Water Research 109 (2017) 144–154

Contents lists available at ScienceDirect

Water Research

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

Bayesian belief network modelling of chlorine disinfection for human pathogenic viruses in municipal wastewater

Guido Carvajal [a], David J. Roser [a], Scott A. Sisson [b], Alexandra Keegan [c], Stuart J. Khan [a, *]

[a] UNSW Water Research Centre, School of Civil & Environmental Engineering, University of New South Wales, NSW 2052, Australia
[b] School of Mathematics & Statistics, University of New South Wales, NSW 2052, Australia
[c] Research and Innovation Services, SA Water Corporation, Adelaide, SA 5000, Australia

ARTICLE INFO

Article history:
Received 17 May 2016
Received in revised form
7 October 2016
Accepted 2 November 2016
Available online 3 November 2016

Keywords:
Bayesian belief networks
Chlorination
Modelling
Multilayer perceptron
Removal efficiency
Virus

ABSTRACT

Chlorine disinfection of biologically treated wastewater is practiced in many locations prior to environmental discharge or beneficial reuse. The effectiveness of chlorine disinfection processes may be influenced by several factors, such as pH, temperature, ionic strength, organic carbon concentration, and suspended solids. We investigated the use of Bayesian multilayer perceptron (BMLP) models as efficient and practical tools for compiling and analysing free chlorine and monochloramine virus disinfection performance as a multivariate problem. Corresponding to their relative susceptibility, Adenovirus 2 was used to assess disinfection by monochloramine and Coxsackievirus B5 was used for free chlorine. A BMLP model was constructed to relate key disinfection conditions (CT, pH, turbidity) to observed Log Reduction Values (LRVs) for these viruses at constant temperature. The models proved to be valuable for incorporating uncertainty in the chlor(am)ination performance estimation and interpolating between operating conditions. Various types of queries could be performed with this model including the identification of target CT for a particular combination of LRV, pH and turbidity. Similarly, it was possible to derive achievable LRVs for combinations of CT, pH and turbidity. These queries yielded probability density functions for the target variable reflecting the uncertainty in the model parameters and variability of the input variables. The disinfection efficacy was greatly impacted by pH and to a lesser extent by turbidity for both types of disinfections. Non-linear relationships were observed between pH and target CT, and turbidity and target CT, with compound effects on target CT also evidenced. This work demonstrated that the use of BMLP models had considerable ability to improve the resolution and understanding of the multivariate relationships between operational parameters and disinfection outcomes for wastewater treatment.

1. Introduction

Disinfection dose requirements for water and wastewater treatment are conventionally expressed as the product of disinfectant concentration and contact time (CT), required to achieve a predetermined reduction in microbial numbers. Achievement of target CT values is dependent upon meeting various factors for each pathogen type, including pH, temperature, disinfectant concentration, ionic strength, and suspended particles (Jensen et al., 1980; LeChevallier and Au, 2004). Consequently, assigning log reduction value (LRV) credits depends on optimising such variables as well as ensuring the primary CT product.

Previous work has sought to establish CT values necessary to achieve LRVs for various pathogens in biologically treated (activated sludge) municipal wastewater (Keegan et al., 2012). The products of that research included a number of linear models and tables which related target CTs to various pH and turbidity combinations (DOH, 2013). This approach to deriving and representing CT-LRV relationships is in line with international best practice for defining the disinfection CT requirements for drinking water (USEPA, 2003). However, as the number of disinfection controlling factors increases, interpolating values between experimental data points and communicating the information in tabular form or as linear models becomes increasingly problematic. Further, the increasing popularity of Quantitative Microbial Risk Assessment

* Corresponding author.
E-mail addresses: guido.carvajal@unsw.edu.au (G. Carvajal), djroser@unsw.edu.au (D.J. Roser), scott.sisson@unsw.edu.au (S.A. Sisson), Alex.Keegan@sawater.com.au (A. Keegan), s.khan@unsw.edu.au (S.J. Khan).

http://dx.doi.org/10.1016/j.watres.2016.11.008

(US-EPA and USDA/FSIS, 2012) means that in future, LRV point estimates may not be sufficient, and measures of model uncertainty and variability will be needed.

Bayesian belief networks (BBNs) offer an alternative approach for relating chlorination LRVs to CT and wastewater quality parameters and incorporate parameter uncertainty and variability. They also offer a convenient means for performing scenario exploration and inference, and hence prediction of disinfection performance under diverse conditions. Continuous BBNs are models which involve the use of continuous variables without the need for discretisation. A BBN's structure is defined by directional connections, known as ‘arcs’, which specify the dependence and conditional independence assumptions i.e. relationships, between random variables, which in BBNs are termed ‘nodes’. These interdependencies in turn determine what information is required to specify the joint probability distribution among the random variables of a network. Through the directed acyclic graph structure, BBNs reduce the quantity of information required to define a joint probability distribution (Korb and Nicholson, 2011).

To facilitate interpolation and CT estimation and prediction, in this study we investigated the use of Bayesian multilayer perceptron (BMLP) models to derive continuous relationships between virus LRV, pH, turbidity and CT, and perform interpolation considering uncertainty in the model parameters. The Bayesian integration within the BMLP model transforms this model into a BBN, introducing features such as stochastic representation of the parameters and predictions, as well as computation of queries on target variables given a set of observations. A multilayer perceptron (MLP) model is a type of neural network composed of layers of neurons (elements that generate a transformation of the inputs) with an input layer, at least one hidden layer, and an output layer (Priddy and Keller, 2005). In MLP models the inputs of the neurons in one layer come from the outputs of neurons in a previous layer. Neurons in one layer are connected to the previous layer through weighted connections. These models can solve non-linear problems and perform prediction with high accuracy in multivariate settings.

In this study we assessed the application of a continuous BBN for estimating chlorination and chloramination LRVs for human virus removal during wastewater treatment, while accounting for the influence of pH, turbidity and CT variance. The broader aim of the study was to investigate the utility of BBNs for quantifying the effectiveness of chlorination of treated wastewater in a multivariate context and to present a simple and practical tool for interpolation of CT values and incorporation of uncertainty and variability.

2. Materials and methods

2.1. Data extraction and model construction

Chlorine disinfection data were obtained from a previously published project report of bench-scale batch experiments using secondary treated wastewater from the Bolivar Wastewater Treatment Plant in Adelaide, South Australia (Keegan et al., 2012). This wastewater was seeded with two viruses, Coxsackievirus B5 for estimating free chlorine LRVs, and Adenovirus 2 for estimating monochloramine LRVs. These viruses were selected since they are known to exhibit high resistance to chlorine and monochloramine inactivation respectively (Liu et al., 1971; Payment et al., 1985). Two parameters, pH and turbidity, were varied to determine CT values for virus inactivation under a range of conditions.

Inactivation experiments were conducted at pH 7.0, 8.0 and 9.0, and at three turbidity values (2, 5, 20 NTU), at a constant low (conservative) temperature of 10 °C. Both viruses were seeded at concentrations of *ca* 10^5 pfu/mL to allow measurement of up to at least 4 logs inactivation. The datasets produced by Keegan et al. (2012) consisted of 226 records for Adenovirus and 154 records for Coxsackievirus. CT values were calculated from the integrals (areas under the curves) of residual free and combined chlorine concentrations vs. sample contact time. Unlike the original study, in the present analysis the disinfection conditions and corresponding LRV data were not obtained from fitted linear models, but by using the raw replicated LRV measurements obtained in that study. The aim for Keegan et al. (2012) was to construct CT tables for specific whole-number LRV values (1, 2, 3 and 4) in line with conventional past practice (USEPA, 2003). In the work presented here, we describe the construction of a BBN, which can produce a target CT value for any combination of input variables.

Prior to analysis, the raw data (Supplementary Material) were extracted and arranged in a table (.CSV file) with columns representing variables and rows presenting the experimental cases. This arrangement was used to facilitate importation to *R* programming (R-project 2014) environment and BMLP model construction. The process of model construction and parameter definition was conducted in *R* using Jags through the freely available *R2jags* package (Plummer, 2013).

The two overall structurally identical but parametrically different models, one for each virus, represented the conventional procedure followed by the US Environment Protection Agency (USEPA, 2003). The models were constructed considering three continuous variables, target LRV, pH and turbidity as predictors of target CT. The BBN model is represented in Fig. 1, which shows the choice of distributions for each identified node using Jags nomenclature. Explanation of each variable in the model is provided in Table 1.

Other water quality parameters may also affect chlorination LRV. However, such variables were controlled throughout these experiments by the use of the same biologically treated wastewater matrix in all experiments. The model thus assumes that target LRV, pH and turbidity are conditionally dependent given target CT. This means that without knowing the value of the target CT, a modification to any of these three nodes will not produce any change to the other two.

A BMLP model was employed to capture the relationships between the variables. BMLP model use enabled consideration of pH and turbidity as continuous variables and therefore facilitated interpolation of values intermediate with the selected experimental settings. In this model, the error estimates in the target CT were assumed normally distributed. Normal distribution was also preferred due to its simplicity. This assumption has been previously used to derive confidence intervals in the predicted outcomes of multilayer perceptron (Chryssolouris et al., 1996; He and Li, 2011). That is

$$CT \sim N\left(\mu; \sigma^2\right) \quad (1)$$

where μ is the predicted dependant value (target CT) as a function of the target LRV, turbidity and pH, and σ is the standard deviation of the model errors. Unlike the conventional neural network approach, the BMLP model considers uncertainty in the model parameters which introduces an additional source of uncertainty for the CT value estimation. The resulting model is presented in section 2.3. Inactivation ratio (IR) and actual LRV (LRV.act) are variables which were computed from the outcomes of stochastic variables using Eqs. (2) and (3) respectively (USEPA, 2003).

$$\mathrm{IR} = \frac{\mathrm{CT.obs}}{\mathrm{CT.target}} \quad (2)$$