CENTRAL ASIAN JOURNAL OF WATER RESEARCH (2024) 10(1): 40-67

https://doi.org/10.29258/CAJWR/2024-R1.v10-1/40-67.eng

Central Asian Journal of Water Research

Assessing water quality in a distribution network based on hydraulic conditions

Tomperi J.[a]

[a] University of Oulu, Pentti Kaiteran katu 1, Oulu, 90570, Finland

ABSTRACT

Abnormalities in hydraulic conditions inside a water distribution network are strongly related to the deterioration of drinking water quality. Leaking pipes and valves cause changes to the hydraulic conditions and allow the entry of impurities into the distribution system. Sudden flow and pressure shocks can detach soft deposits and biofilms from the pipe surface, resulting in deterioration of water quality. Online water quality measurements in a distribution network are scarce, but more common online flow and pressure measurements reveal the changes in hydraulic conditions in a distribution network and can be utilized to assess the water quality continuously and near real-time via modelling. Here, a data-driven model based solely on the online flow and pressure measurements in a distribution network for assessing the water quality at the end of an urban district metered area is presented. With the accuracy of R^2 0.77, the developed data-driven model is able to assess the level of and the changes in potable water quality in a non-laborious and cost-effective way, also enabling proactive operations to ensure the distribution of high-quality drinking water to the consumers.

ARTICLE HISTORY

Received: January 12, 2023
Accepted: December 21, 2023
Published: February 6, 2024

KEYWORDS

drinking water, flow, modelling, pressure, data-driven modelling, turbidity

CONTACT Jani Tomperi jani.tomperi@oulu.fi, University of Oulu, P.O. Box 4300, Oulu, FI-90014, Finland

Published by Kazakh German University, Almaty/Kazakhstan

1. Introduction

Providing potable water with appropriate quality and quantity through the water supply system, the infrastructure that collects, treats, stores, and distributes water between the raw water sources and consumers' taps, is essential for the general well-being and health. In addition to potential health-related risks, extra financial costs and energy consumption occur when potable water is contaminated. Disturbances in potable water quality are most commonly dealt with in a reactive way by the water company, i.e., by cleaning parts of the distribution network (water flushing, water/air scouring, and swabbing/pigging) (Vreeburg & Boxall, 2007) and providing clean water for consumers using tanker trucks, etc. The deterioration of potable water quality strongly relates to problems in a water treatment process or abnormalities and sudden changes in the hydraulic conditions inside a water distribution network (WDN) (Clark & Haught, 2005).

In a WDN, microbiological growth, the re-suspension, and mobilization of precipitation, or breakdown of pipes and valves allowing the entry of particles and microbes, can deteriorate the water quality. When distributing potable water, biofilms will inevitably grow on the inner surfaces of the pipes, and soft deposits consisting of organic and inorganic matter and several metals will accumulate in the pipelines. Rapid changes in water flow or pressure can detach the biofilms and soft deposits from the surfaces of the pipes and deteriorate the water quality, which is shown, among others as the elevated concentrations of bacteria, metals, and turbidity in water (Lehtola et al., 2006b). Mustonen et al. (2008) found that pressure shocks temporarily increase the number of particles, turbidity, and electrical conductivity in water. The rapid change in the flow creates high shear stresses which cause particle mobilization from sediments and biofilms along the pipe affecting the water quality. The mobilization occurs in the first dynamic surge of the transient. The events causing the water quality deterioration may be rapid and occur over a short time. The changes in water quality typically have characteristics of a sharp rise that reduces within a few hours (Vreeburg & Boxall, 2007). New insight into how the hydraulic transients, that occur within drinking water distribution networks, can mobilize material adhered to the pipe wall and hence cause unacceptable water quality, and customer dissatisfaction is provided by Weston et al. (2021). In addition, unsteady hydraulic conditions can lead to a significant impact on the disinfectant residual as a significant decay in the disinfectant residual is attributed to the mobilization and entrainment of particles and biofilms, which affect the bulk and wall reactions during the unsteady hydraulic conditions (Aisopou et al., 2012). The lack of pressure and slow water flow lead to long detention times, increasing the biofilm growth and bacterial regrowth, which may ultimately lead to water-borne diseases as the long travel and detention times contribute to the loss of disinfection residual. Long detention times are a significant contributing factor in the formation of disinfection by-products (DBPs), which are formed when the disinfectant reacts with organic and inorganic substances in water. Many DBPs have toxic properties and can be mutagenic and genotoxic. (Clark & Haught, 2005; Ghebremichael et al., 2008; Mains, 2008; Li, 2017; Manasfi, 2017)

Properties that affect the water quality can be physical, chemical, or biological factors. Turbidity is one of the critical water quality parameters in environmental monitoring, industrial process operation, and water treatment and distribution. Turbidity, an optical measure of clarity describing the physical transparency of liquid, and total suspended solids in water are strongly related. Still, turbidity is not a direct measurement of the total suspended materials in water as other factors also affect the measured turbidity value (Tomperi et al., 2022). In a water distribution system, suspended particles that cause the water turbidity are usually from raw water sources due to inadequate water treatment or from re-suspension of sediments in a distribution system. Turbidity and water flow are causally related as the high flow prevents particles from settling, and significant changes in velocity can increase turbidity and corrosion in the distribution system (Chapman, 1996). Despite its unpleasant appearance, turbid water is not harmful to health. Still, the increase in turbidity can often indicate potential pollution as pollutants such as dissolved metals and pathogens can attach to suspended particles. The number of bacteria in water correlates with the turbidity and the number of particles in water (Lehtola et al., 2006a). High turbidity also hinders the effect of disinfection, and the suspended particles can carry impurities, protect microorganisms from the effects of disinfection and stimulate bacterial growth. (Chapman, 1996; Anderson, 2005; WHO, 2008; Mukundan et al., 2013)

Typically, water utilities monitor the quality of potable water daily. Still, the monitoring is often carried out by manual sampling and laboratory analysis, and only at a few key locations of the WDN, as mounting and maintaining online sensors for water quality monitoring in an old WDN is laborious and expensive. Laboratory analyses are also expensive and time-consuming, and the results always present the past. Therefore, the current commonly used water quality monitoring frequency and broadness are not sufficient to detect abrupt and transient events in real time nor discovering the slow and long-term changes in the water quality in a wide district metered area (DMA). Often the deterioration of water quality is noticed by the customers. Hence, there is lack of important real-time information on the state of the water quality and the conditions of the network in wide parts of the WDN. On the other hand, water distribution networks include at least some amount of online pressure and flow measurements, which show the hydraulic conditions continuously and in real-time. It is advisable to utilize this existing information via a modelling approach to increase the knowledge on water quality at the parts of a distribution network where no stand-alone water quality measurements are available, with a faster and more cost-effective way against the laboratory analysis.

Several water quality models have been developed for design and operational management purposes and to help ascertain the quality of potable water in the water distribution systems. For example, an integrated pressure-dependent hydraulic model based on the well-known Epanet 2 model was used to simulate the operating condition, including normal and subnormal pressure (Seyoum & Tanyimboh, 2013). A full-scale hydraulic model of the whole water supply system was built by Sunela & Puust (2015) using an extended version of EPANET including simulated quality parameter modeling. On the other hand, a data-driven turbidity forecasting method capable of aiding operational staff and enabling proactive management strategies was developed by Meyers et al. (2017). Development of a hydraulic model includes many challenges as they require detailed information on the distribution network and may therefore include some assumptions which cause uncertainties to the model performance. Significant spatial and temporal variations in hydraulic conditions can occur inside the water distribution network. While the pressure is nearly constant in stable conditions, the water flow rate varies within a day, from day to day and season to season. The dynamic behavior poses challenges as the water usage is time-dependent and tied to the type of water users, for example, residential or industrial. In addition, the development and maintenance of hydraulic models are time-consuming and expensive. Data-driven models, on the other hand, require only the measured data, and therefore they present more accurately the current situation. However, the accuracy of the data-driven model strongly depends on the quality of measured data and the selected model variables.

In this paper, the development of a data-driven model for assessing the water quality, namely turbidity, in an urban distribution network is presented. Water turbidity in the end part of the DMA is modelled based on only the online hydraulic condition measurements along the early part of the WDN. The presented work is a revision for the unfinished work introduced in Tomperi (2020). With the simple data-driven modelling approach, the water quality information is provided in a non-laborious and cost-effective way utilizing the exiting measurement data. The received water quality information could be utilized as an early warning of changes and for proactive management strategies by operation personnel to ensure the distribution of high-quality potable water.

2. Material and Methods

2.1. The research site and collected data

The site of this research work was an urban DMA, where several online monitoring stations were mounted along with the water distribution system (Figure 1), including a raw water source pumping station (marked as I), pressure increase pumping stations along with the DMA, and distribution pipelines at the length of tens of kilometers. For confidential reasons, the exact locations or detailed map of the DMA cannot be presented here. Hydraulic condition parameters were monitored in several places (marked as circles), and the water quality measurements were performed at two locations: at a raw water pumping station (marked as I) and one monitoring station at the end part of the distribution network (marked as O). Water quality measurements were performed with YSI 6920V2, multi-parameter probes for continuous water quality monitoring, which included turbidity, dissolved oxygen, oxygen reduction potential, pH, salinity, specific conductivity, and temperature sensors. Water quality and hydraulic conditions data were collected from the WDN during two separate campaigns, in spring/early summer, and in fall, covering in total data from a period of circa 120 days. The data were initially logged at a minute interval but were averaged for this research.

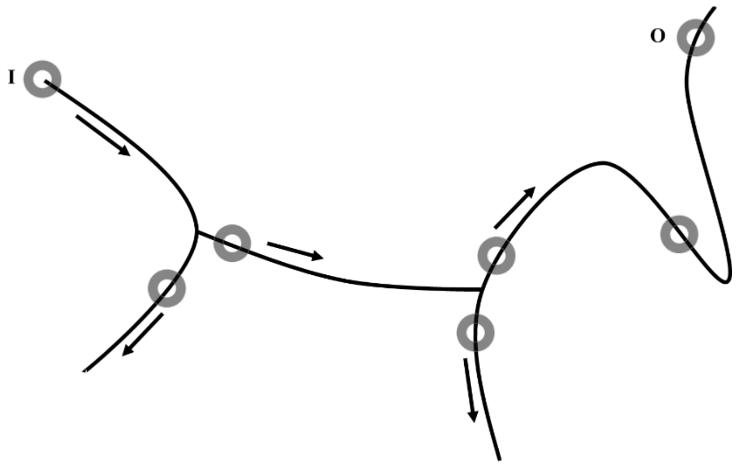

Figure 1. A rough illustration of the WDN and the measurement locations.

2.2. Data analysis and modelling

Datasets generally include irrelevant variables to develop a model for a specific purpose. Selecting the proper amount and the optimal input variables for a model from a large variable set is one of the crucial steps in model development. Too few input variables or noisy and uninformative variables lead to a model with poor performance. On the other hand, using too many input variables increases the risk of developing an over-fitted model with excellent training results but poor prediction abilities with new data. The selection can be done manually or using various automated variable selection methods. In this research, a combination of expert knowledge and a stepwise regression method was used to select the optimal input variables for the water quality model. Stepwise regression is a modified forward selection method, which adds the best variable to or deletes the worst variable from a variable subset at each round. Adding and deleting are based on the variable's statistical significance in regression. It starts with an initial model and continues until either no further model changes occur over one complete round, or a pre-set number of variable selections and deletions occur.

As the amount of collected data for this research was relatively low including two individual clusters of turbidity values (as shown in the Results section), it was not feasible to divide the data traditionally into training and validation subsets. Instead, cross-validation was used. Cross-validation is an efficient resampling method to predict the fit of a model when the dataset is small. In k-fold cross-validation, the whole data set is used for training and validating the model by dividing the dataset randomly into k equal size partitions, and in turn, using k-1 partitions to train the model and one partition for validation. The procedure is repeated k times until every partition is used once for validation. The final estimation is produced by averaging over the k models' prediction results. (Arlot & Celisse, 2010)

In this study, multivariable linear regression (MLR) was used to estimate the output variable as a linear combination of selected input variables as in Equation (1).

$$y=b+p_1 x_1+p_2 x_2....+p_n x_n, \tag{1}$$

where y is the estimate output variable, x is the selected input variable, p is a model parameter, and b is the bias value defined from the data. Root Mean Square Error (RMSE) and coefficient of determination (R^2) are used to evaluate the relative performance of the model.

3. Results and Discussion

The delays between the monitoring locations in the DMA were studied both mathematically using auto-correlation analysis and inspecting the figures of flow, pressure, and water quality measurements visually to compare the temporal location of the peaks and level changes. The auto-correlation analysis measures the similarity between the variables as a function of the lag and indicates the delays between measurement locations. Both mathematical and visual inspections showed that variables from the consecutive monitoring stations in the WDN have logical delays. Based on the mathematical analysis, the quality measurements in the end part of the network have a delay of up to 20 hours to the beginning of the distribution network.

Water turbidity was modelled with the aforementioned MLR modelling method. Even though the linear model does not find the nonlinear relations between input and output variables, it has a simple structure, and it is easy to understand and implement. MLR models are also suggested being used to avoid over-fitting, and in some cases, they can outperform the more complex and computationally heavier nonlinear models (Hastie et al., 2009; Montgomery et al., 2012). In Table I is shown the performance values R^2 (1.0 means the perfect match) and RMSE together with the selected input variables (x), the bias (b), and the parameter (p) values of the developed model. The measured turbidity and the modelled turbidity with the 95% confidence interval are shown in Figure 2. For confidential reasons, no exact dates or turbidity values are shown, but the values are scaled to the range [0 1]. As seen, there are two main clusters of turbidity values due to the two separate data collecting periods in spring/early summer and late summer/fall. The developed model is not perfect, and some turbidity values fall out the 95% confidence interval.However, the model is able to determinate both the higher and lower levels of water turbidity even though the lower turbidity values seem to be harder to model correctly. As seen in the model structure presented in Table I, the turbidity is highly dependent on the one flow (F) measurement and one pressure (P) measurement at the beginning of the network and less dependent on the three pressure measurements at the later part of the network. In Mustonen et al. (2008), it was reported that pressure shocks temporarily increase the turbidity in water.

When evaluating the presented results, it is essential to take into account that only flow and pressure measurements were used as input variables, and no other measurements such as water quality were utilized as inputs. For example, the temperature, which differs by many degrees of Celsius between and during the two data collecting campaigns, is an essential factor when assessing the water quality as the temperature can alter the physical and chemical properties of water and influences several other quality parameters. Therefore, the achieved performance of the developed turbidity model for assessing the water turbidity can be considered reasonable. Also, due to the limited resources available, no external long-term testing data were available for proofing the functionality of the model in various and varying conditions. However, based on the presented results it is reasonable to state that it seems that the level of and changes in turbidity at the end part of the WDN can be estimated based on only the existing online hydraulic condition measurements. The model-based estimation on water quality requires no other information, manual sampling, or mounting and maintaining any additional sensors. Against the laborious, expensive and infrequent manual sampling and laboratory analysis, the presented data-driven modelling approach enables assessing the water quality in a non-laborious and cost-effective way utilizing only the data from the existing online flow and pressure measurements. Hence, utilizing the developed model, essential information on the water quality and the state of the distribution network can be achieved also from the areas where no stand-alone water quality measurements exist. The developed model could be used as an early warning tool to indicate the upcoming changes in water quality. This information would enable proactive operation, and the distribution of potable water with the required quality could be achieved. In addition, the overall costs of the water utility could be decreased as possible at least some of the manual sampling and laboratory analyses performed in the normal conditions could be replaced by the modelling approach monitoring.

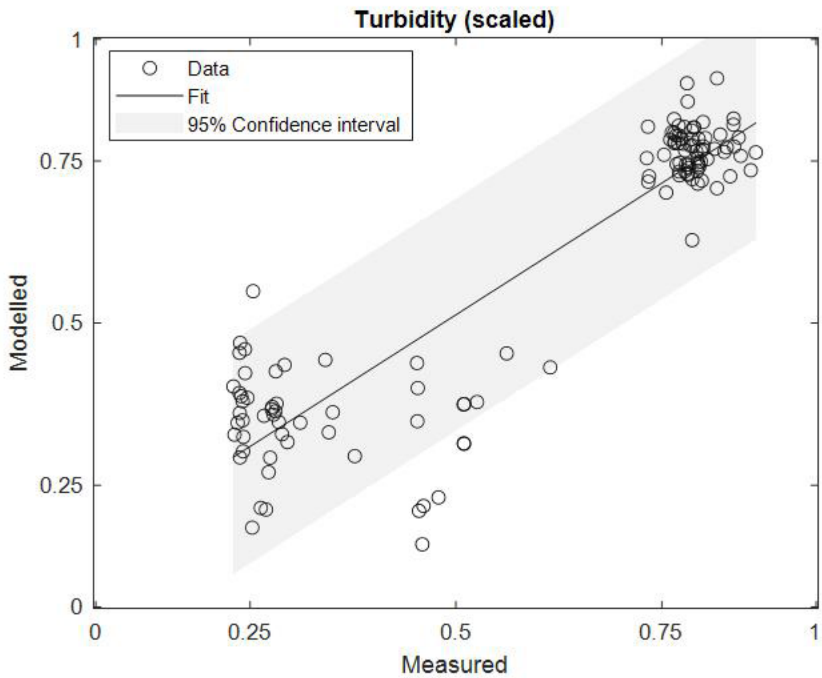

Figure 2. The measured and modelled turbidity in scaled values with 95% confidence interval.

Table I. The performance values, selected input variables, and the parameter values of the developed water turbidity model.

	Turbidity model
R^2	0.77
RMSE	0.18
$b+p_1*x_1 +...+ p_n*x_n$	$-2.13 + 2.15*F_1 + 2.91*P_2 + 0.33*P_3 - 0.42*P_4 - 0.33*P_5$

4. Conclusion

In this study, a water quality model for assessing the turbidity in the end part of the distribution network was developed using only the online flow and pressure measurements of a WDN as input variables. The online monitoring data were collected at several monitoring stations along the urban WDN during two separate periods. Based on the presented modelling result, it seems that the level of and changes in water turbidity can be estimated utilizing only the hydraulic condition information of the distribution network collected by existing online sensors. However, as the model uses only flow and turbidity measurements as inputs, any changes in turbidity caused by other factors (for instance, temperature) may not be picked up by the model which may affect the model performance.

Mounting and maintaining online sensors for water quality monitoring in old WDN is laborious and expensive. Utilizing the data from existing sensors via modelling approach does not increase the purchase, installation or maintenance costs of the water utility but could increase the essential information on the water quality and the state of the network in areas where no water quality measurements are available. On the contrary, the modelling approach could reduce the overall costs as in the best situation the number of reactive actions could be lower and possible some of the laborious, expensive and time-consuming manual sampling and laboratory analyses could be replaced by the model-based water quality information. As the data analysis also showed, the delays between the monitoring stations and the water quality monitoring point were several hours. Therefore, the information received from the model could be utilized also as an early warning tool of the changes in water quality contrary to the infrequent manual sampling and laboratory analyses.

Acknowledgments

This research was carried out as part of the Water-M project (ITEA 2). The research data is a property of Kuopion Vesi and it was downloaded from SaMi - Savonia Measurements site, and used in the research with a permission

References

Aisopou, A., Stoianov, I. & Graham, N.J.D. 2012 In-pipe water quality monitoring in water supply systems under steady and unsteady state flow conditions: A quantitative assessment. *Water Research*, 46, 235-246. doi: https://doi.org/10.1016/j.watres.2011.10.058

Anderson, C. W. 2005 Chapter A6.7 Turbidity. in Wilde, F.D. (ed.) U.S. Geological Survey, National field manual for the collection of water-quality data: U.S. *Geological Survey Techniques of Water-Resources Investigations*, book 9, chaps. A1-A10, available online at http://pubs.water.usgs.gov/twri9A

Arlot S., Celisse A. 2010 A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40–79.

Chapman, D. 1996 Water Quality Assessments - A Guide to Use of Biota, Sediments and Water in Environmental Monitoring. Chapman D. (ed.) – 2nd edn, *E&FN Spon, an imprint of Chapman & Hall*. Cambridge, Great Britain.

Clark, R.M. & Haught, R.C. 2005 Characterizing pipe wall demand: implications for water quality modelling. *Journal of Water Resources Planning and Management*, 131(3), 208–217.

Ghebremichael, K., Gebremeskel, A. & Trifunovic, N. 2008 Modelling disinfection by-products: coupling hydraulic and chemical models. *Water Science and Technology: Water Supply*, 8(3), 289–295.

Hastie, T., Tibshirani, R. & Friedman, J. 2009 The elements of statistical learning: Data mining, inference and prediction. Springer Series in Statistics. *Springer-Verlag*, New York. 2nd ed.

Lehtola, M.J., Laxander, M., Miettinen, I.T., Hirvonen, A., Vartiainen, T., Martikainen, P.J. 2006a. The effects of changing water flow velocity on the formation of biofilms and water quality in pilot distribution system consisting of copper or polyethylene pipes. *Water Research*, 40 (11), 2151–2160.

Lehtola, M.J., Miettinen, I.T., Hirvonen, A., Vartiainen, T., Martikainen, P.J. 2006b. Resuspension of biofilms and sediments to water from pipelines as a result of pressure shocks in drinking water distribution system. In: I*nternational Conference (IWA) Biofilm Systems VI*, Amsterdam, September 24–27, 2006, CD-Rom.

Li, C., Wang, D., Xu, X., Xu, M. & Wang, Z. 2017 Spatial variations in the occurrence of potentially genotoxic disinfection by-products in drinking water distribution systems in China. *Environmental Pollution*, 231(2), 1463-1468. doi: https://doi.org/10.1016/j.envpol.2017.09.008

Mains, C. 2008 Biofilm control in distribution systems. T*he national environmental services center at west Virginia university*, 8(2).

Manasfi, T., De Méo, M., Di Giorgio, C., Coulomb, B. & Boudenne, J-L. 2017 Assessing the genotoxicity of two commonly occurring byproducts of water disinfection: Chloral hydrate and bromal hydrate. *Mutation Research*, 813, 37–44. doi: https://doi.org/10.1016/j.mrgentox.2016.11.009

Meyers, G., Kapelan, Z., Keedwell, E. 2017 Short-term forecasting of turbidity in trunk main networks. *Water Research*, 124, 67-76. doi: https://doi.org/10.1016/j.watres.2017.07.035

Montgomery, D.C., Peck, E.A. & Vining, G.G. 2012 Introduction to linear regression analysis. 5th edition. *A JohnWiley & Sons*, Inc, Hoboken, New Jersey.

Mukundan, R., Pierson, D. C., Schneiderman, E. M., O'Donnell, D. M., Pradhanang, S. M., Zion, M. S., et al. 2013 Factors affecting storm event turbidity in a New York City water supply stream. *Catena*, 107, 80–88.

Mustonen, S.M., Tissari, S., Huikko, L., Kolehmainen, M., Lehtola, M.J. & Hirvonen, A. 2008 Evaluating online data of water quality changes in a pilot drinking water distribution system with multivariate data exploration methods. *Water Research*, 42, 2421-2430. doi: https://doi.org/10.1016/j.watres.2008.01.015

Seyoum, A.G., Tanyimboh, T.T. 2013 Pressure-dependent network water quality modelling. Water Management. *Proceedings of the Institution of Civil Engineers*. doi: https://doi.org/10.1680/wama.12.00118

Sunela, M.I., Puust, R. 2015 Real Time Water Supply System Hydraulic and Quality Modeling – A Case Study. *Procedia Engineering*, 119, 744-752. https://doi.org/10.1016/j.proeng.2015.08.928

Tomperi, J. 2020 Modelling the potable water quality in a distribution network based on the hydraulic conditions, *Drink. Water Eng. Sci. Discuss.* [preprint]. https://doi.org/10.5194/dwes-2020-20

Tomperi, J., Isokangas, A., Tuuttila, T., Paavola, M. 2022 Functionality of turbidity measurement under changing water quality and environmental conditions. *Environmental Technology,* 43:7, 1093-1101. doi: https://doi.org/10.1080/09593330.2020.1815860

Weston, S.L., Collins, R.P., Boxall, J.B. (2021) An experimental study of how hydraulic transients cause mobilisation of material within drinking water distribution systems. *Water Research*, 194. https://doi.org/10.1016/j.watres.2021.116890

Vreeburg, I.J.H.G., Boxall, D.J.B. 2007 Discolouration in potable water distribution systems: a review. Water Research, 41, 519-529. doi: https://doi.org/10.1016/j.watres.2006.09.028

World Health Organization, 2008. Guidelines for drinking-water quality: incorporating 1st and 2nd addenda, Vol. 1, Recommendations, 3rd Edn, World Health Organization, Geneva.