/L)	0.027

Theoretical contributions

Although different contamination source identification models have been developed by different investigators, the CZM suggested in this paper can be considered as a new generation of contamination source identification, which is more effective when there is suspicion of a contamination event. In theory, this study proposed and tested a new combined methodology. The novel contribution of the current study is to explore the capabilities of rough set theory in monitoring drinking water distribution networks.

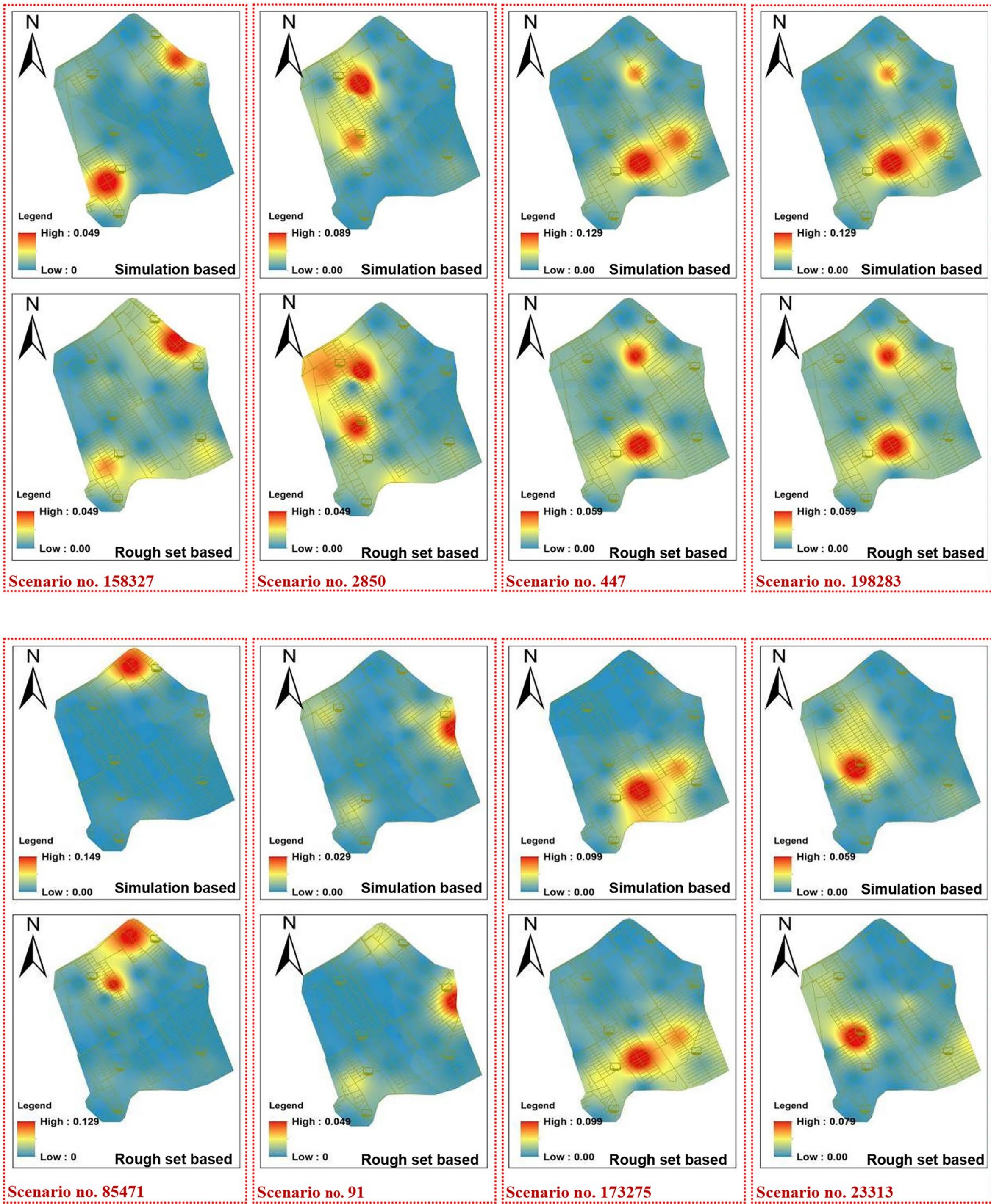

Fig. 9 The simulation-based CZM (upper figures), and the corresponding forecasted RST-based CZM (lower figures) for 16 randomly selected contamination events

Moreover, different from previous investigations that are mostly focused on single-point contamination to reduce the complexity of the developing surveillance, and monitoring programs, this paper demonstrates the capabilities of the proposed methodology in including multi-point contaminations.

Practical implications

Based upon the obtained results for the studied real-world water distribution network, the proposed model is robust in generating CZM and has the capability to be applied effectively to any WDS. Decision-makers can utilize the generated CZM for making proper decisions for implementing rapid and effective responses to the detected intentional or accidental contamination disasters of WDS. As an instance, the CZMs can be utilized for planning some special safety restrictions on the critical zones once a contamination disaster is detected, or implementing an effective rapid contaminant flushing program for the polluted WDS.

Limitations

The results of applying the proposed methodology to a real-world case study revealed no significant difference between the forecasted and the simulated water CZMs. Nevertheless, the location and the number of selected check-points as well as the number of simulated contamination scenarios included in the Monte Carlo simulation can affect the accuracy of the generated CZM. Selecting more check-points and more contamination scenarios is practically feasible. However, a larger set of check-points and contamination scenarios does not necessarily lead to a more accurate CZM. Future works may include optimizing the number and the pattern of the check-points as well as the uncertainties in the demand of consumers.

Acknowledgements We acknowledge the contribution of Afagh Company in providing us with the data and information for simulating the studied WDS. This research is financially supported by the Ontario Ministry of Economic Development, Job Creation and Trade's Early Researcher Awards.

Data availability The data used to support the findings of this study are included in the article.

References

Abbas, Z., & Burney, A. (2016). A survey of software packages used for rough set analysis. *Journal of Computer and Communications, 4*(9), 10–18

Abed-Elmdoust, A., & Kerachian, R. (2012). Wave height prediction using the rough set theory. *Ocean Engineering, 54*, 244–250

Adedoja, O., Hamam, Y., Khalaf, B., & Sadiku, R. (2018). Towards development of an optimization model to identify contamination source in a water distribution network. *Water, 10*(5), 579

Adendoja, O. S., Hamam, Y., Khalaf, B., & Sadiku, R. (2019). Sensor placement stratefies for contamination identification in water distribution networks: A review. *WIT Transactions on Ecology and the Environment, 229*, 79–90

Ahsan, M. S., Akber, M. A., Islam, M. A., Kabir, M. P., & Hoque, M. I. (2017). Monitoring bacterial contamination of piped water supply in rural coastal Bangladesh. *Environmental monitoring and assessment, 189*(11), 597

Barbagallo, S., Consoli, S., Pappalardo, N., Greco, S., & Zimbone, S. M. (2006). Discovering reservoir operating rules by a rough set approach. *Water Resources Management, 20*(1), 19–36

Bazargan-Lari, M. R. (2014). An evidential reasoning approach to optimal monitoring of drinking water distribution systems for detecting deliberate contamination events. *Journal of cleaner production, 78*, 1–14

Bazargan-Lari, M. R. (2018). Real-time response to contamination emergencies of urban water networks. *Iranian Journal of Science and Technology, Transactions of Civil Engineering, 42*(1), 73–83

Birkett, D. (2017). Water critical infrastructure security and its dependencies. *Contemporary Voices: St Andrews Journal of International Relations, 8*(2)

Chen, Y., Feng, Y., Zhang, F., & Wang, L. (2019). Assessing water resources vulnerability by using a rough set cloud model: A case study of the Huai River Basin, China. *Entropy, 21*(1), 14

Ciaponi, C., Creaco, E., Di Nardo, A., Di Natale, M., Giudicianni, C., Musmarra, D., & Santonastaso, G. F. (2019). Reducing impacts of contamination in water distribution networks: A combined strategy based on network partitioning and installation of water quality sensors. *Water, 11*(6), 1315

Ciaponi, C., Creaco, E., Nardo, A. D., Natale, M. D., Giudicianni, C., Musmarra, D., & Santonastaso, G. F. (2018). *Optimal sensor placement in a partitioned water distribution network for the water protection from contamination.* Paper presented at the Multidisciplinary Digital Publishing Institute Proceedings

Cristo, C. D., & Leopardi, A. (2008). Pollution source identification of accidental contamination in water distribution networks. *Journal of Water Resources Planning and Management, 134*(2), 197–202

Dai, J., Wei, B., Zhang, X., & Zhang, Q. (2017). Uncertainty measurement for incomplete interval-valued information systems based on α-weak similarity. *Knowledge-Based Systems, 136*, 159–171

de Winter, C., Palleti, V. R., Worm, D., & Kooij, R. (2019). Optimal placement of imperfect water quality sensors in water distribution networks. *Computers & Chemical Engineering, 121*, 200–211

Design-criteria-of-urban-and-rural-water-supply-and-distribution-systems. (2013). *Design criteria of urban and rural water supply and distribution systems: Publication No. 117–3*. Retrieved from Tehran, Iran

Emam, O., Farhan, M., & Abohany, A. (2017). Faults repairing analysis using rough sets after implementation of labor force redistribution algorithm: A case study in Telecom Egypt. *Information Sciences Letter, 6*, 39–48

Emamjomeh, H., Jazany, R. A., Kayhani, H., Hajirasouliha, I., & Bazargan-Lari, M. R. (2020). Reliability of water distribution networks subjected to seismic hazard: Application of an improved entropy function. *Reliability Engineering & System Safety,* 106828. https://doi.org/10.1016/j.ress.2020.106828

Forest, J. J., Howard, R. D., & Sheehan, M. A. (2013). *Weapons of mass destruction and terrorism*: McGraw-Hill New York

Gogoi, P., Bhattacharyya, D. K., & Kalita, J. K. (2013). A rough set–based effective rule generation method for classification with an application in intrusion detection. *International Journal of Security and Networks, 8*(2), 61–71

Grbčić, L., Kranjčević, L., & Družeta, S. (2021). Machine learning and simulation-optimization coupling for water distribution network contamination source detection. *Sensors, 21*(4), 1157

Grbčić, L., Lučin, I., Kranjčević, L., & Družeta, S. (2020). Water supply network pollution source identification by random forest algorithm. *Journal of Hydroinformatics, 22*(6), 1521–1535

Guan, J., Aral, M. M., Maslia, M. L., & Grayman, W. M. (2006). Identification of contaminant sources in water distribution systems using simulation–optimization method: Case study. *Journal of Water Resources Planning and Management, 132*(4), 252–262

Hema, R., & Sundararajan, M. (2019). *Development, Analysis and Evaluation of Arsenic (III) Sensor.* Paper presented at the 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT).

Hu, C., Li, M., Zeng, D., & Guo, S. (2018). A survey on sensor placement for contamination detection in water distribution systems. *Wireless Networks, 24*(2), 647–661

Ip, W., Hu, B., Wong, H., & Xia, J. (2007). Applications of rough set theory to river environment quality evaluation in China. *Water Resources, 34*(4), 459–470

Kansal, M. L., Dorji, T., Chandniha, S. K., & Tyagi, A. (2012). Identification of optimal monitoring locations to detect accidental contaminations. In World Environmental and Water Resources Congress 2012: Crossing Boundaries (pp. 758–776)

Kessler, A., Ostfeld, A., & Sinai, G. (1998). Detecting accidental contaminations in municipal water networks. *Journal of Water Resources Planning and Management, 124*(4), 192–198

Khan, M. A. I., & Banik, B. K. (2017). Contamination source characterization in water distribution network. *Global Science and Technology Journal, 5*(1), 44–55

Khorshidi, M. S., Nikoo, M. R., Ebrahimi, E., & Sadegh, M. (2019). A robust decision support leader-follower framework for design of contamination warning system in water distribution network. *Journal of cleaner production, 214*, 666–673

Khorshidi, M. S., Nikoo, M. R., & Sadegh, M. (2018). Optimal and objective placement of sensors in water distribution systems using information theory. *Water research, 143*, 218–228

Klosterman, S., Murray, R., Szabo, J., Hall, J., & Uber, J. (2010). Modeling and simultation of arsenate fate and transport in a distribution system simulator *Water Distribution Systems Analysis 2010* (pp. 655–669)

Komorowski, J., Øhrn, A., & Skowron, A. (2002). *Case studies: public domain, multiple mining tasks systems: ROSETTA rough sets.* Paper presented at the Handbook of data mining and knowledge discovery

Lei, L., Chen, W., Xue, Y., & Liu, W. (2019). A comprehensive evaluation method for indoor air quality of buildings based on rough sets and a wavelet neural network. *Building and Environment, 162*, 106296

Li, J., & Cercone, N. (2005). *A rough set based model to rank the importance of association rules.* Paper presented at the International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing

Liu, L., Ranjithan, S. R., & Mahinthakumar, G. (2010). Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure. *Journal of Water Resources Planning and Management, 137*(2), 183–192

Liu, L., Zechman, E. M., Brill, J., E Downey, Mahinthakumar, G., Ranjithan, S., & Uber, J. (2008). *Adaptive contamination source identification in water distribution systems using an evolutionary algorithm-based dynamic optimization procedure.* Paper presented at the Water Distribution Systems Analysis Symposium 2006

Lučin, I., Grbčić, L., Čarija, Z., & Kranjčević, L. (2021). Machine-learning classification of a number of contaminant sources in an urban water network. *Sensors, 21*(1), 245

Luo, J., Fujita, H., Yao, Y., & Qin, K. (2020). On modeling similarity and three-way decision under incomplete information in rough set theory. *Knowledge-Based Systems, 191*, 105251

Naserizade, S. S., Nikoo, M. R., & Montaseri, H. (2018). A risk-based multi-objective model for optimal placement of sensors in water distribution system. *Journal of hydrology, 557*, 147–159

Neppolian, B., Doronila, A., & Ashokkumar, M. (2010). Sonochemical oxidation of arsenic (III) to arsenic (V) using potassium peroxydisulfate as an oxidizing agent. *Water research, 44*(12), 3687–3695

Pawlak, Z. (1998). Rough set theory and its applications to data analysis. *Cybernetics & Systems, 29*(7), 661–688

Perelman, L., & Ostfeld, A. (2012). Bayesian networks for source intrusion detection. *Journal of Water Resources Planning and Management, 139*(4), 426–432

Rajesh, R. (2020). Sustainability performance predictions in supply chains: grey and rough set theoretical approaches. *Annals of Operations Research,* 1–30. https://doi.org/10.1007/s10479-020-03835-x

Rajesh, R., & Rajendran, C. (2019). Grey-and rough-set-based seasonal disaster predictions: an analysis of flood data in India. *Natural Hazards, 97*(1), 395–435

Rathi, S., & Gupta, R. (2014). Monitoring stations in water distribution systems to detect contamination events. *ISH Journal of Hydraulic Engineering, 20*(2), 142–150

Riaz, M., Davvaz, B., Firdous, A., & Fakhar, A. (2019). Novel concepts of soft rough set topology with applications. *Journal of Intelligent & Fuzzy Systems, 36*(4), 3579–3590

ROSETTA. (2017). A rough set toolkit for analysis of data. *V1.4.41, Knowledge SystemsGroup at NTNU, Norway, and the Logic Group at Warsaw University, Poland,* <http://bioinf.icm.uu.se/rosetta/downloads.php> *Accessed 03.01.17*

Rossman, L. A. (2000). *EPANET 2: users manual*. Cincinnati, OH.

Kumar, A., Kansal, M. L., Arora, G., Ostfeld, A., & Kessler, A. (1999). Detecting accidental contaminations in municipal water networks. *Journal of Water Resources Planning and Management, 125*(5), 308–310

Rutkowski, T., & Prokopiuk, F. (2018). Identification of the contamination source location in the drinking water distribution system based on the neural network classifier. *IFAC-PapersOnLine, 51*(24), 15–22

Saravanan, K., Anusuya, E., & Kumar, R. (2018). Real-time water quality monitoring using Internet of Things in SCADA. *Environmental monitoring and assessment, 190*(9), 556

Scheili, A., Delpla, I., & Rodriguez, M. J. (2020). Development of a drinking water quality index based on a participatory procedure using mixed multicriteria methods. *Environmental Monitoring and Assessment, 192*(8), 1–14

Seth, A., Klise, K. A., Siirola, J. D., Haxton, T., & Laird, C. D. (2016). Testing contamination source identification methods for water distribution networks. *Journal of Water Resources Planning and Management, 142*(4), 04016001. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000619

Shang, F., Uber, J., & Murray, R. (2007). *Uncertainty of contamination source characterization within water distribution systems.* Paper presented at the World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat

Sharif, M. N., Farahat, A., Haider, H., Al-Zahrani, M. A., Rodriguez, M. J., & Sadiq, R. (2017). Risk-based framework for optimizing residual chlorine in large water distribution systems. *Environmental Monitoring and Assessment, 189*(7), 1–19

Sudha, M. (2017). Intelligent decision support system based on rough set and fuzzy logic approach for efficacious precipitation forecast. *Decision Science Letters, 6*(1), 95–106

Sulaiman, S., Rahim, N. A. A., & Pranolo, A. (2016). Generated rules for AIDS and e-learning classifier using rough set approach. *International Journal of Advances in Intelligent Informatics, 2*(2), 103–122

Szul, T., & Kokoszka, S. (2020). Application of rough set theory (rst) to forecast energy consumption in buildings undergoing thermal modernization. *Energies, 13*(6), 1309

Tripathy, B. (2016). Application of rough set based models in medical diagnosis *Handbook of Research on Computational Intelligence Applications in Bioinformatics* (pp. 144–168): IGI Global.

USEPA. (2001). National primary drinking water regulations; arsenic and clarifications to compliance and new source contaminants monitoring; final rule. *Federal Register, Part VIII, 40 CFR Parts 9, 141, and 142. US Government Printing Office, Washington, DC*

Wagner, D., & Neupauer, R. (2013). *Probabilistic contaminant source identification in water distribution systems with incomplete mixing at pipe junctions.* Paper presented at the World Environmental and Water Resources Congress 2013: Showcasing the Future

Wang, H., & Harrison, K. W. (2012). Improving efficiency of the Bayesian approach to water distribution contaminant source characterization with support vector regression. *Journal of Water Resources Planning and Management, 140*(1), 3–11

Wei, W., & Liang, J. (2019). Information fusion in rough set theory: An overview. *Information Fusion, 48*, 107–118

Wu, H. Q., Huang, Q., Xu, W., & Xi, S. F. (2015). Application of K-means cluster and rough set in classified real-time flood forecasting. *Advanced Materials Research, 1092–1093*, 734–741

Xu, J., Johnson, M. P., Fischbeck, P. S., Small, M. J., & VanBriesen, J. M. (2010). Robust placement of sensors in dynamic water distribution systems. *European Journal of operational research, 202*(3), 707–716

Yan, X., Zhao, J., Hu, C., & Wu, Q. (2016). Contaminant source identification in water distribution network based on hybrid encoding. *Journal of Computational Methods in Sciences and Engineering, 16*(2), 379–390

Yang, X., & Boccelli, D. L. (2016). Model-based event detection for contaminant warning systems. *Journal of Water Resources Planning and Management, 142*(11), 04016048

Yang, X., & Boccelli, D. L. (2017). Integrated systemwide model-based event detection algorithm. *Journal of Water Resources Planning and Management, 143*(8), 04017047

Yao, Y., & She, Y. (2016). Rough set models in multigranulation spaces. *Information Sciences, 327*, 40–56

Zhang, C., Li, D., Kang, X., Song, D., Sangaiah, A. K., & Broumi, S. (2020). Neutrosophic fusion of rough set theory: An overview. *Computers in Industry, 115*, 103117