Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 89 (2014) 173 – 180

Procedia Engineering

www.elsevier.com/locate/procedia

16th Conference on Water Distribution System Analysis, WDSA 2014

Multivariate Data Mining for Estimating the Rate of Discoloration Material Accumulation in Drinking Water Systems

S.R. Mounce[a]*, S.P. Husband[a], W.R. Furnass[a], J.B. Boxall[a]

[a] *Pennine Water Group, Department of Civil and Structural Engineering, University of Sheffield, S1 3JD, UK.*

Abstract

It has been demonstrated that particulate material accumulates as cohesive layers on pipeline surfaces in drinking water distribution systems (WDS) and that when mobilized this material can cause discoloration and other water quality issues. This paper investigates the factors involved in this accumulation rate from real world field data. A data-driven modelling approach is adopted, whereby two machine-learning methods are applied for multivariate data mining based on the observed phenomena. The results highlight bulk water iron concentration, pipe material and looped network areas as key descriptive parameters. Such understanding and expressions are important for pro-active network management.

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

Keywords: Material accumulation; Discolouration; Turbidity; Operation and maintenance strategies; Evolutionary Polynomial Regression; Self Organising Maps.

1. Introduction

Discolored water is the most apparent water quality issue to customers and the major cause of customer complaints for water service providers. It has been shown that particulate material accumulates as cohesive layers on pipeline surfaces in drinking water distribution systems (WDS) as a ubiquitous process and that when mobilized this material can cause discoloration and other water quality issues. The cohesive nature and variable shear strength properties of material layers shown to cause discoloration, and how these developing layers are conditioned by the hydraulic regime,

* Corresponding author. Tel.: +44 (0)114 222 5750; fax: +44 (0)114 222 5700.
E-mail address: s.r.mounce@sheffield.ac.uk

Peer-review under responsibility of the Organizing Committee of WDSA 2014
doi:10.1016/j.proeng.2014.11.174

have been demonstrated [1, 2]. Factors influencing this accumulation (also referred to here as regeneration rate) might include local pipe properties such as age, material or diameter, whilst hydraulic conditions and bulk water quality are also considered important. Fig. 1 captures some of the possible factors the literature suggests are important. Fieldwork results have suggested that accumulation rates are a linear function of time and are dominated by the supplied water quality [3, 4]. There has been limited work on how discoloration material accumulates and the key factors in the accumulation rate. Note that discoloration risk and material accumulation rate are not the same. Risk is hydraulically mediated, while the accumulation rate is more based on water quality and pipe material. Models that can provide site-specific predictions of regeneration rates do not yet exist but a basic bi-variate categorical breakdown of discoloration rates are presented by Husband and Boxall [1]. They showed that the development of material layers is a reproducible and repetitive process. Given the complex, interrelated nature of the physical, chemical and biological reactions that are considered to contribute to discoloration material regeneration (see Fig. 1) it could be expected that predictive models of regeneration rates will only be engendered through the application of multi-variate, regressive methods to sufficient volumes of representative data. McClymont et al. [5] presented an approach for the multi-objective optimization of water distribution systems using a new hyper-heuristic called the Markov-chain hyper-heuristic, for which one of the objectives was discoloration risk. They specifically sought to examine the impact of pipe diameter on discoloration risk. Trading off various considerations including optimizing network design and rehabilitation costs along with discoloration risk is possible, with constraints such as pipe velocities and node heads. However, specific models of material regeneration are simplistic and are not sufficiently representative of reality to allow for prediction. This paper investigates how accumulation/regeneration rate can be correlated with other system information such as source water quality and pipe material.

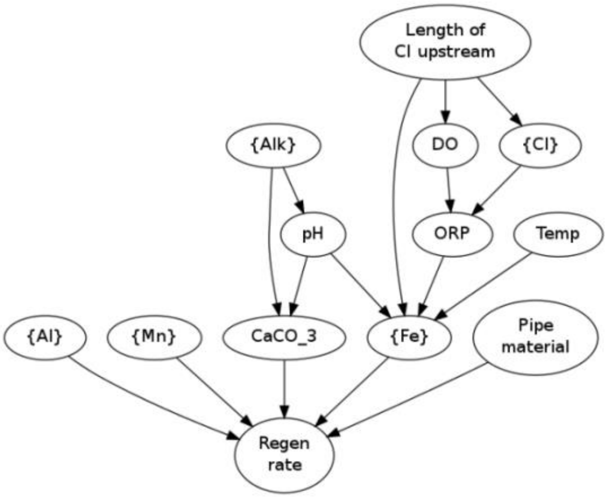

Fig. 1. Potential factors determining material accumulation rate

A data-driven modelling approach is adopted for this paper, whereby two machine learning methods, Kohonen Self-Organising Maps (SOMs) and Evolutionary Polynomial Regression (EPR), make use of a set of real world data for multivariate data mining based on the observed phenomena. The focus was on knowledge discovery of casual factors rather than actual prediction and validation of a model in this preliminary study.

2. Methods

2.1. Data mining for water resources knowledge discovery

Machine learning techniques are capable of identifying complex nonlinear relationships between inputs (factors potentially affecting regeneration) and output (regeneration rate). Based on the discovered relationships within