International Conference on Hydroinformatics

2014

# Predicting Iron And Manganese Accumulation Potential In Water Distribution Networks Using Artificial Neural Network

Prasad Devi Tumula

Ebenezer Danso-Amoako

# PREDICTING IRON AND MANGANESE ACCUMULATION POTENTIAL IN WATER DISTRIBUTION NETWORKS USING ARTIFICIAL NEURAL NETWORK

T. DEVI PRASAD (1), EBENEZER DANSO-AMOAKO (1)
*(1): Department of Civil Engineering, The University of Salford, The Crescent, Salford, M5 4WT, United Kingdom*

In April 2010, the Water Services Regulation Authority in England and Wales, OFWAT, introduced the Service Incentive Mechanism (SIM) which rates water companies on their performance based on customer satisfaction and either reward or penalise them. In view of this, it has become extremely important for water companies to lower customer complaints due to drinking water discolouration; which is approximately 34% of all customer complaints. Presently, most water companies identify high discolouration risk regions in water distribution networks (WDNs) by selecting areas in the network with high Iron (Fe) and Manganese (Mn) concentrations from their random sampling. With about 315,000 km of water mains in England and Wales, monitoring Fe and Mn concentrations will always be a very difficult and expensive task. In this paper, an artificial neural network (ANN) model was developed to predict Fe and Mn accumulation potential using relevant biological, chemical, hydraulic and pipe-related parameters. The model is able to predict Fe and Mn accumulation potential for each node in a given water supply zone (WSZ). It was observed that most regions in the network with high Fe and Mn accumulation potential from the risk maps generated by the model for each of the WSZs also had high customer complaints due to discolouration. This model can be used as a tool to assist in reducing discolouration and customer complaints by helping water resource engineers to identify the high risk regions, investigate the causes of high Fe and Mn accumulation potential in those regions and if possible find solutions to them.

## INTRODUCTION

The processes influencing the accumulation and release of Fe and Mn in WDNs are highly complex, unpredictable, not fully understood, and at present, difficult to model mathematically. Concentrations of Fe and Mn change frequently with time and space as water moves from the treatment plant to customers. The variability of source materials, hydraulics, biological and chemical reactions that occur within a network contribute toward creating a very complex environment that is difficult to understand. The parameters that influence accumulation can be grouped into three categories: (a) chemical parameters representing the chemical reactions within a WDN; (b) biological parameters that aid the accumulation; and (c) physical parameters such as the condition and age of infrastructure such as pipes and network hydraulics. A comprehensive review on factors that influence Fe and Mn accumulation has been published by Prasad and Danso-Amoako [1].

Understanding the processes that influence Fe and Mn accumulation and identifying the regions in WDNs that have high risks of Fe and Mn failures are of paramount importance to all water utilities. Despite their efforts to comply with the standards for drinking water, they continue to receive customer complaints related to water quality. Given the nonlinear nature of water quality data and their interactions that occur within them, artificial-intelligence based methods may be the best method to model water discolouration. A number of researchers have attempted to use artificial neural networks (ANNs) to model water discolouration. Stanley *et al.* [2] used ANNs to forecast turbidity and colour removal, through enhanced coagulation. Whiles Zhang and Stanley [3]used ANNs in raw water colour forecasting.

The rest of this paper is organized as follows: The next section shows how the model parameters were calculated. The model development section presents the ANN model for predicting Fe and Mn accumulation potential using relevant biological, chemical, hydraulic and pipe-related parameters. The results generated by the model were discussed in the last section.

## DATA PREPARATION

The selection of appropriate input variables are very important for the development of any type of model. It reduces the cost of collecting unwanted data and improves the performance of the model. In ANNs, including inappropriate input variables confuse the training process. A five-year dataset of 37 water quality parameters covering 14 water supply zones (WSZs) which consists of 176 different District Metered Areas (DMAs) from a UK water company was used. In the context of this research, the sites of interest included WSZs with high, medium and low levels of customer complaints in order for the model to capture all levels of discolouration and to remove any form of bias. Hydraulic parameters and pipe-related parameters like pipe type and pipe age were also extracted from the hydraulic files obtained. The following sections show how the model's input parameters were computed.

### Maximum daily shear stress at node

To investigate the influence of shear stress on Fe and Mn accumulation, a computer program was written as an add-on code to the Epanet software to extract all pipe and node parameters. From the program, the shear stress was computed every 15 minutes for 24 hours for each pipe in the network and the maximum daily shear stress for each pipe recorded. The maximum daily shear stress in a pipe can be mathematically expressed as Eq. (1). In view of the fact that shear stress is a pipe property, a methodology was devised to calculate the maximum daily shear stress at nodes. The maximum daily shear stress at a given node was calculated by summing the maximum daily shear stress of pipes connected to the node and divided by number of pipes connected to it. Mathematically it can be expressed as Eq. (2).

$$\tau_{\mathrm{pipe(max)}} = \frac{\rho g d H}{4L} \tag{1}$$

$$\tau_{\mathrm{node(max)}} = \frac{\sum_{i=1}^{n} \tau_{\mathrm{pipe(max)}}}{N} \tag{2}$$

where $\tau_{(pipe)}$ = link boundary shear stress  $\rho$ = density of water
g = acceleration due to gravity  H = head loss  L = length of pipe  d = diameter
$\tau_{node(max)}$ = maximum daily shear stress at the node  n = the $n^{th}$ pipe connected to a node
$\tau_{pipe(max)}$ = maximum daily shear stress of the pipe  N = Number of pipes connected to a node

**Variation in daily shear stress at node**

The diurnal variation in WDNs due to the continuous variation of drinking water demand causes the shear stress in pipes to vary from time to time. The shear stress acting on the walls at peak demand times are higher than off peak times. A program was written as an extension to the Epanet software to compute variation in daily shear stress at each node.

**Water age**

The age of water in a WDN, often referred to as residence time, is the time taken for treated water to travel from the treatment plant to a given node. It is a vital parameter which can determine the extent of disinfectant loss. It may range from a few seconds to weeks. A computer program was written as an add-on code to the Epanet software to calculate the water age at all the nodes in the network after 72 hours of simulation.

**Hydraulic distance from source of water supply**

Hydraulic distance from source of water supply is the distance taken for a particle of water to travel from a source of water supply to a given node within a WDN. To measure this vital parameter, a Matlab program was written as an add-on code to Epanet software to calculate the hydraulic distance of each node from source of water supply. The program makes use of two main algorithms; namely Particle Backtracking Algorithm (PBA) and a Shortest Route Algorithm (SRA).

The PBA that was used was adopted from Shang *et al*. [4] and modified to suit this model. PBA is an algorithm that is able to track particle movements in water between any two points in a WDN even with multiple water sources. This algorithm is able to track changes of discrete particles of water as they travel through the network. Unlike the traditional simulation approaches in calculating delays, the PBA is able to give the numerous flow paths that water travels between a given source node and output node as well as their corresponding time delays. PBA uses Lagrangian time-driven method which runs in reverse time for its computations; that means it runs opposite to the hydraulic simulation time. It has the ability to calculate the number of paths, their delays and impact factors. A detailed algorithm with and without multiple storage tanks has been published by Shang *et al*. [4]. The SRA was used to calculate the shortest distance from each node to all the reservoirs / tanks. While the PBA was used to find which of the reservoirs / tanks supplied the nodes with water.

**Water quality parameters**

Yearly averages were calculated for phosphorus, aluminium, turbidity, temperature, hardness, free chlorine residual (FCR) and water colour in each of the district metered areas (DMA). Each node in a given DMA was assigned these yearly averages in the model.

**MODEL DEVELOPMENT**

The ANN model tries to unravel the complex process of Fe and Mn accumulation by taking a holistic approach to combine all the parameters. The input parameters for the model were hardness, free chlorine residual, temperature, alkalinity, aluminium, water colour, temperature, phosphorus, turbidity, pipe material, pipe age, water age, maximum daily shear stress at node, variation in daily shear stress at node, and hydraulic distance from source of water supply. The predicted parameter is the Fe and Mn accumulation potential. The measured Fe and Mn accumulation potential values were calculated by dividing Fe and Mn concentrations by their respective MCLs which are 200 µ/l and 50µ/l respectively. They were then summed and scaled between 0 and 1.

These parameters were used to develop a four-layered ANN using the commercial statistics software JMP 10 Pro (Figure 1). Generally, a neural network consists of three layers of neurons: input layer, hidden layer and output layer. However, for this model an additional hidden layer was added to improve the model's prediction accuracy. The input layer receives the data and passes them on to the hidden layer, and finally the output layer generates the predicted values (Fe and Mn accumulation potential). All neurons in their layers were fully weighted and connected, and the weights were adjusted during the training phase. Considering that different variables have different scales, all the data were normalised to a unique scale (between 0 and 1). To get unbiased partitioning of a small-sized data set, the commonly used 4-fold-cross-validation method was applied to the data. This means that the whole data set is split into 4 folds, any four folds are used for training and the other left for testing, and the validation results are averaged over the rounds.

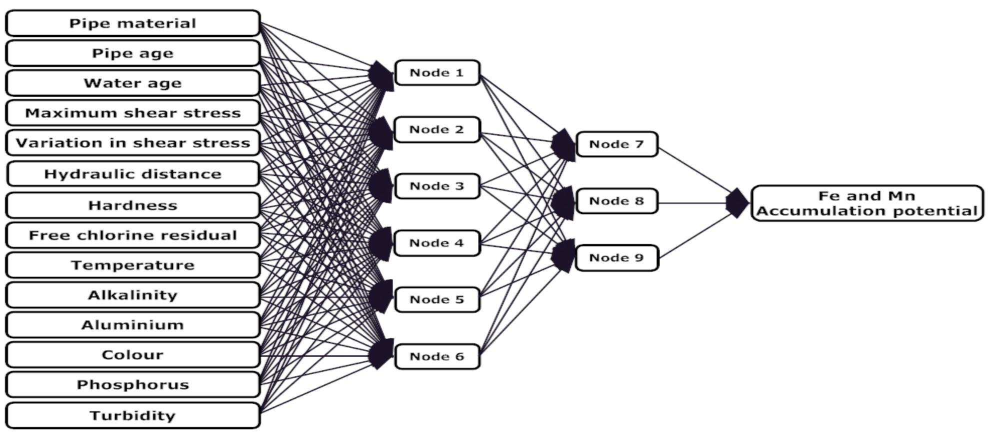


Figure 1. The ANN model

**RESULTS AND DISCUSSIONS**

**Analysis of input parameters**

Fe and Mn concentrations were plotted against maximum shear stress at node for each of the WSZs (Figure 2(a)). Low maximum daily shear stress regions were found in sections of the pipe network that have dead ends and redundant loops. While high maximum daily shear stress regions were mainly found on trunk mains and regions with high water demand. From the graphs, it was observed that areas with high maximum daily shear stress had low Fe and Mn concentrations. This is because Fe and Mn precipitates are unable to accumulate on the pipe walls under these high conditions.

It was observed that Fe and Mn had high concentrations at low variation in daily shear stress and low concentrations at high variation in daily shear stress. Figure 2(b) shows a graph of Mn against variation in daily shear stress at nodes. Nodes with low variation in daily shear stress generally have low disturbance in WDNs. This implies any deposition of Fe and Mn precipitates in these regions are more likely to be attached to the walls of the pipes. This caused high levels of Fe and Mn at regions with low variation in daily shear stress. Conversely, nodes with high variation in daily shear stress generally have high disturbance in the WDNs. This means Fe and Mn precipitates in these regions are more likely not to accumulate on the pipe walls. This results in low concentrations of these water quality parameters in these regions. In general, pipes with high variation in daily shear stress have low Fe and Mn accumulation potential. While pipes with low variation in daily shear stress have high Fe and Mn accumulation potential.

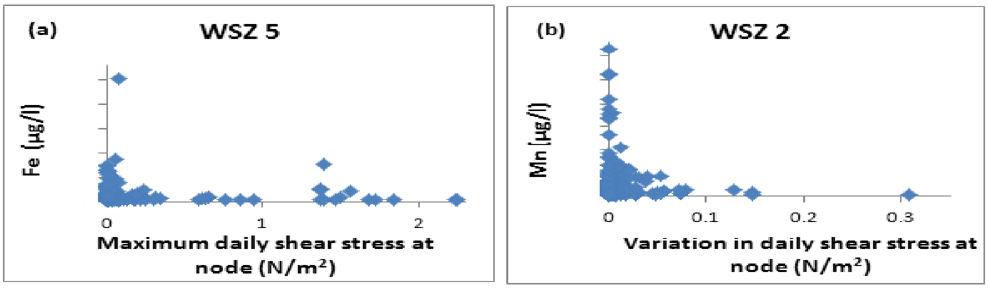


Figure 2. Variation of Fe against maximum daily shear stress at nodes and Mn against variation in daily shear stress at nodes.

Water age is a major factor in water quality deterioration within the distribution system. The distribution of water age in all the WSZs revealed that high water ages were predominantly found in regions with dead ends and redundant loops. The water age in a WDN also depends on its mode of operation as well as physical parameters such as the flow rate, pipe size, configuration, water demand, system design and amount of storage. WDNs with high flow rates and small pipe sizes will have a lower water age. Water quality problems that can be associated with water age include, poor taste, bad odour, increased microbial growth, discoloration and increased water temperature.

Graphs of Fe and Mn concentrations plotted against hydraulic distance from source of water supply for all the WSZs showed a gradual increase in Fe and Mn concentrations as hydraulic distance from source increases. Figure 3 shows a sample graph of Fe concentrations plotted against hydraulic distance from source of water supply for WSZ 2. In general, the further water travels through a WDN, the higher the residence time and the more chlorine dissipates in the system. Since chlorine is a disinfectant it suppresses the growth or kill Fe and Mn oxidising bacteria, preventing them to biologically oxidising soluble Fe and Mn to insoluble Fe and Mn. Hence, regions with short hydraulic distance from source of water supply have low Fe and Mn concentration. On the other hand, regions with long hydraulic distances from source have low concentrations of FCR. This increases microbial growth which causes biological oxidation of Fe and Mn and subsequently leads to increased Fe and Mn concentrations.

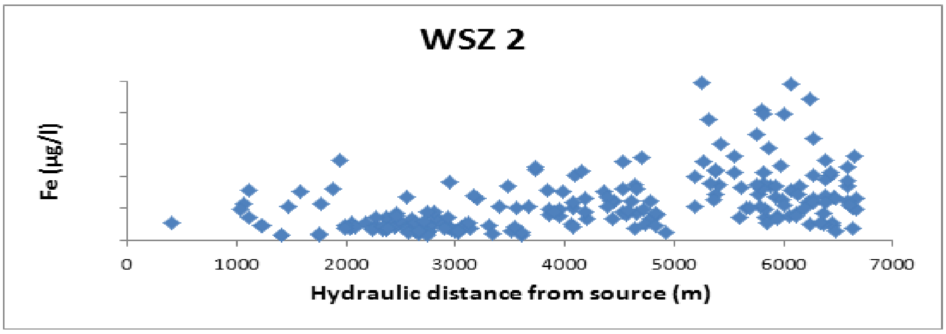


Figure 3. Iron concentrations plotted against hydraulic distance from source of water supply

**Analysis of model results**

The high cross-validation $R^2$ values observed in both the training and validation data sets respectively shows that the predicted values are similar to the actual values; an indication that the model is likely to predict Fe and Mn accumulation potential reasonably well on new datasets (Table 1). Similarly, the low RMSE values for the training and validation data sets respectively indicates that the model is likely to predict well on new data sets.

Table 1. Performance indicators for the model

| WSZ | Training $R^2$ | Validation $R^2$ | Training RMSE | Validation RMSE | Customer complaints (2005-2009) | Number of high risk nodes predicted by model |
|---|---|---|---|---|---|---|
| WSZ 1 | 0.80 | 0.80 | 0.20 | 0.17 | 153 | 198 |
| WSZ 2 | 0.83 | 0.82 | 0.16 | 0.16 | 328 | 466 |
| WSZ 3 | 0.83 | 0.76 | 0.20 | 0.20 | 131 | 59 |
| WSZ 4 | 0.87 | 0.75 | 0.19 | 0.14 | 497 | 571 |
| WSZ 5 | 0.86 | 0.72 | 0.21 | 0.13 | 254 | 146 |

Most water utilities use increased Fe and Mn concentrations from their sampled data and customer complaints data to determine high discolouration risk regions. This can sometimes be imprecise; since it will be impossible to sample every node in a large WSZ. Regions which have high Fe and Mn concentrations that are not sampled will not be detected. Furthermore, according to studies conducted by Ewan and Williams [5] in the United Kingdom, approximately 30% of customers that experienced discoloured water event actually complain. This means that there is a high tendency that certain regions in WSZs with high customer complaints (Fe and Mn concentrations) can go undetected. The model is able to solve this problem by predicting Fe and Mn accumulation potential for each node in a given WSZ.

A matlab program was written to plot the risk maps for the predicted Fe and Mn accumulation potential. Figure 4 shows customer complaints due to discolouration in WSZ 2. While Figure 5 and 6 show the corresponding predicted risk map generated by the model and the measured risk map respectively. It was observed that most of the regions in the network with high Fe and Mn accumulation potential from the risk maps generated by the model for each of the WSZs also had high customer complaints due to discolouration. These results conforms to research conducted by Slaats [6] who observed that the gradual accumulation or sudden increase in Fe and Mn particles in WDNs was the most common cause of water discolouration and customer complaints. This also explains why some researches have used Fe and Mn concentrations as

Key Performance Indicators (KPIs) in customer complaints studies (Ewan and Williams [5], Gauthier *et al*. [7] and Bernal *et al*. [8]).

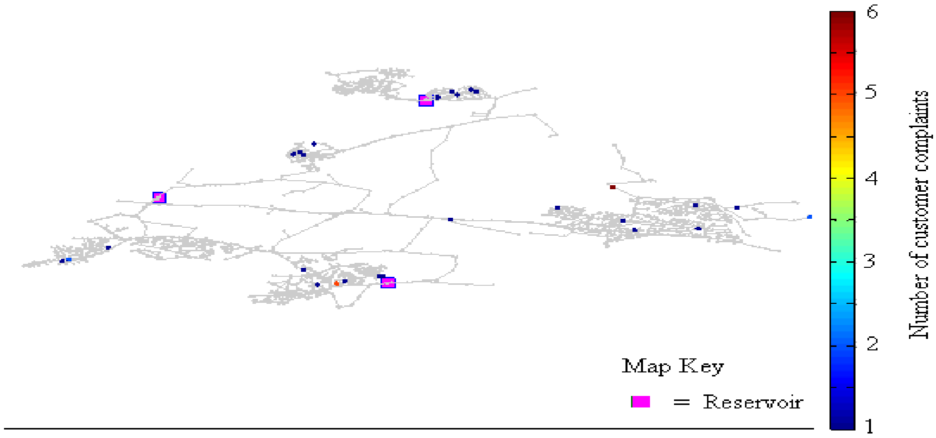


Figure 4. Customer complaints risk map for WSZ 2.

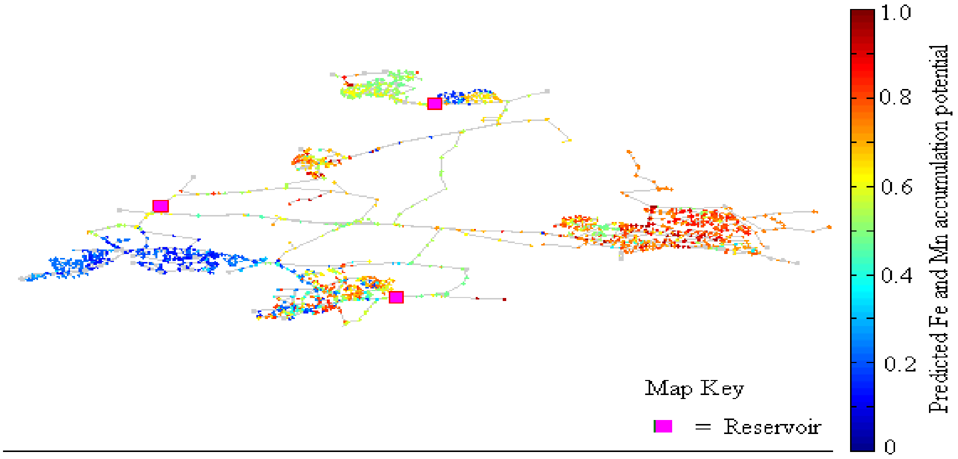


Figure 5. Predicted Fe and Mn accumulation potential for WSZ 2.

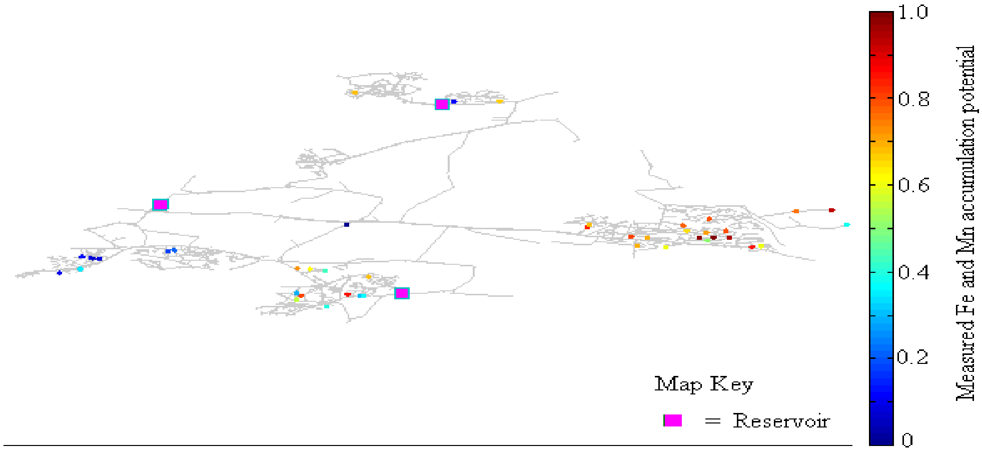


Figure 6. Measured Fe and Mn accumulation potential for WSZ 2.

## CONCLUSIONS AND RECOMMENDATIONS

The design of WDN is very important to reduce water discolouration due to increased water age. To reduce this, it is recommended that dead ends be prevented or looped, reservoir turnover be increased, oversized mains be reduced and stagnant zones be routinely flushed.

The developed model can be used as a tool to assist in reducing discolouration and customer complaints by helping water resource engineers to identify the high risk regions, investigate the causes of high Fe and Mn accumulation potential in those regions and if possible find solutions to them. Even though the ANN model is able to predict areas in the network with high or low Fe and Mn accumulation potential, it is unable to explain the reasons why they are high or low because of the black box nature of ANN.

## REFERENCES


*[1] Prasad T. D., and Danso-Amoako E., "Influence of chemical and biological parameters on iron and manganese accumulation In water distribution networks," in Proceedings of the 12th International conference CCWI 2013 "Informatics for water systems and smart cities", Perugia, Italy, (2013).*

*[2] Stanley S., Baxter C., Zhang Q., and Shariff R., "Process Modelling and control of enhanced coagulation," AWWA Research Foundation and American Water Works Association, (2000).*

*[3] Zhang Q., and Stanley S. J., "Forecasting raw-water quality parameters for the North Saskatchewan River by neural network modeling," Water Research, Vol. 31, No. 9, (1997), pp. 2340-2350.*

*[4] Shang, F., Uber J. and Polycarpou M., "Particle backtracking algorithm for water distribution system analysis," Journal of Environmental Engineering, Vol. 128, No. 5, (2002), pp. 441-450.*

*[5] Ewan V. J. and Williams S., "DoE Contract E15 interim report, monitoring water quality deterioration in distribution systems, final report," Water Research Centre, Engineering. Wiltshire, UK., (1986).*

*[6] Slaats N., "Processes involved in generation of discoloured water," American Water Works Association Research Fundation, The Netherlands, 2002.*

*[7] Gauthier V., Gérard B., Portal J.-M., Block J.-C. and Gatel D., "Organic matter as loose deposits in a drinking water distribution system," Water Research, Vol. 33, No. 4, (1999), pp. 1014-1026.*

*[8] Bernal A., Cardenoso R., Fabrellas C., Matia L. and Salvatella, N., "An aesthetic quality index for Barcelona's water supply," Water Science and Technology, Vol. 40, No. 6, (1999), pp. 23-29.*