Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 89 (2014) 189 – 196

Procedia Engineering

www.elsevier.com/locate/procedia

16th Conference on Water Distribution System Analysis, WDSA 2014

Water Quality Supervision of Distribution Networks Based on Machine Learning Algorithms and Operator Feedback

C. Kühnert[a]*, T. Bernard[a], I. Montalvo Arango[b], R. Nitsche[c]

[a]Fraunhofer IOSB, 76131 Karlsruhe, Germany
[b]3S Consult Gmbh, 76137 Karlsruhe, Germany
[c]Technologiezentrum Wasser (TZW), 01326 Dresden, Germany

Abstract

Water Distribution Networks contain lots of quality sensors placed in the network. Generally, the analysis of this sensor data, e.g. to check for contaminations, is performed manually by the operators and not by data-driven methods. This has several reasons: First, the parameterization of these methods is time consuming; second, many false positive alarms are generated due to special operational actions. This paper addresses both problems: An alarm detection method is presented needing only a few parameters for configuration and the amount of false alarms is reduced, by using known events for training. The approach is tested on a laboratory plant.

Peer-review under responsibility of the Organizing Committee of WDSA 2014

Keywords: Machine Learning, Time Series Analysis, Condition Monitoring

1. Introduction

Water Distribution Networks (WDNs) are critical infrastructures that are exposed to deliberate or accidental chemical, biological or radioactive contamination. During the last years powerful multi-parameter sensors for water quality monitoring in WDNs have been developed in order to monitor the water quality. These sensors measure several physical and chemical water parameters like conductivity, pH, free chlorine, redox potential, diffusion, and turbidity.

* Corresponding author. Tel.: +49 721 6091-511; fax: +49 721 6091-413.
E-mail address: christian.kuehnert@iosb.fraunhofer.de

Peer-review under responsibility of the Organizing Committee of WDSA 2014
doi:10.1016/j.proeng.2014.11.176

2. Aim of the alarm generation module

The aim of the alarm generation module is the online monitoring of the water distribution network. On the basis of measured historical sensor data, machine learning approaches can be applied to automatically generate a model for monitoring the water quality and quantity. The advantage of this approach is that no analytically-formulated expertise is needed a priori, and thus the user is not hampered by "unsafe" assumptions about the physical / biochemical behavior of the drinking water.

An alarm is generated, if a novelty occurs in the acquired measurements. In that case measurements can describe one station in the WDN which covers several sensors or sensors placed in the network itself. In both contexts, a novelty is defined as a detected unknown state of the WDN. Unknown states can be detected using data from the water quantity as well as water quality. The output of the module is a continuous alarm index value. This value quantifies the difference from the trained model to the acquired measurement data. If the alarm index passes a certain threshold, an alarm is detected.

In the past, several approaches have been investigated and implemented. In [7] an event detection software called CANARY is proposed which contains several statistically based algorithms. The company whitewater has developed a software tool called BlueBox [10], which can be used to perform an enhanced data analysis on water quality and quantity data. Other approaches are presented in [4] by using Support Vector Machines and in [1] by using genetic algorithms. Still, all proposed approaches have not be widely applied in WDNs. Two possible reasons for that are:

- The parameterization of an alarm generation software is complex and time consuming. Appropriate parameters for the machine learning algorithm need to be selected. Additionally, the selection of the alarm threshold is not an easy task.
- A lot of abnormalities detected by the software in the data are due to special operational actions. These leads to false alarms, reducing the credibility of the module. Examples of operational actions which lead to false alarms are sensor calibrations, flushing of pipes or rapid changes of water quality due to mixing of different water resources.

In this paper a new approach is presented which reduces the impact of both problems. The development is part of the project SMaRT-OnlineWDN [8]. The main objective of this project is the development of an online security management toolkit for WDNs that is based on sensor measurements of water quality as well as water quantity.

The paper is structured in two parts:

In the first part, the methodology of the alarm generation module is explained. This covers the preprocessing step, the calculation of the alarm index and the calculation of the alarm threshold. In the second part, results of experiments performed on a laboratory plant are presented. The paper closes with a short summary and proposals for future research.

3. Methodology of the alarm generation module

The proposed alarm generation module covers several steps. Initially, a normalization of the measurements is performed. As multivariate statistical method the principal component analysis (PCA) is used. The PCA is applied to the normalized measurements. In the following sections, it is explained in detail how the alarm index and finally the generation of the alarm threshold are obtained.

3.1. Normalization

Since the event detection module works with different types of quality and quantity parameters (e.g. pressure, conductivity) the measurements need to be normalized. For normalization of the data, the z-score normalization [5] is used, being defined as

$$z[k] = \frac{x[k] - \mu}{\sigma} \tag{1}$$