Environmental Modelling & Software 77 (2016) 71–80

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Spatial event classification using simulated water quality data

Nurit Oliker, Ziv Ohar, Avi Ostfeld*

Faculty of Civil and Environmental Engineering, Technion — Israel Institute of Technology, Haifa 32000, Israel

ARTICLE INFO

Article history:
Received 23 February 2015
Received in revised form
4 October 2015
Accepted 20 November 2015
Available online 14 December 2015

Keywords:
Water distribution systems
Multiple sensors
Water security
Event detection
Contamination simulation
EPANET-MSX

ABSTRACT

This study deals with the integration of contamination simulations and a spatial event detection model. The simulation of contaminant intrusion includes detailed chemical-specific reactions within a multi-species water quality model. This set-up generates a scenario of contaminant distribution and produces a continuous multiple sensor stations database. Three organophosphates pesticides, Chlorpyrifos, Malathion, and Parathion, are modeled as possible contaminants. The event detection model comprises both local and spatial data analysis. The local model applies a previously developed single-sensor event detection model with a higher alert threshold that reduces false alarm rates. The spatial model considers upstream sensor datasets which are examined for their uniqueness and mutual resemblance in a sliding time window. The model utilizes outlier detection, data analyses, and network hydraulics for the detection of suspicious spatial trends. The proposed algorithm is capable of detecting events with low contamination signatures and spatial influence. Two case studies are explored and compared to the single sensor model. The proposed methodology resulted in a lower number of false alarms compared to the previous single sensor event detection modeling approach.

1. Introduction

Securing water infrastructure is surely an eminent task, being vital for population welfare. A water distribution system (WDS) is a vulnerable infrastructure, comprising numerous elements which can be easily penetrated. Thus, the threat of either accidental or deliberate contaminant intrusion to the WDS arouses growing concern worldwide. Accordingly, water security has become one of the most investigated areas in the field of water distribution systems analysis during the last two decades. Substantial resources are being invested in the development of contamination warning systems.

General water quality parameters, which are routinely measured by utilities, may indicate the presence of a contaminant in the network (Hall et al., 2007). Parameters like total organic carbon (TOC), oxidation reduction potential (ORP), chloride, free chlorine and conductivity have shown significant reaction to most groups of pollutants (Panguluri et al., 2009). These findings triggered the development of automated event detection system (EDS) that utilize online measurements of water quality parameters for the detection of contamination events. Meriam-Webster defines "event" as "a noteworthy happening", in the contexts of this work an "event" is considered as a measurable deviation from typical quality behavior of one or more water quality parameters, duo to a contamination intrusion.

The application of EDS in water distribution networks entails two main tasks: (1) allocating sensors around the network, and (2) analyzing the measured data. The problem of sensor placement in WDS is one of the most explored problems in the field during the last two decades and was approached by different optimization techniques. Hart and Murray (2010) reviewed over 90 studies that include different optimization strategies for the sensor placement problem. Nearly all models have used the unrealistic assumption of perfect sensor, which assumes that a sensor detects any measured contaminant regardless of its concentration and influence. A slightly more realistic assumption is that the sensor detects any contamination above a certain concentration or threshold (Hart and Murray, 2010). However, the complementary problem of data analysis is far from being sufficiently addressed and requires further research.

Several models for the detection of contamination events, based on the database of a single-site sensor in the network have been developed. Klise and McKenna (2006) and McKenna et al. (2008) used the nearest neighbor method to analyze the parameters' multivariate space, and classified events according to a fixed threshold. Perelman et al. (2012) and Arad et al. (2013) developed

* Corresponding author.
E-mail address: ostfeld@tx.technion.ac.il (A. Ostfeld).

http://dx.doi.org/10.1016/j.envsoft.2015.11.013
1364-8152/

an artificial neural networks (ANN) model for the detection of outlier measurements according to the deviation from the ANN predictions. A Bayesian decision rule was used to evaluate event probability. Eliades et al. (2014) developed a real time EDS, based on an ANN model analyzing changes in chlorine concentration solely. Edthofer (2010) provided a database management tools combined with event detection software. HACH GuardianBlue™ is a commercial hardware and software for event detection applicable in WDS. The CANARY software (Hart et al., 2007; Murray et al., 2010) provides a freely available tool for contamination event detection.

To date, contamination event detection modeling focused mainly on a single-site measured database. The development of EDS based on a multiple sensor stations database was hardly dealt. O'Halloran et al. (2006) characterized the relation between two upstream data sets measurements, according to their measurements pattern, without using any hydraulic information. Koch and McKenna (2011) developed a spatial analysis tool, integrating single location event detection results. They simulated the spread of the contaminant plume, and applied a space-time clustering of the measurements. Using this approach, they defined a group of sensors that were expected to detect the contaminant. By applying a statistical test on the results they examine the alert reliability, where a partial detection by these sensors was suspected to indicate a false alarm. This is a generic method that can be applied for any single station EDS. However the EDS results is assumed to be a binary indication of event/typical operation conditions. The commonly used event likelihood measure is not used in the evaluation of the alert reliability. Yang (2013) integrated the binary signals of each sensor's independent event detection, with a backtracking algorithm. By cross-referencing sensors alerts, he evaluated event occurrence probability. Oliker and Ostfeld (2015) developed the basic model used in this study, features a spatial analysis of artificial on-line measurements, taken from multiple sensor stations. Their study included the integrated of upstream located sensors into a unified event alert system. However, in the absent of multiple sensor stations database, the application in Oliker and Ostfeld (2015) was based on a data set of a single sensor station that was shifted in the average advection time and included randomly generated Gaussian noise that was added to the measurements. The presented study features the model integration with a recently developed contamination simulation model, performing the required adaptations and calibration. The study examined the model capabilities and limitations for a more realistic physicochemical based database.

Unfortunately, for event detection modeling there are no records of real-time event measurements in WDS. Event examples are necessary for training and testing the models. Therefore, all of the aforementioned models applied some random disturbances to recorded operational dataset. These randomized disorders were superimposed on the measured routine patterns in order to represent contamination events.

Several algorithms for simulating water quality in WDS were published in the 1980s and early 1990s (Liou and Kroon, 1987; Rossman et al., 1993; Boulos et al., 1994). One of the most widely used simulation tools is EPANET (USEPA, 2013), which provides the ability to track the fate and transport of a single water-quality constituent or specie. Water quality modeling of multi-species has gained momentum during the past several years through development and application of EPANET-MSX (Shang et al., 2008). This software enables the modeling of multiple constituents throughout a WDS and the chemical reaction mechanism between them. EPANET-MSX describes the chemical behavior based on a set of equilibrium (algebraic) and kinetic (differential) equations, while the constituents advection is based on the EPANET hydraulic solution. Among studies using a multi-species approach were those of Helbling and VanBriesen (2009) that simulated the changes in residual chlorine resulting from microbial contamination events; Klosterman et al. (2012) described an Arsenic contamination event; Husband et al. (2012) utilize the software to described turbidity in main water trunk; Ohar and Ostfeld (2014) used free chlorine-trihalometans model to optimize the location and operation of booster chlorination stations; and Schwartz et al. (2014) formularized the reaction kinetics of chlorpyrifos and parathion in WDS.

Organophosphates (OP) are widely used pesticides that can cause varying degrees of health impacts from nervous system damage to death. Three widely used insecticides of the OP group are chlorpyrifos (CP), parathion (PA), and malathion (MA). Toxicological information for CP, PA, and MA can be found in ATSDR (1997, 2014, and 2003). These three compounds are considered as possible contaminants for this study.

The present study aims at integrating the simulation of contaminant intrusion detailed chemical reactions in a multi-species water-quality model with a multiple-sensor stations spatial event detection model. The simulation produces the database of measurable water quality parameters taken at selected pipeline junctions (model nodes) in the network and thereby offer the ability to integrate the multiple data sets analyses, traditionally assessed as separate problems. Differing from most EDS that are completely generic, the proposed event detection model utilizes network hydraulics and thus, places the model into the context of water distribution system networks. These steps are aimed at promoting the development of a more physically realistic and inclusive EDS.

2. Methodology

The proposed methodology incorporates detailed chemical modeling of contaminant intrusion into a WDS, with a spatial event detection model based on a multiple-sensor station database. The general scheme of the proposed methodology is shown in Fig. 1, which consists of two main stages: (1) contamination simulation and (2) event detection. The first stage includes a network water-quality analysis that simulates water-quality values under typical non-contaminated operating conditions, followed by a contamination simulation that yields a detailed scenario of contaminant spread in a network. The simulation considers the contaminant chemical reactions, degradation by-products, dilution processes and total influence on measurable water quality parameters.

The output of the simulation process includes a continuous database of measurable water quality parameters from selected nodes (i.e., sensor station locations), and a complete hydraulic database description of the simulation. The event detection model analyzes each station's data set separately for an outlier detection procedure, conducted in parallel for all sensor stations.

The initial data analysis process is followed by a sequence analysis for the classification of both local and spatial events. In addition to the sequence analysis, the spatial event classification utilizes the original measurements and the hydraulic data of the network for the detection of suspicious spatial trends. A detailed description of the two-stage model methodology is presented in subsequent sections of this paper.

2.1. Database formation

Events calcification by EDS is based on analysis of datasets containing water-quality values time series. As mentioned, it is custom to use measured sensors data with events, which are generated by random disturbances, as the models database. In this work, the generation of the water-quality database was obtained by conducting a water quality simulation that is based on an explicit