

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Dissolved oxygen prediction using a possibility-theory based fuzzy neural network

U. T. Khan and C. Valeo

Mechanical Engineering, University of Victoria, P.O. Box 1700, Stn. CSC,
Victoria, BC, V8W 2Y2, Canada

Received: 16 October 2015 – Accepted: 13 November 2015 – Published: 26 November 2015

Correspondence to: C. Valeo (valeo@uvic.ca)

Published by Copernicus Publications on behalf of the European Geosciences Union.

12311

Abstract

A new fuzzy neural network method to predict minimum dissolved oxygen (DO) concentration in a highly urbanised riverine environment (in Calgary, Canada) is proposed. The method uses abiotic (non-living, physical and chemical attributes) as inputs to the model, since the physical mechanisms governing DO in the river are largely unknown. A new two-step method to construct fuzzy numbers using observations is proposed. Then an existing fuzzy neural network is modified to account for fuzzy number inputs and also uses possibility-theory based intervals to train the network. Results demonstrate that the method is particularly well suited to predict low DO events in the Bow River. Model output and a defuzzification technique is used to estimate the risk of low DO so that water resource managers can implement strategies to prevent the occurrence of low DO.

1 Introduction

The City of Calgary is a major economic hub in western Canada. With a rapidly growing population, currently estimated in excess of 1 million, the City is undergoing expansion and urbanisation to accommodate the changes. The Bow River is a relatively small river (with an average annual flow of $90 \text{ m}^3 \text{ s}^{-1}$, with an average width of 100 m, and depth of 1.5 m) that flows through the City and provides approximately 60% of the residents with potable water (Khan and Valeo, 2015a, b). In addition to this, water is diverted from within the City for irrigation, is used as a source for commercial and recreational fisheries, and is the source of drinking water for communities downstream of Calgary (Robinson et al., 2009; Bow River Basin Council, 2015). This highlights the importance of the Bow River, not just as a source of potable water, but also as a major economic resource.

However, urbanisation has the potential to reduce the health of the Bow River, which is fast approaching its assimilative capacity and is one of the most regulated rivers in

12312

Alberta (Bow River Basin Council, 2015). Three wastewater treatment plants (shown in Fig. 1) and numerous stormwater outfalls discharge their effluent into the River and are considered to be a major cause of water quality degradation in the River (He et al., 2015). This highlights some of the major impacts on the Bow River from the surrounding urban area. A number of municipal and provincial programs are in place to reduce the loading of nutrients and sediments into the river such as the Total Loadings Management Plan and the Bow River Phosphorus Management Plan (Neupane et al., 2014) as well as modelling efforts – namely the Bow River Water Quality Model (Tetra Tech, 2013; Golder, 2004) – to predict the impact of different water management programs on the water quality.

One of the major concerns is that low dissolved oxygen (DO) concentration has occurred on a number of occasions over the last decade in the Bow River within the City limits. DO is an indicator of overall health of the aquatic ecosystem (Dorfman and Jacoby, 1972; Hall, 1984; Canadian Council of Ministers of the Environment, 1999; Kannel et al., 2007; Khan and Valeo, 2014a, 2015a), and low DO – which can be caused by a number of different factors (Pogue and Anderson, 1995; Hauer and Hill, 2007; He et al., 2011; Wen et al., 2013) – can impact various organisms in the waterbody. While the impact of long-term effects of low DO are largely unknown, acute events can have devastating effects on aquatic ecosystems (Adams et al., 2013). Thus, maintaining a suitably high DO concentration, and water quality in general, is of utmost importance to the City of Calgary and downstream stakeholders, particularly as the City is being challenged to meet its water quality targets (Robinson et al., 2009).

A number of recent studies have examined the DO in the Bow River, and the factors that impact its concentration. Iwanyshyn et al. (2008) found the diurnal variation in DO and nutrient (nitrate and phosphate) concentration was highly correlated, suggesting that biogeochemical processes (photosynthesis and respiration of aquatic vegetation) had a dominant impact on nutrient concentration rather than wastewater treatment effluent. Further, Robinson et al. (2009) found that the DO fluctuations in the River were primarily due to periphyton rather macrophyte biogeochemical processes. In both stud-

12313

ies, the seasonality of DO, nutrients, and biological concentration, and external factors (e.g. flood events) were demonstrative of the complexity in understanding river processes in an urban area, and that consideration of various inputs, outputs and their interaction is important to fully understand the system. He et al. (2011) found that seasonal variations in DO in the Bow River could be explained by a combination of abiotic factors (such as climatic and hydrometric conditions), as well as biotic factors. The study found that while photosynthesis and respiration of biota are the main drivers of DO fluctuation, the role of nutrients (from both point and non-point sources) was ambiguous. Neupane et al. (2014) found that organic materials and nutrients from point and non-point sources influence DO concentration in the River. The likelihood of low DO was highest downstream of wastewater treatment plants, and that non-point sources have a significant impact in the open-water season. Using a physically-based model, Neupane et al. (2014) predicted low DO concentration more frequently in the future in the Bow River owing to higher phosphorus concentration in the water, as well as impacts of climate change.

A major issue of modelling DO in the Bow River is that rapid urbanisation within the watershed has resulted in substantial changes to land-use characteristics, sediment and nutrient loads, and to other factors that govern DO. Major flood events (like those in 2005 and 2013) completely alter the aquatic ecosystem, while new wastewater treatment plants (e.g. the Pine Creek wastewater treatment plant) added in response to the growing population further increases the stress downstream. These types of changes in a watershed increase the complexity of the system, making DO trends and variability more challenging to model. The interaction of numerous factors, over a relatively small area and across different temporal scales means that DO trends in urban areas are more difficult to predict and evaluating water quality in urban riverine environments is a difficult task (Hall, 1984; Niemczynowicz, 1999).

The implication of this is that the simplistic representation described in conceptual, physically-based models is not suitable for complex systems, i.e. where the underlying physical mechanisms behind the factors that govern DO are still not clearly under-

12314

stood, in a rapidly changing urban environment. Physically-based models require the parameterisation of a several different variables which may be unavailable, expensive and time consuming (Antanasijević et al., 2014; Wen et al., 2013; Khan et al., 2013). In addition to this, the increase in complexity in an urban system proportionally increases the uncertainty in the system. This uncertainty can arise as a result of vaguely known relationships among all the factors that influence DO, in addition to the inherent randomness in the system (Deng et al., 2011). The rapid changes in an urban area render the system *dynamic* as opposed to *stationary*, which is what is typically assumed for many probability-based uncertainty quantification methods. Thus, not only is DO prediction difficult, it is beset with uncertainty, hindering water resource managers from making objective decisions.

In this research, we propose a new method to predict DO concentration in the Bow River using a data-driven approach, as opposed to a physically-based method, that uses *possibility* theory and fuzzy numbers to represent the uncertainty rather than the more commonly used probability theory. Data-driven models are a class of numerical models based on generalised relationships, links or connections between input and output datasets (Solomatine and Ostfeld, 2008). These models can characterize a system with limited assumptions and are useful in solving practical problems, especially when there is lack of understanding of the underlying physical process, the time series are of insufficient length, or when existing models are inadequate (Solomatine et al., 2008; Napolitano et al., 2011).

1.1 Fuzzy numbers and data-driven modelling

Possibility theory is an information theory that is an extension of fuzzy sets theory for representing uncertain, vague or imprecise information (Zadeh, 1978). Fuzzy numbers are an extension of fuzzy set theory, and express an uncertain or imprecise *quantity*. These types of numbers are particularly useful for dealing with uncertainties when data are limited or imprecise (Bárdossy et al., 1990; Guyonnet et al., 2003; Huang et al., 2010; Zhang and Achari, 2010) – in other words when epistemic uncertainty

12315

exists. This type of uncertainty is in contrast to aleatory uncertainty that is typically handled using probability theory. Possibility theory and fuzzy numbers are thus useful when a probabilistic representation of parameters may not be possible, since the exact values of parameters may be unknown, or only partial information is available (Zhang, 2009). Thus, the choice of using a data-driven approach in combination with possibility theory lends itself well to the constraints posed by the problem in the Bow River: the difficulty in correctly defining a physically-based model for a complex urban system and the use of possibility theory to model the uncertainty in the system when probability theory based methods may be inadequate.

Data-driven models, such as neural networks, regression-based techniques, fuzzy rule-based systems, and genetic programming, have seen widespread use in hydrology, including DO prediction in rivers (Shrestha and Solomatine, 2008; Solomatine et al., 2008; Elshorbagy et al., 2010). Wen et al. (2011) used artificial neural networks (ANN) to predict DO in a river in China using ion concentration as the predictors. Antanasijević et al. (2014) used ANNs to predict DO in a river in Serbia using a Monte Carlo approach to quantify the uncertainty in model predictions and temperature as a predictor. Chang et al. (2015) also used ANNs coupled with hydrological factors (such as precipitation and discharge) to predict DO in a river in Taiwan. Singh et al. (2009) used water quality parameters to predict DO and BOD in a river in India. Other studies (e.g. Heddiam, 2014; Ay and Kisi, 2012) have used regression to predict DO in rivers using water temperature, or electrical conductivity, amongst others, as inputs. In general, these studies have demonstrated that there is a need and demand for less complex DO models, has led to an increase in the popularity of data-driven models (Antanasijević et al., 2014), and that the performance of these models is suitable. Recent research into predicting DO concentration in the Bow River in Calgary using abiotic factors (these are non-living, physical and chemical attributes) as inputs have shown promising results (He et al., 2011; Khan et al., 2013; Khan and Valeo, 2015a). The advantage of using readily available data (i.e. the abiotic inputs) in these studies is that if a suitable relationship between these factors and DO can be found, changing

12316

the factors (e.g. increasing the discharge rate downstream of a treatment plant) can potentially reduce the risk of low DO.

While fuzzy set theory based applications, particularly applications using fuzzy *logic* in neural networks, have been widely used in many fields including hydrology (Bárdossy et al., 2006; Abrahart et al., 2010), the use of fuzzy *numbers* and possibility theory based applications has been limited in comparison (Bárdossy et al., 2006; Jacquin, 2010). Some examples include maps of soil hydrological properties (Martin-Clouaire et al., 2000), remotely sensed soil moisture data (Verhoest et al., 2007), climate modelling (Mujumdar and Ghosh, 2008), subsurface contaminant transport (Zhang et al., 2009), and streamflow forecasting (Alvisi and Franchini, 2011). Khan et al. (2013) and Khan and Valeo (2015a) have introduced a fuzzy number based *regression* technique to model daily DO in the Bow River using abiotic factors with promising results. Similarly, Khan and Valeo (2014a) used an autoregressive time series based approach combined with fuzzy numbers to predict DO in the Bow River. In these studies the use of fuzzy numbers meant that the uncertainty in the system could be quantified and propagated through the model. However, due to the highly non-linear nature of DO modelling, the use of an ANN based method is of interest since these types of models are effective for modelling complex, nonlinear relationships without the explicit understanding of the physical phenomenon governing the system. A fuzzy neural network method proposed by Alvisi and Franchini (2011) for streamflow prediction that uses fuzzy weights and biases in the network, is further refined in this research for predicting DO concentration.

1.2 Objectives

Given the importance of DO concentration as an indicator of overall aquatic ecosystem health, there is a need to accurately model and predict DO in urban riverine environments, like that in Calgary, Canada. In this research a new data-driven method is proposed that attempts to address the issues that plague numerical modelling of DO concentration in the Bow River. The FNN method proposed by Alvisi and Franchini (2011) is adapted and extended in two critical ways. The existing method uses crisp (i.e. non-fuzzy) inputs and outputs to train the network, producing a set of fuzzy number weights and biases, and fuzzy outputs. The method is adapted to be able to handle fuzzy number inputs to produce fuzzy weights and biases, and fuzzy outputs. The advantage is that the uncertainties in the input observations are also captured within the model structure. To do this, a new method of creating fuzzy numbers from observations is presented based on a probability-possibility transformation. Second, the existing training algorithm is based on capturing a predetermined set of observations (e.g. 100, 95 or 90 %) within the fuzzy outputs. The selection of the predetermined set of observations in the original study was an arbitrary selection. A new method that exploits the relationship between possibility theory and probability theory is defined to create a more objective method of training the FNN. A consequence of this is that the resulting fuzzy number outputs from the model can then be directly used for risk analysis, specifically to quantify the risk of low DO concentration. This information is extremely valuable for managing water resources in the face of uncertainty. Following previous research for this river, two abiotic inputs (daily mean water temperature, T and daily mean flow rate, Q) will be used to predict daily minimum DO. An advantage of using these factors is that they are routinely collected by the City of Calgary, and thus, a large dataset is available. Also, their use in previous studies has shown that they are good predictors of daily DO concentration in this river basin (He et al., 2011; Khan et al., 2013; Khan and Valeo, 2015a). The following sections outline the background of fuzzy numbers and existing probability-possibility transformations. This is followed by the development of the new method to create fuzzy numbers from observations. Then, the new FNN method using fuzzy inputs is developed mathematically using new criteria for training, also based on possibility theory. Lastly, a method to measure the risk of low DO is described.

chini (2011) is adapted and extended in two critical ways. The existing method uses crisp (i.e. non-fuzzy) inputs and outputs to train the network, producing a set of fuzzy number weights and biases, and fuzzy outputs. The method is adapted to be able to handle fuzzy number inputs to produce fuzzy weights and biases, and fuzzy outputs. The advantage is that the uncertainties in the input observations are also captured within the model structure. To do this, a new method of creating fuzzy numbers from observations is presented based on a probability-possibility transformation. Second, the existing training algorithm is based on capturing a predetermined set of observations (e.g. 100, 95 or 90 %) within the fuzzy outputs. The selection of the predetermined set of observations in the original study was an arbitrary selection. A new method that exploits the relationship between possibility theory and probability theory is defined to create a more objective method of training the FNN. A consequence of this is that the resulting fuzzy number outputs from the model can then be directly used for risk analysis, specifically to quantify the risk of low DO concentration. This information is extremely valuable for managing water resources in the face of uncertainty.

Following previous research for this river, two abiotic inputs (daily mean water temperature, T and daily mean flow rate, Q) will be used to predict daily minimum DO. An advantage of using these factors is that they are routinely collected by the City of Calgary, and thus, a large dataset is available. Also, their use in previous studies has shown that they are good predictors of daily DO concentration in this river basin (He et al., 2011; Khan et al., 2013; Khan and Valeo, 2015a). The following sections outline the background of fuzzy numbers and existing probability-possibility transformations. This is followed by the development of the new method to create fuzzy numbers from observations. Then, the new FNN method using fuzzy inputs is developed mathematically using new criteria for training, also based on possibility theory. Lastly, a method to measure the risk of low DO is described.

2 Methods

2.1 Data collection

The Bow River is 645 km long and averages a 0.4 % slope over its length (Bow River Basin Council, 2015) from its headwaters at Bow Lake in the Rocky Mountains to its confluence with the Oldman River in Southern Alberta, Canada (Robinson et al., 2009; Environment Canada, 2015). The river is supplied by snowmelt from the Rocky Mountains, rainfall and discharge from groundwater. The City of Calgary is located within the Bow River Basin and the river has an average annual discharge of $90 \text{ m}^3 \text{ s}^{-1}$, an average width and depth of 100 and 1.5 m, respectively (Khan and Valeo, 2014b, 2015b).

The City of Calgary routinely samples a variety of water quality parameters along the Bow River to measure the impacts of urbanisation, particularly from three wastewater treatment plants and numerous stormwater runoff outfalls that discharge into the River. DO concentration measured upstream of the City is generally high throughout the year, with little diurnal variation (He et al., 2011; Khan et al., 2013; Khan and Valeo, 2015a). The DO concentration downstream of the City is lower and experiences much higher diurnal fluctuation. The three wastewater treatment plants are located upstream of this monitoring site, and are thought to be responsible, along with other impacts of urbanisation, for the degradation of water quality (He et al., 2015).

For this research, nine years of DO concentration data was collected from one of the downstream stations from 2004 to 2012. The monitoring station was located at Pine Creek and sampled water quality data every 30 min (from 2004 to 2005), and every 15 min (from 2006 to 2007). The station was then moved to Stier's Ranch and sampled data every hour (in 2008) and every 15 min (2009 to 2011). The monitoring site was moved further downstream to its current location (at Highwood) in 2012 where it samples every 15 min. During this period a number of low DO events have been observed in the River and are summarised below in Table 1 corresponding to different water quality guidelines.

12319

Note that even though daily minimum DO was observed to be below 5 mgL^{-1} on several occasions in 2004 and 2006 (in Table 1), the minimum DO was below 9.5 mgL^{-1} only 107 and 164 days, respectively, for those two years. In contrast, in 2007 and 2010, no observations below 5 mgL^{-1} are seen yet 182 and 180 days, respectively, below the 9.5 mgL^{-1} guideline were seen for those years. The total amount of days below 9.5 mgL^{-1} constitute approximately 90 % of all observations for those years. This highlights that despite no DO events below 5 mgL^{-1} , generally speaking minimum DO on a daily basis was quite low in these two years. The implication of this is that only using one guideline for DO might not be a good indicator of overall aquatic ecosystem health.

A YSI sonde is used to monitor DO and T , and the sonde is not accurate in freezing water, thus only data from the ice free period was considered, which is approximately from April to October for most years (YSI Inc., 2015). Since low DO events usually occur in the summer (corresponding to high water temperature and lower discharge), the ice-free period dataset contains the dates that are of interest for low DO modelling.

Daily mean flow rate, Q , was collected from the Water Survey of Canada site "Bow River at Calgary" (ID: 05BH004) for the same period. This data is collected hourly throughout the year, thus, data where considerable shift corrections were applied (usually due to ice conditions) were removed from the analysis. The mean annual water temperature ranged between 9.23 and 13.2 °C, and the annual mean flow rate was between 75 and $146 \text{ m}^3 \text{ s}^{-1}$, for the selected period.

2.2 Probability-possibility transformations

Fuzzy sets were proposed by Zadeh (1965) in order to express imprecision in complex systems, and can be described as a generalisation of classical set theory (Khan and Valeo, 2015a). In classical set theory, an element x either belongs or does not belong to a set A . In contrast, using fuzzy set theory, the elements x have a degree of membership, μ , between 0 and 1 in the fuzzy set A . If μ equals 0, then x does not belong in

12320

A , and $\mu = 1$ means that it completely belongs in A , while a value $\mu = 0.5$ means that it is only a partial member of A .

Fuzzy numbers express uncertain or imprecise quantities, and represent the set of all possible values that define a quantity rather than a single value. A fuzzy number is defined as a specific type of fuzzy set: a normal and convex fuzzy set. Normal implies that there is at least one element in the fuzzy set with a membership level equal to 1, while convex means that the membership function increases monotonically from the lower support (i.e., $\mu = 0^L$) to the modal element (i.e. the element(s) with $\mu = 1$) and then monotonically decreases to the upper support (i.e., $\mu = 0^R$) (Kaufmann and Gupta, 1985).

Traditional representation of a fuzzy numbers has been using symmetrical, linear membership functions, typically denoted as triangular fuzzy numbers. The reason for selecting this type of membership function has to do with its simplicity: given that a fuzzy number must, by definition, be convex and normal, a minimum of three elements are needed to define a fuzzy number (two elements at $\mu = 0$ and one element at $\mu = 1$). For example, if the most credible value for DO concentration is 10 mgL^{-1} ($\mu = 1$), with a support about the modal value between ($\mu = 0^L$) and 12 mgL^{-1} ($\mu = 0^R$). This implies that the simplest membership function is triangular, though not necessarily symmetrical. Also, as we demonstrate below, in some probability-possibility frameworks, a triangular membership function corresponds to a uniform probability distribution – the least specific distribution in that any value is equally probable and hence, represents the most uncertainty (Dubois and Prade, 2015; Dubois et al., 2004).

However, recent research (Khan et al., 2013; Khan and Valeo, 2014a, b, 2015a, b) has shown that such a simplistic representation may not be appropriate for hydrological data, which is often skewed, and non-linear. This issue is further highlighted if the probability-possibility framework mentioned above is used: it implies that for a triangular membership function, the fuzzy number bounded by the support $[8 \text{ } 12] \text{ mgL}^{-1}$, has a uniform probability distribution bounded between 8 and 12 mgL^{-1} with a mean value of 10 mgL^{-1} , suggesting that values between the support are equally likely to occur. It

12321

not difficult to see that this an over-simplification of hydrological data, which often have skewed, non-symmetrical distributions. In many cases enough information (i.e. from observations) is available to define the membership function with more specificity, and this information should be used to define the membership function.

Multiple frameworks exist to transform a probability distribution to a possibility distribution, and vice versa; a comparison of different conceptual approaches are provided in Klir and Parvais (1992), Oussalah (2000), Jaquin (2010), Mauris (2013) and Dubois and Prade (2015). However, a major issue of implementing fuzzy number based methods in hydrology is that there is no consistent, transparent and objective method to convert observations (e.g. time series data) into fuzzy numbers, or generally speaking to construct the membership function associated with fuzzy values (Abrahart et al., 2010; Dubois and Prade, 1993; Civanlar and Trussel, 1986).

A popular method (Dubois et al., 1993, 2004) converts a probability distribution to a possibility distribution by relating the area under a probability density function to the membership level (Zhang, 2009). From this point of view, the possibility is viewed as the upper envelope of the family of probability measures (Jacquin, 2010; Ferrero et al., 2013; Betrie et al., 2014). There are two important considerations for this transformation, first it guarantees that something must be possible before it is probable; hence, the degree of possibility cannot be less than the degree or probability – this is known as the consistency principle (Zadeh, 1965). Second is order preservation, which means if the possibility of x_i is greater than the possibility of x_j then the probability of x_i must be greater than the probability of x_j (Dubois et al., 2004). For a discrete system, this can be represented as:

$$\text{if } p(x_1) > p(x_2) > \dots > p(x_n),$$

then the possibility distribution of x ($\pi(x)$), follows the same order, that is:

$$\pi(x_1) > \pi(x_2) > \dots > \pi(x_n).$$

The transformation is given by:

12322

ements to have equal $\mu = 1$. The proposed algorithm is described in the proceeding section.

A new algorithm to create fuzzy numbers

Shimazaki and Shinomoto (2007) proposed a method to find the optimum bin-size of a histogram when the underlying distribution of the process is unknown. The basic premise of the method is that the optimum bin-size (D_{opt}) is one that minimises the error between the theoretical (but unknown) probability density and the histogram generated using the D_{opt} . The error metric used by Shimazaki and Shinomoto (2007) is the mean integrated squared error (E_{MISE}) which is frequently used for density estimation problems. It is defined as:

$$E_{MISE} = \frac{1}{P} \int_0^P E[f_n(t) - f(t)]^2 dt \quad (2)$$

where $f(t)$ is the unknown density function, $f_n(t)$ is the histogram estimate of the density function, t denotes time and P is the observation period, and $E[\cdot]$ is the expectation. In practice, E_{MISE} cannot be directly calculated since the underlying distribution is unknown and thus, an estimate of the E_{MISE} is used in its place (see C_D below). Thus, $f_n(t)$ can be found without any assumptions of the type of distribution (e.g. class, unimodality, etc.); the only assumption is that the number of events (i.e. the counts k_i) in the i th bin of the histogram follow a Poisson point process. This means that the events in two disjoint bins (e.g. the i th and $i + 1$ th bin) are independent, and that mean (k) and variance (v) of the k_i in each bin are equal, due to the assumption of a Poisson process (Shimazaki and Shinomoto, 2007).

Using this property, the optimum bin-size can be found as follows. Let X be the input data vector for the observation period (P), e.g., a $[24 \times 1]$ vector corresponding to hourly samples for a given day. The elements in X are binned into N bins of equal bin-size

12325

D . The number of events k_i in each i th bin are then counted and the mean (k) and variance (v) of the k_i are calculated as follows:

$$k = \frac{1}{N} \sum_{i=1}^N k_i \quad (3)$$

$$v = \frac{1}{N} \sum_{i=1}^N (k_i - k)^2. \quad (4)$$

The k and v are then used to compute the *cost-function* C_D , which is defined as:

$$C_D = \frac{2k - v}{D^2}. \quad (5)$$

This cost-function is a variant of the original E_{MISE} listed in Eq. (2) and is derived by removing the terms from E_{MISE} that are independent of the bin-size D , and by replacing the unobservable quantities (i.e. $E[f(t)]$) with their unbiased estimators (details of this derivation can be found in the original paper by Shimazaki and Shinomoto, 2007). The objective then is to search for D_{opt} : the value of D that minimises C_D . To do this two systematic modification are made: first, C_D is recalculated at different *partitioning positions*, and secondly, the entire process is repeated for different values of N and D , until a “reliable” estimate of minimum C_D and thus D_{opt} is found. Using different partitioning positions means that the variability in k_i resulting from the position of the bin (rather than the size of the bin) can be quantified. Repeating the analysis at different N and D accounts for the variability due to different bin-sizes. Both these techniques are ways of accounting for the uncertainty associated with estimating the histogram.

Partitioning positions are defined as the first and last point that define a bin. The most common way of defining a partitioning position is to centre it on some value a , e.g. the bin defined at $[a - D/2, a + D/2]$ is centred on a and has a bin-size D . Variations of this partitioning position can be found by using a moving-window technique, where the

12326

bin-size D is kept constant, but the first and last points are perturbed by a small value δ : $[a - D/2 + \delta, a + D/2 + \delta]$, where δ ranges incrementally between 0 and D . Using these different values of δ whilst keeping D constant will result in different values of k_i and hence unique values of C_D . Thus, for a single value of D , multiple values of C_D are possible.

For this research this bin-size optimisation algorithm is implemented to determine the optimum histogram for the two input variables, Q and T . The array of daily data, \mathbf{X} , (at hourly or higher frequency, see Sect. 2.1 for details regarding the sampling frequency of both inputs) for each variable was collected for the nine year period. The bin-size was calculated for each day as follows:

$$D = \frac{X_{\max} - X_{\min}}{N} \quad (6)$$

where the x_{\max} and x_{\min} are the maximum and minimum sampled values for \mathbf{X} , respectively, and N is the number of bins. As described above, a number of different D were considered to find the optimum C_D . This was done by selecting a number of different values of N , ranging from N_{\min} to N_{\max} . The minimum value N_{\min} , was set at 3 for all days; this is the necessary number of bins to define a fuzzy number (two elements for $\mu = 0$, and one element for $\mu = 1$). The highest value, N_{\max} was calculated as:

$$N = \frac{X_{\max} - X_{\min}}{2r} \quad (7)$$

where $2r$ is the measurement resolution of the device used to measure either Q or T , set at twice the accuracy (r) of the device. The rationale for this decision is that as N increases D necessarily decreases (as per Eq. 6). However, D cannot be less than the measurement resolution; this constraint (i.e. $N \leq N_{\max}$) ensures that the optimum bin-size is never less than what the measurement devices can physically measure. For this research, the accuracy for T is listed as $\pm 0.1^\circ\text{C}$, and thus, the resolution ($2r$) is 0.2°C (YSI Inc., 2015). For Q all measurements below $99 \text{ m}^3 \text{ s}^{-1}$ have an accuracy of

12327

$\pm 0.1 \text{ m}^3 \text{ s}^{-1}$ and thus, a resolution of $0.2 \text{ m}^3 \text{ s}^{-1}$, while measurements above $99 \text{ m}^3 \text{ s}^{-1}$ have an accuracy of $\pm 1 \text{ m}^3 \text{ s}^{-1}$, and thus a resolution of $2 \text{ m}^3 \text{ s}^{-1}$. This is based on the fact that all data provided by the Water Survey of Canada is accurate to three significant figures. Note that for the case where x_{\min} equals x_{\max} (i.e. no variance in the daily observed data) then $D = 2r$, which means that the only uncertainty considered is due to the measurement.

Once the N_{\max} is determined, the bin-size D was calculated for each N between N_{\min} and N_{\max} . Then, starting at the largest D (i.e. $D = (x_{\max} - x_{\min})/N_{\min}$), the cost-function C_D is calculated at the first partitioning positing, where the first bin is centred at x_{\min} , $[(x_{\min} - D/2)(x_{\min} + D/2)]$, and the N th bin is centred on x_{\max} , $[(x_{\max} - D/2), (x_{\max} + D/2)]$. Then, C_D is calculated at the next partitioning positing, where the first bin is $[(x_{\min} - D/2 + \delta)(x_{\min} + D/2 + \delta)]$, and the N th bin is $[(x_{\max} - D/2 + \delta), (x_{\max} + D/2 + \delta)]$. The value of δ ranged between 0 and D at $(D/100)$ intervals. Thus, for this value of D , 100 values of C_D were calculated since 100 different partitioning positions were used. The mean value of these C_D was used to define the final cost-function value for the given D .

This process is then repeated for the next N between N_{\min} and N_{\max} , using the corresponding D at 100 different partitioning positions, and so on until the smallest D (at N_{\max}). This results in $[N_{\max} - N_{\min}]$ values of mean C_D : the value of D corresponding to the minimum value of C_D is considered to be the optimum bin-size D_{opt} . This D_{opt} is then used to construct the optimum histogram of each daily observation. This histogram can be used to calculate a discretised probability density function ($p(x)$), where for each x (an element of \mathbf{X}), the $p(x)$ is calculated by dividing the number of events in each bin by the total number of elements in \mathbf{X} . The x and $p(x)$ can then be used to calculate the possibility distribution using the transformation described in Eq. (1).

First, the $p(x)$ are ranked from highest to lowest, and the x corresponding to the highest $p(x)$ is has a membership level of 1. Then the $\pi(x)$ values for the remaining x are calculated using Eq. (1). For cases where multiple elements have equal $p(x)$, the highest $\pi(x)$ is assigned to each x . For example, if $p(x_i) = p(x_j)$, and $x_i > x_j$, then

$\pi(x_j) = \pi(x_i)$. This means that in some cases, for each calculated membership level, $\pi(x)$, there exists an α -cut interval $[a, b]_{\mu=\pi(x)}$ where all the elements between a and b have equal $p(x)$ and hence equal $\pi(x)$. By definition of α -cut intervals, all values of x within the interval $[a, b]$ have *at least* a possibility of $\pi(x)$. A special case of this occurs when multiple x have joint-equal maximum $p(x)$, meaning that multiple elements have a membership level of $\mu = 1$. Thus, an α -cut interval is created for the $\mu = 1$ case, creating a trapezoidal membership function, where the modal value of the fuzzy number is defined by an interval rather than a single element.

Once all the $\pi(x)$ are calculated for each element x in X , a *discretised* empirical membership function of the fuzzy number X can be constructed using the calculated α -cut intervals. That is, the fuzzy number is defined by a number of intervals at different membership levels. The upper and lower limit of the intervals at higher membership levels define the extent of the limits of the intervals at lower membership levels. This way the constructed fuzzy numbers maintain convexity (similar to a procedure used by Alvisi and Franchini, 2011), where the widest intervals have the lowest membership level. For example, the interval at $\mu = 0.2$ will contain the interval $\mu = 0.4$, and this interval will contain the interval at $\mu = 0.8$.

In creating this discretised empirical membership function this way (rather than assuming a shape of the function) means that this function best reflects the possibility distribution of the observed data. However, it also means that all fuzzy numbers created using this method are not guaranteed to be defined at the same $\pi(x)$, nor have an equal number of $\pi(x)$ intervals used to define the fuzzy number. Thus, direct fuzzy arithmetic between multiple fuzzy numbers using the extension principle is not possible since it requires each fuzzy number to be defined at the same α -cut intervals (Kaufmann and Gupta, 1985). Thus, linear interpolation is used to define each fuzzy number at a pre-set α -cut interval using the empirical $\pi(x)$ calculated using the transformation. To select the pre-set α -cut intervals it is illustrative to see the impact of selecting two extreme cases: (i) if only two levels are selected (specifically $\mu = 0$ and 1) the constructed fuzzy number will reduce to a triangular fuzzy number. As discussed above

12329

there are important implications of using triangular membership functions that make it undesirable for hydrological data, (ii) if a large number of intervals (e.g. 100 intervals between $\mu = 0$ and 1) are selected, there is a risk that the number of pre-set intervals is much larger than the empirical $\pi(x)$, which means not enough data (empirical α -cut levels intervals) to conduct interpolation, leading to equal interpolated values at multiple α -cut levels. For this research, results (discussed in the following section) of the bin-size optimisation showed that most daily observations for T and Q resulted in 2 to 10 unique $p(x)$ values. Based on this, six pre-set α -cut intervals were selected: 0, 0.2, 0.4, 0.6, 0.8 and 1. The empirical $\pi(x)$ can then be converted to a standardised function at pre-defined membership levels using linear interpolation.

2.3 Fuzzy neural networks

2.3.1 Background on artificial neural networks

Artificial neural networks (ANN) are a type of data-driven model that are defined as a massively parallel distributed information processing system (Elshorbagy et al., 2010; Wen et al., 2013). As a predictive model, ANNs can capture complex, nonlinear relationships that may exist between variables without the explicit understanding of the physical phenomenon (Alvisi and Franchini, 2011; Antanasijević et al., 2014). This has resulted in significant use of ANN models in hydrology when the complexity of the physical systems is high owing partially to an incomplete understanding of the underlying process, and the lack of availability of necessary data (He et al., 2011; Kasiviswanathan et al., 2013). Further, ANNs arguably require less data and do not require an explicit mathematical description of the underlying physical process (Antanasijević et al., 2014), making it a simpler and practical alternative to traditional modelling techniques.

Multilayer Perceptron (MLP) is a type of feedforward ANN and is one of the most commonly used in hydrology (Maier et al., 2010). One of the reasons for the popularity of MLPs is that a trained network can be used as universal approximators with only a single hidden layer (Hornik et al., 1989). This means that models are relatively

12330

simple to develop, and theoretically have the capacity of approximating any linear or nonlinear mapping (ASCE 2000; Elshorbagy et al., 2010; Napolitano et al., 2011; Kasiviswanathan et al., 2013). Further, the popularity of MLP has meant that subsequent research has continued to use MLP (He and Valeo, 2009; Napolitano et al., 2011) and thus, form a reference for the basis of comparing ANN performance (Alvisi and Franchini, 2011).

In the simplest case, an MLP consists of an input layer, a hidden layer, and an output layer as shown in Fig. 2. Each layer consists of a number of neurons (or nodes) that each receive a signal, and on the basis of the strength of the signal, emit an output. Thus, the final output layer is the synthesis and transformation of all the input signals from both the input and the hidden layer (He and Valeo, 2009).

The number of neurons in the input (n_I) and output (n_O) layers corresponds to the number of variables used as the input and the output, respectively and the number of neurons in the hidden layer (n_H) are selected based on the relative complexity of the system (Elshorbagy et al., 2010). A typical MLP is expressed mathematically as follows:

$$\mathbf{y}_i = f_{\text{HID}}(\mathbf{W}_{\text{IH}}\mathbf{x}_i + \mathbf{B}_\text{H}) \quad (8)$$

$$\mathbf{z}_i = f_{\text{OUT}}(\mathbf{W}_{\text{HO}}\mathbf{y}_i + \mathbf{B}_\text{O}) \quad (9)$$

where \mathbf{x}_i is the i th observation (an $n_I \times 1$ vector) from of a total of n observations, \mathbf{W}_{IH} is a $n_H \times n_I$ matrix of weights between the input and hidden-layer, \mathbf{B}_H is a vector ($n_H \times 1$) of biases in the hidden-layer, and \mathbf{y}_i is the i th output (an $n_H \times 1$ vector) of the input signal through the hidden-layer transfer function, f_{HID} . Similarly, \mathbf{W}_{HO} is an $n_O \times n_H$ matrix of weights between the hidden and output-layers, \mathbf{B}_O is an $n_O \times 1$ vector of biases in the output-layer, and f_{OUT} the final transfer function to generate the i th modelled output \mathbf{z}_i (an $n_O \times 1$ vector).

The values of all the weights and biases in the MLP are calculated by training the network by minimising the error – typically mean squared error (E_{MSE}) (He and Valeo, 2009) – between the modelled output and the target data (i.e. observations). A num-

12331

ber of training algorithms can be used, and one of the most common methods is the Levenberg–Marquardt algorithm (LMA) (Alvisi et al., 2006). In this method, the error between the output and target is back-propagated through the model using a gradient method where the weights and biases are adjusted in the direction of maximum error reduction. The LMA is well-suited for ANN problems that have a relatively small number of neurons. To counteract potential over-fitting issues, an early-stopping procedure is used (Alvisi et al., 2006; Maier et al., 2010), which is a form of regularisation where the data is split into three subsets (for training, validation and testing) and the training is terminated when the error on the validation subset increases from the previous iteration. Most ANNs have a deterministic structure without a quantification of the uncertainty corresponding to the predictions (Alvisi and Franchini, 2012; Kasiviswanathan and Sudheer, 2013). This means that users of these models may have excessive confidence in the forecasted values, and misinterpret the applicability of the results (Alvisi and Franchini, 2011). This lack of uncertainty quantification is one reason for the limited appeal of ANN by water resource managers (Abrahart et al., 2012; Maier et al., 2010). Without this characterisation, the results produced by these models have limited value (Kasiviswanathan and Sudheer, 2013).

In this research, two methods are proposed to quantify the uncertainty in MLP modelling to predict DO in the Bow River. First, the uncertainty in the input data (daily mean water temperature and daily mean flow rate) is represented through the use of fuzzy numbers. These fuzzy numbers are created using the probability-possibility transformation discussed in the previous section. Second, the *total uncertainty* (as defined by Alvisi and Franchini, 2011) in the weights and biases of an MLP are quantified using a new possibility theory-based FNN. The total uncertainty represents the overall uncertainty in the modelling process, and not of the individual components (e.g. randomness in observed data). The following section describes the proposed FNN method.

2.3.2 FNN with fuzzy inputs and possibility-based intervals

Alvisi and Franchini (2011) proposed a method to create a FNN, where the weights and biases, and by extension the output, of the neural network are fuzzy numbers rather than crisp (non-fuzzy) numbers. These fuzzy numbers quantify the total uncertainty of the calibrated parameters. While fuzzy set theory based applications of ANN have been limited in hydrology, most have used fuzzy logic, e.g. the widely used Adaptive Neuro-Fuzzy Inference System, where automated IF-THEN rules are used to create crisp outputs (Abrahart et al., 2010; Alvisi and Franchini, 2011). Thus, the advantage of fuzzy outputs is that it provides the uncertainty of the predictions as well, while the fuzzy parameters reflect the uncertainty in the model structure. This uncertainty quantification can be used to by end users to assess the value of the model output.

In their FNN, the MLP model is presented in Eqs. (5) and (6) is modified to predict an interval rather than a single value for the weights, biases and output, corresponding to an α -cut interval (at a defined membership level μ). This is repeated for several α -cut levels, thus building a discretised fuzzy number at a number of membership levels. This is done by using a stepwise, constrained optimisation approach:

$$\left[y_i^L y_i^U \right] = f_{\text{HID}} \left(\left[\mathbf{W}_{\text{IH}}^L \mathbf{W}_{\text{IH}}^U \right] x_i + \left[B_{\text{H}}^L B_{\text{H}}^U \right] \right) \quad (10)$$

$$\left[z_i^L z_i^U \right] = f_{\text{OUT}} \left(\left[\mathbf{W}_{\text{HO}}^L \mathbf{W}_{\text{HO}}^U \right] \times \left[y_i^L y_i^U \right] + \left[B_{\text{O}}^L B_{\text{O}}^U \right] \right) \quad (11)$$

where all the variables are as described as before, and the superscripts "U" and "L" represent the upper and lower limits of the α -cut interval, respectively. The constraints are defined so that the upper and lower limits of each weight and bias (in both layers) minimise the width of the predicted interval:

12333

$$\begin{aligned} & \min \left(\sum_{i=1}^n (z_i^L - z_i^U) \right) \\ & \frac{1}{n} \sum_{i=1}^n (\delta_i) > P_{\text{CI}} \quad (12) \\ & \delta_i = \begin{cases} 1, & \text{if } z_i^L < t_i < z_i^U \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where t is that target (observed data) and P_{CI} is a predefined percentage of data. Alvisi and Franchini (2011) defined P to be 100 % at $\mu = 0$, 99 % at $\mu = 0.25$, 95 % at $\mu = 0.5$ and 90 % at $\mu = 0.75$. This algorithm was built starting at $\mu = 0$ and moving to higher membership levels to maintain convex membership functions of the generated fuzzy numbers by using the results of the previous optimisation as the upper and lower limit constraints for the proceeding optimisation. Lastly, at $\mu = 1$, the interval collapses to a singleton, represent the crisp results from non-fuzzy ANN. Therefore, these α -cut intervals of the FNN output quantify the uncertainty around the crisp prediction, within which is expected to contain P_{CI} percentage of data.

In this research, this method is modified in two ways. First, the inputs x are also fuzzy numbers, which means that Eqs. (10) and (11) are revised as follows:

$$\left[y_i^L y_i^U \right] = f_{\text{HID}} \left(\left[\mathbf{W}_{\text{IH}}^L \mathbf{W}_{\text{IH}}^U \right] \times \left[x_i^L x_i^U \right] + \left[B_{\text{H}}^L B_{\text{H}}^U \right] \right) \quad (13)$$

$$\left[z_i^L z_i^U \right] = f_{\text{OUT}} \left(\left[\mathbf{W}_{\text{HO}}^L \mathbf{W}_{\text{HO}}^U \right] \times \left[y_i^L y_i^U \right] + \left[B_{\text{O}}^L B_{\text{O}}^U \right] \right). \quad (14)$$

Note that now the input vector is represented by its upper and lower limits. The major impact on this is that the training algorithm for the FNN needs to accommodate this fuzzy α -cut interval, which requires the implementation of fuzzy arithmetic principles (Kaufmann and Gupta, 1985). The cost function for the optimisation remains unchanged.

12334

The second modification of the original algorithm is related to the selection of the percent of data included in the predicted interval (P_{CI}). In the original, the selection is arbitrary and end-users of this method may be interested in the events that are not included in the selected P_{CI} . Thus, a full spectrum of possible values for a given prediction is required. Thus, the Alvisi and Franchini (2011) approach is further refined by utilising the same relationship between probability and possibility that was used to define the input fuzzy numbers, giving a more objective means of designing FNNs with fuzzy weights, biases and output.

In the adopted possibility-probability framework, the interval $[a b]_{\alpha}$ created by the α -cut at a $\mu = \alpha$ implies that:

$$[p(x < a) + p(x > b)] = \alpha. \quad (15)$$

This can be used to calculate the probability:

$$[p(a < x < b)] = (1 - \alpha). \quad (16)$$

This means that there is a probability of $(1 - \alpha)$ that the random variable x falls within the interval $[a b]_{\alpha}$. In other words the α -cuts of a possibility distribution (at any μ) correspond to the $(1 - \alpha)$ confidence interval of the probability distribution of the same variable (Serrurier and Prade, 2013). This principle is used to select the different P for the optimisation constraints rather than the predetermined P_{CI} selected by Alvisi and Franchini (2011) These are shown in Table 2.

Note that for practical purposes, P_{CI} was selected as 99.50 % at $\mu = 0$ to prevent over-fitting. The implication of this selection is that at $\mu = 0$, *nearly-all* the observed data should fall within this predicted FNN interval, reflecting the highest uncertainty in the prediction. The uncertainty decreases as μ increases. For the $\mu = 1$ case the values of the weights and biases were determined to be the mid-point of the interval at $\mu = 0.8$ to maintain convexity of the produced fuzzy numbers, and the difficulty in finding an interval containing 0 % of the data.

12335

2.3.3 Network architecture and implementation

For this research a three layer, feedforward MLP architecture was selected to model minimum daily DO (the output) using fuzzified daily flowrate (Q) and fuzzified daily water temperature (T) as the inputs. The three layers consist of an input layer, an output layer, and a hidden layer (with 5 neurons based on a trial-and-error search procedure). This architecture was selected for three reasons: it is one of the most commonly used in hydrology (Maier et al., 2010), it can be used as a universal approximator (Hornik et al., 1989), and as reference for comparing performance with previous research (He and Valeo, 2009; Napolitano et al., 2011). In particular, a previous study modelling minimum DO in the Bow River used a three-layer MLP feedforward network (see He et al., 2011). Two transfer functions are required for FNN implementation: the hyperbolic tangent sigmoid function was selected for f_{HID} , and a pure linear function for f_{OUT} . Both function selections follow Alvisi and Franchini (2011), Wen et al. (2012) and Elshorbagy et al. (2010), and are described as follows:

$$f_{HID} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (17)$$

$$f_{OUT} = x. \quad (18)$$

The LMA method was used to train the network, minimising E_{MSE} . The input and output data was pre-processed before training, validating and testing: the data was normalised so that all input and output data fell within the interval $[-1 1]$. Further the data were randomly divided into training, validation and testing subsets, following a 50–25–25 % split.

This FNN optimisation algorithm was implemented in MATLAB (version 2015a). First, the built-in MATLAB Neural Network Toolbox was used to estimate the value of weights and biases using the midpoint of the interval at $\mu = 1$. The results from this were used as the constraints to solve the FNN optimisation (Eqs. 12 to 14) at subsequent lower membership levels. The Shuffled Complex Evolution algorithm (commonly known as

12336

SCE-UA, Duan et al., 1992) was used to find the optimisation solution. The optimisation is run such that the intervals at higher membership levels govern the upper and lower bounds of the predicted interval in order to preserve the convexity of fuzzy numbers.

2.4 Risk analysis using defuzzification

5 Risk analyses for complex systems is challenging for a number of reasons, including an insufficient understanding of the failure mechanisms (Deng et al., 2011). The use of imprecise information (e.g. fuzzy numbers) is an effective method of conducting a risk analysis (Deng et al., 2011). However, communicating uncertainty is an important, yet difficult task, and many different frameworks exist to do so; water quality indices (Sadiq et al., 2007; Van Steenberg et al., 2012) are one example. Since water resource managers often prefer to use probabilistic measures (rather than possibilistic ones), it is important to convert the possibility of low DO to a comparable probability for effective communication of risk analysis. Note that the linguistic parameters (e.g. “most likely”) that are often used to convey risk or uncertainty (Van Steenberg et al., 2012) have
10 a probability-based meaning – in this case “most likely” is a measure of *likelihood*.
15

In this research, a *defuzzification* procedure is used to convert the possibility of low DO to a probability measure, to represent the risk of observing low DO (below a given threshold) in the Bow River. This method uses the inverse of the transformation described in Eq. (1); however, instead of calculating the probability of one element, $p(x)$, which is of limited value in most applications, it is generalised to calculated $P \{ \{X < x\} \}$, as follows (from Khan and Valeo, 2014a, 2015b): for any x in the support (defined as the α -cut interval at $\mu = 0$) of a fuzzy number $[a b]$ we have the corresponding μ and the paired value x' which shares the same membership level. The value μ is the sum of the cumulative probability between $[a, x]$ and $[x', b]$, labelled P_L and P_R , respectively:
20

$$\mu(x) = P_L + P_R \tag{19}$$

where P_L represents the cumulative probability between a and x which is assumed to equal the probability $P \{ \{X < x\} \}$, since the fuzzy number defines any values to less

than a to be impossible (i.e. $\mu = 0$). Given the fact that the fuzzy number is not symmetrical, the lengths of the two intervals $[a, x]$ and $[x', b]$ can be used to establish a relationship between P_L and P_R . Then, P_L can be estimated as:

$$P\{X < x\} = P_L = \frac{\mu}{1 + \frac{(b-x')}{(x-a)}} \tag{20}$$

5 Thus, Eq. (20) gives the probability that the predicted minimum DO for a given day is below the threshold value x . For example, if the lowest acceptable DO concentration for the protection of aquatic life for cold water ecosystems (6.5 mgL^{-1} , Canadian Council of Ministers of the Environment, 1999) is selected, then this transformation can be used to calculate the probability that the predicted fuzzy DO will be below 6.5 mgL^{-1} .

3 Results and discussion

3.1 Probability-possibility transformation using bin-size optimisation

The bin-size optimisation and the probability-possibility transformation algorithms were applied to the collected Q and T data for the nine year period. The constructed fuzzy numbers were then used to calibrate the FNN model. This section compares the results of constructing a discretised probability distribution with and without the bin-size optimisation algorithm and its impact on the resulting membership function of the fuzzy number. The comparison is illustrated through five examples each for Q and T as a means of illustrating the advantages of using the proposed approach.
15

Figure 3 shows sample results of converting hourly Q observations to fuzzy numbers for five cases. The left most column in the figure shows the raw data, i.e. the observations sampled over the course of 24 h. The resulting histogram-based probability functions are shown for both the optimised (D_{opt} , illustrated with circles) and original ($D_{orig} = 2r$; see Sect. “A new algorithm to create fuzzy numbers” for the definition, illustrated with squares) bin-sizes in the second column. The third and fourth columns in
20

Fig. 3 show the resulting discretised empirical membership function using each of the histograms. The five examples selected here represent a full spectrum of results for the bin-size optimisation. The first row shows an example of when the optimum result was equal to the measurement resolution ($D_{\text{orig}} = D_{\text{opt}} = 2$), followed by cases where the D_{opt} was 4, 4.5, 10 and 20 times greater than the initial bin-size.

The example in the first row illustrates cases where the bin-size optimisation algorithm calculates an optimum bin-size, corresponding to the minimum cost-function C_D , which is equal to the instrument measurement resolution. Thus, the resulting probability distributions for both cases are equivalent, as are the membership functions. In most cases, this occurred when the calculated minimum C_D would result in a D_{opt} smaller than $D_{\text{orig}} = 2r$, and since this is not physically feasible (measurable) the algorithm did not consider any bin-sizes below $2r$. Of note in this example is that the transformation of the probability distribution results in five empirical membership levels. Only one element was found to have a membership level equal to 1 (at $Q = 161 \text{ m}^3 \text{ s}^{-1}$). Thus, the α -level cut at this level is a simple singleton: $[161]_{\mu=1}$. The next membership level was calculated as 0.58; again the resulting α -cut level only has one element at $Q = 149$ (which is less than the modal value). However, at this level the upper and lower limits of α -cuts at higher membership levels define the upper and lower limits of α -cuts at lower levels. Thus, using the information from the α -level cut at $\mu = 1$, the level at $\mu = 0.58$ was defined as $[149 161]_{\mu=0.58}$. The next membership level calculated was 0.46, and four elements had equal membership levels, ranging between 147 and 165. The α -cut interval at this level was defined as: $[147 165]_{\mu=0.46}$. Note that in this case, this interval captures both the intervals at higher membership levels within its limits, i.e. the lower limit is less than the lower limits of higher intervals, and the upper limit is greater than then upper limits at higher levels. The next membership level was calculated to be 0.125, and three elements between 157 and 171 were assigned this value. However, the lower limit at $\mu = 0.46$ (the next higher membership level) was 147, which is less than 157, and thus, for the α -cut level at this membership, the interval is then revised to: $[147 171]_{\mu=0.125}$ rather than $[157 171]_{\mu=0.125}$ to maintain convexity. Again, the reason

12339

here is that if something is possible at $\mu = 0.46$, it must be possible (by definition) at $\mu = 0.125$. The last membership level found for this particular example was $\mu = 0$, with six elements sharing this value, ranging from 145 to 173, resulting in an α -cut level of $[145 173]_{\mu=0}$. Together, these five membership levels define a discretised membership function of the fuzzy number for Q on 27 July 2008. Following this, linear interpolation was conducted to find the elements corresponding to the six predefined membership levels of $\mu = 0, 0.2, 0.4, 0.6, 0.8$ and 1. The results are not explicitly shown in the figure for clarity, but can be essentially located on the dashed line in the last column on the corresponding membership levels.

The second row in Fig. 3 shows the results for 20 August 2009, where the optimum bin-size was found to be four times higher than the original bin-size ($D_{\text{opt}} = 0.8$ vs. $D_{\text{orig}} = 0.2$). The impact of this change is clearly evident in the distribution functions in the second column. The original histogram is multi-modal, and with multiple candidates as the modal value (where $\mu = 1$), whereas the post-optimisation histogram is considerably smoother, with a definitive modal value at $Q = 91.4 \text{ m}^3 \text{ s}^{-1}$. The impact of this increase in bin-size is that the resulting membership function is defined at four membership levels (0, 0.25, 0.54 and 1), whereas the original function was defined at six levels, including an interval (rather than singleton) at $\mu = 1$. This decrease in membership levels in this case has a consequence of smoothing out the membership function, as can be seen by comparing the shapes of the functions in columns three and four. The overall impact of this smoothing out of both the distribution and the membership functions is that the heightened specificity of the original function at $\mu = 0.54$ and above is reduced to a more generalised shape.

Since the objective of the bin-size algorithm was to reduce the error between the histogram created using the D_{opt} and the unknown theoretical distribution, then the density function plotted in Fig. 3 represents the closest distribution to the unknown distribution. Hence, the membership function generated using this optimum distribution better reflects the underlying phenomenon than the membership function generated using D_{orig} . Thus, in comparing columns 3 and 4 for the second row, the smoother membership

12340

function representing D_{opt} is preferred. Linear interpolation is then performed on this membership function to get values of Q at the six predefined membership levels.

Similar results can be seen in the third row in Fig. 3, where the optimised bin-size is 4.5 times greater than the original bin-size, ($D_{\text{opt}} = 9$ vs. $D_{\text{orig}} = 2$). Again, the original histogram is extremely uneven, whereas the post-optimisation histogram is considerably smoother with a definitive modal value at $Q = 277 \text{ m}^3 \text{ s}^{-1}$. The overall impact of this smoothing is that the specificity of the function at $\mu = 0.6$ and higher of the original function is reduced to a more general shape in the optimised function.

The fourth row shows a different phenomenon, where instead of smoothing out the original membership function, the combined bin-size optimisation and transformation algorithm, creates a membership function with more specificity. In this case D_{opt} is ten times higher than D_{orig} , and the consequence of this increase is the smoother probability density function with one clear modal value (at $Q = 70 \text{ m}^3 \text{ s}^{-1}$). In contrast, the original histogram had six elements with joint-equal $p(x)$, resulting in a membership function that is shaped similarly to a uniform distribution (column 3) and defined with only 3 membership levels. This means that all values are considered equally-possible and represents maximum vagueness. However, using the optimised value, this is no longer the case and the modal value is assigned a membership level of 1, and the remaining elements defined at three other membership levels. This suggests that this modal value is more possible (since it has a higher possibility), and this is reflected in the observations. This example illustrates that the method can not only generalise the data to smoother functions (as shown in the first three examples) but can also be more specific when the underlying data demonstrates this but this is not captured by the non-optimised bin-width distribution function.

The last example for Q in Fig. 3 is an example of a case where the number of membership levels for both the original and optimised membership function are equal (four in this case), however the bin-size is 20 times greater for the optimised case. In this case, an optimum bin-size was found that did not change the specificity of the membership function, i.e. it is still defined with the same number of intervals but at different mem-

12341

bership levels. In this case, the probability for D_{orig} is extremely uneven, but smoothed out to a unimodal function with the D_{opt} . The final membership function for D_{opt} is defined more generally (smoothly) especially at higher membership levels compared to the one defined by D_{orig} . This example again demonstrates the utility of the new coupled optimisation-transformation method to create fuzzy numbers for data where the underlying distribution is unknown.

Figure 4 shows similar results for the five water temperature examples, where the D_{opt} was equal to the D_{orig} (the first example on the top row), or increased by a factor of 1.5, 2.5, 3 or 5. The first example shows a case with very little T variation over a given day and the water temperature falls between 5.2 and 6.2°C for the entire day. This lack of variability is responsible for the minimal bin-size selection D_{opt} : a unimodal distribution is best constructed using smaller bin-sizes for these cases. The second example shows another case where D_{opt} is only slightly greater than the original, resulting in a somewhat smoother probability function, and a slightly smoother membership function.

A major difference between the T and Q data is that the former is strongly diurnal, increasing after sunrise in the morning, peaking in late afternoon, and then decreasing through the night. This temporal trend is seen for all examples in Fig. 4, but most significantly in the bottom three examples. A major implication for this in developing a probability density function for this data is that the resulting shape will have a tendency to be bimodal. This means that the resulting membership functions might be trapezoidal or near-trapezoidal (and hence most vague) in shape, which is clearly demonstrated in the functions created using D_{orig} in the bottom three examples. However, in each case the optimised bin-size creates a smoother probability distribution with a clearer modal value, resulting in membership functions that are no longer trapezoidal.

Thus, without using the bin-size optimisation algorithm there is a risk that the resulting membership functions will be too vague and do not represent the information that can be gained from the observations. It is worth nothing that for these three examples, if linear interpolation is used on the original membership function, the resulting

12342

A sample of the fuzzy weights and biases produced through the optimisation are shown in Fig. 5. Note that the membership functions are assumed to be piecewise linear (following similar assumptions made in Alvisi and Franchini, 2011; Khan et al., 2013; Khan and Valeo, 2015a), i.e. that the intervals at each membership levels can be joined to create a fuzzy number. This can be confirmed by the fact that each of the weights and biases are *convex* where intervals at lower levels are wider than intervals at higher levels, and are *normal* with at least one element with $\mu = 1$. Note that each *weight* and *bias* has a non-linear membership function, i.e. none of the functions produced follow the typical triangular functions and are not necessarily symmetric about the modal value.

The figure demonstrates that enough α -cut levels (i.e. six levels equally spaced between 0 and 1) have been selected to *completely* define the shape of the membership functions. A smaller number of levels e.g. two levels, one at $\mu = 0$ and one at $\mu = 1$, the fuzzy number collapses to a triangular fuzzy number, which is not desirable for this research, as discussed in previous sections. When only two levels are selected, the figures demonstrate that significant differences exist between those simple functions and the ones generated using six membership levels: the decrease in the width of the intervals with an increase in membership level is not linear as is in triangular shaped function. Similarly, a higher number of intervals e.g. 100 intervals, equally spaced between 0 and 1, could be selected. The risk in selecting many intervals is that as the membership level increases (closer to 1) the intervals become narrower as a consequence of convexity. This will result in numerous closely spaced intervals, with essentially equal upper and lower bounds, making the extra information redundant. This is demonstrated in the sample membership functions in Fig. 5 for \mathbf{W}_{IH} number 5 and \mathbf{B}_O , where the intervals at the higher membership levels collapse to a singleton, or are extremely narrow. Thus, defining more uncertainty bands between the existing levels would not add more detail but would merely replicate the information already calculated.

12345

Connecting this back to the results in Table 4, these two particular weights and biases show why the percentage of data calculated at $\mu = 0.2$ (for training) cannot be improved by further optimisation. At some point, if the intervals at $\mu = 0.2$ for the various weights and biases collapse to a single element, no further refinement in the model is possible (since all the constraints are met) and the minimum interval width of the predicted DO whilst capturing *at least* P_{CI} amount of data has been reached. It is worth emphasizing here that the uncertainty represented by these fuzzy number weights and biases is not the uncertainty of the particular weight or bias, but is the total forecasting uncertainty defined by the quantifying bands around the crisp predicted value.

Tables 3 and 4, and Fig. 5 demonstrates the overall success of the proposed approach to calibrate an FNN model. The optimisation algorithm is defined based on the principles of possibility theory (i.e. defining the amount of data to include in each interval) and is a transparent, repeatable and objective (not arbitrary) method to create the fuzzy numbers for the FNN model.

The observed vs. crisp predictions (black dots) and fuzzy predictions at $\mu = 0$ (grey lines) for daily minimum DO for the three different data subsets (training, validation and testing) are shown in Fig. 6. The figure shows that *nearly-all* (specifically, 99.4, 98.8 and 99.0% of the training, validation, and testing subsets, respectively) of the observations fall within the $\mu = 0$ interval, since the observed values (black dots) tend to fall inside the grey lines. This figure also highlights one of the major advantages of the FNN over a simple non-fuzzy ANN: almost all of the fuzzy results intersect the 1 : 1 line whereas many of the crisp results are quite far from that line, especially at low DO values (which is marked at 6.5 mgL^{-1} on the figure). In other words, while the fuzzy number prediction may not predict the observed value exactly, they provide at least some *possibility* of the observed value within its various α -cut intervals, but the crisp results do not provide this additional information. This figure illustrates that E_{NSE} (listed in Table 2 for the $\mu = 1$ case only) is not representative of the entire fuzzy number predictions, since it does not capture the performance at different membership levels.

12346

Thus, there is a need to develop an equivalent performance metric when comparing crisp observations to fuzzy number predictions.

Figure 6 also demonstrates the benefit of the FNN approach as compared to the crisp approach with respect to predicting low DO (i.e. when DO is less than 6.5 mg L^{-1}). The FNN predicts more of the low DO events within its intervals as compared to the crisp method. The figure demonstrates that both the crisp ($\mu = 1$) and fuzzy predictions tend to over predict the low DO events (since they fall above the 1 : 1 line), but the fuzzy intervals are closer to the observations (i.e. they intersect the 1 : 1 line for the majority of low DO events), and therefore predict some *possibility* (even if it is a low probability) the low DO events occur. Thus, generally speaking the ability of the FNN to capture *nearly-all* of the data within its predicted intervals guarantees that most of the low DO events are successfully predicted. This is a major improvement over conventional methods used to predict low DO.

Trend plots of observed minimum DO and predicted fuzzy minimum DO for the years 2004, 2006, 2007 and 2010 are illustrated in Figs. 7 and 8. These years were selected due to the high number of low DO occurrences in each year (as listed in Table 1), and highlight the utility of the proposed method to predict minimum DO using abiotic factors in the absence of a complete understanding of the physical mechanisms that govern DO in the Bow River. Note that for each year, 50 % of the data are training data, 25 % are validation and 25 % are testing data. However, for clarity this difference is not individually highlighted for each data point in these figures.

In Figs. 7 and 8, the predicted minimum DO at equivalent membership levels (e.g. 0^L or 0^R) at different times steps are joined together creating bands representative of the predicted fuzzy numbers calculated at each time step. In doing so, it is apparent that all the observed values fall within the $\mu = 0$ interval for the years 2006, 2007 and 2010, and all but one observation in 2004. The width of each band represents the amount of uncertainty associated with each membership level. For example, the bands are the widest at $\mu = 0$, meaning the results have the most vagueness associated with it. Narrower bands are seen as the membership level increases until $\mu = 1$. This reflects

12347

a decrease in vagueness, increase in credibility, or less uncertainty of the predicted value, as the membership level increases. Note that the majority of the predictions at $\mu = 1$ are single elements but some predictions are α -cut intervals (e.g. $[a b]_{\mu=1}$). This means that when not enough information is available, the fuzzy prediction collapse to trapezoidal membership functions.

In each of the years shown, the majority of the observations tend to fall within the $\mu = 0.2$ interval or higher, with only the low DO (i.e. $< 5 \text{ mg L}^{-1}$) falling within the $\mu = 0$ and $\mu = 0.2$ bands. This suggests that the low DO events are predicted with less certainty compared to the occasions when DO concentration is high. Also note that the interval at $\mu = 0$ is highly skewed towards the lower limit ($\mu = 0^L$), i.e. the modal value is not at the centre of the interval. This shows that the FNN has been trained to capture these low DO events, but predicts them with lower credibility. Compared to the crisp results (i.e. those at $\mu = 1$), for these low DO events, the proposed method provides some possibility of low DO, whereas the crisp results do not predict a possibility of low DO. Thus, the ability to capture the full array of DO observations within different intervals is an advantaged of the proposed method over existing methods.

The trend plot for 2004 shows that observed DO decreases rapidly from late June to late July, followed by a few days of missing data and near-zero observations, before increasing to higher concentrations. Details of this trend are shown in Fig. 9 which shows magnified versions of important periods for each year. The reason for this rapid decrease in 2004 is unclear and may be related to problems with the real-time monitoring device which was in its first year of operation that year. However, it demonstrates that the efficacy of data-driven methods is dependent on the quality of the data. Since the proposed method was calibrated to capture *nearly-all* the observations (including outliers like those seen in 2004) within the least certain band at $\mu = 0$, the resulting network predicts results to include these outliers, but at low credibility levels. As the data length increases (i.e. the addition of more data and the FNN is subsequently updated), the number of these types of outliers included within the $\mu = 0$ band will decrease because the optimisation algorithm (Eqs. 12 to 14) searches for the smallest width of the

12348

interval whilst including 99.5 % of the data. Thus, with more data, it is expected that these extreme events (i.e. the outliers seen in 2004) will no longer be captured within the $\mu = 0$ band.

The time series plot for 2006 shows that all the observations fall within the predicted intervals, and that the predicted trend generally follows the observed trend. The majority of the 25 low DO events ($< 5 \text{ mgL}^{-1}$) occur from mid-July and continue occasionally until mid-September. Details of some of these low DO events are plotted in Fig. 9. Figure 7 demonstrates these low DO events are captured between $\mu = 0$ and 0.2 intervals, similar to the 2004 case, meaning that the credibility of these predictions is the lowest. However, unlike the 2004 case, Fig. 9 demonstrates that in 2006 the predicted intervals tend to follow the same trend as the observations for these low DO events, even if it is predicting them at a low credibility.

In contrast to the results from 2004 and 2006, the majority of observations are captured at higher membership levels (i.e. greater than $\mu = 0.2$) in 2007 as shown in Fig. 8. That is, only a limited number of observations are captured within the lowest credibility band. More importantly, 26 out of the 27 low DO ($< 6.5 \text{ mgL}^{-1}$) events are captured at a membership level greater than 0.2. Meaning that the low DO predictions in 2007 for the 6.5 mgL^{-1} guideline are predicted with higher credibility than the 2004 and 2006 cases. Another difference for the results from this year is that many of the low DO observations for 2007 are more evenly scattered around the $\mu = 1$ predictions (as seen in Fig. 9) in contrast to the 2004 and 2006 cases, where the low DO events were always predicted to be closer to lower bounds of the intervals.

The trend plot for 2010 is shown in Fig. 8 and it is clear that all observations fall within the $\mu = 0.2$ interval or at higher α -level intervals, meaning that the predictions capture the observations with higher certainty. This is likely due to the lack of DO events below 5 or 6.5 mgL^{-1} in 2010. Also of note for this year is that *all* observations are less than 10 mgL^{-1} , and about 90 % of all observations are below the 9.5 mgL^{-1} guideline (as listed in Table 1). The trend plot again illustrates that the FNN generally reproduces the overall trend of observed minimum DO. This can be seen in a period in early May

12349

where DO falls from a high of 10 mgL^{-1} to a low of 7 mgL^{-1} , and all the predicted intervals replicate the trend. This is an indication that the two abiotic input parameters are suitable parameters for predicting minimum DO in this urbanised watershed.

Figure 9 shows details of a low DO ($< 9.5 \text{ mgL}^{-1}$) event in 2010 in late July through late August. The bulk of low DO events are captured between the $\mu = 0.6^R$ and 0.2^R intervals – demonstrating that these values are predicted with higher credibility than the other low DO cases in 2004 and 2006, and are predicted closer to the upper end of the interval. All of the low DO ($< 9.5 \text{ mgL}^{-1}$) observations in this plot are under predicted by the crisp method (though not with the FNN method since they are captured within a fuzzy interval). This shows that the crisp ANN results tend to over predict extremely low DO events (i.e. $< 5 \text{ mgL}^{-1}$) while under predicting the $\text{DO} < 9.5 \text{ mgL}^{-1}$ events.

The analysis of the trend plots for these four sample years show that the proposed FNN method is extremely versatile in capturing the observed daily minimum DO in the Bow River using Q and T as inputs. The crisp case (at $\mu = 1$) cannot capture the low DO events (as shown in Figs. 7 and 8), however the FNN is able to capture these low DO events. Generally speaking the training method selected for the FNN has been successful in creating nested-intervals to represent the predicted fuzzy numbers. The widths of the predicted intervals correspond to the certainty of the predictions (i.e. larger intervals for more uncertainty). The utility of this method is further demonstrated in the proceeding section, where the risk of low DO is estimated using a possibility-probability (i.e. defuzzification) technique.

3.3 Risk analysis for low DO events

The utility of the FNN method is illustrated through an analysis of the ability of the model to predict low DO events, and then a possibility-probability transformation is used to assess the risk of these low DO events. The number of occasions when observed DO was below any of the three guidelines used for this research are summarised in Table 1. The FNN model was able to capture 100 % of all low DO events (i.e. below 5, or 6.5

12350

or 9.5 mgL^{-1}) within the predicted intervals. In comparison, the crisp network (i.e. at $\mu = 1$) did not predict DO to be less than 5 mgL^{-1} on any of the 51 occasions. Similarly, it predicted DO to be less than the more conservative limit of 6.5 mgL^{-1} in only 53 % of the 184 occurrences. For the last case, the 9.5 mgL^{-1} limit, the ANN method still trailed the FNN method, by predicting 96 % of these low DO events. This illustrates that not only can the FNN method capture more low DO events within its predicted intervals, it performs exceptionally better for the highest risk case ($\text{DO} < 5 \text{ mgL}^{-1}$). In general, more days were correctly identified when there was a risk of low DO using FNN rather than the typical ANN approach. This is one of the major advantages of using a fuzzy number based uncertainty analysis component to low DO prediction.

Once all the low DO events were identified, the inverse transformation (defuzzification) described in Sect. 2.4 was used to estimate the probability of low DO. The primary reason for converting from possibility to probability is to improve the communication of the risk of low DO. For each low DO event (i.e. at 5, or 6 or 9.5 mgL^{-1}), the predicted membership function was used to determine the possibility of low DO, i.e. identify the membership level where the membership function intersects either of the low DO guidelines (some examples of low DO events are shown in Fig. 10). Once these were identified, the defuzzification technique was used to predict the probability of low DO (e.g. $P(\{\text{DO} < 5 \text{ mgL}^{-1}\})$).

For the first case, $P(\{\text{DO} < 5 \text{ mgL}^{-1}\})$, the probability ranged between 11.5 and 16.6 % for the 51 events, with a median value of 14 %. This means that on days when DO was observed to be below 5 mgL^{-1} , the FNN results identified the possibility of low DO and the probability of DO to be below the 5 mgL^{-1} guideline was ~ 14 %. Thus, the FNN method predicts a probability of low DO (even if it is relatively small) on days when the crisp ANN does not predict a low DO event. This value can be used as a threshold by water resource managers for estimating the risk of low DO. For example, if forecasted water temperature and flow rate are used to predict minimum fuzzy DO using the calibrated model, if the risk of low DO reaches 14 %, the event can be flagged.

12351

Appropriate defence mechanisms can then be implemented to prevent the occurrence of low DO.

For the 184 cases where DO was observed to be less than 6.5 mgL^{-1} , the probability-possibility transformation estimated the risk of low DO to be between 13.7 and 92.9 %, with a median value of 73.4 %. Compared to the first case, the probability of low DO for this threshold is higher and more variable. The low probabilities are associated with predictions of low DO at lower credibility levels at the lower limit of the intervals (i.e. L), whereas the higher probabilities are associated with predictions corresponding to the upper limits of the intervals (R). For 43 out of the 184 low DO events, the probability of low DO was less than 21 % – these events correspond to predictions of low DO at low credibility levels at the lower limits. For the majority of events (107 out of 184), the risk was high, more than 65 %. It is worth noting that the crisp network only predicted 53 % of these low DO events, and of those correctly identified, the majority were over-predicted.

For the last, most conservative case, the probability of predicting DO to be less than 9.5 mgL^{-1} (1179 events) varied between 21.9 and 100 %, with a median value of 98.1 %. Only 46 out of the 1179 events had a probability of less than 70 %; the majority of events had a high risk of low DO: more than 80 % of the events had a risk of low DO of more than 90 %. This shows that the FNN can predict with high probability, the events where minimum daily DO is observed to be below the 9.5 mgL^{-1} limit.

The predicted membership functions of minimum DO for nine examples are shown in Fig. 10, along with the observed minimum DO (the vertical dashed line). Three samples are shown for each low DO guideline: 5, 6.5, or 9.5 mgL^{-1} ; along with the associated risk of low DO calculated using the defuzzification technique. Note that the membership functions of the predicted fuzzy numbers show that each is uniquely shaped, *convex* and *normal*, highlighting the fact that the proposed optimisation algorithm successfully produces nested intervals at each membership levels (as it does for the weights and biases shown in Fig. 5). For the predicted fuzzy DO, the intervals are largest at $\mu = 0$, which decrease in size as the membership level increases. The shape of the mem-

12352

selecting confidence intervals), and (ii) using fuzzy number inputs rather than crisp inputs. This research also proposes a new two-step method to construct these fuzzy number inputs using observations. First a bin-size optimising algorithm is used to find the optimum histogram (as an estimate of the underlying but unknown probably density function of the observations). Then a probability-possibility transformation is used to determine the shape of the fuzzy number membership function.

The results demonstrate the network training algorithm proposed can be successfully implemented. Model results demonstrate that low DO events are better captured by the fuzzy network as compared to a non-fuzzy network. A defuzzification technique is then used to calculate the risk of low DO events. Generally speaking, the method demonstrates that a data-driven approach using abiotic inputs is a feasible method for predicting minimum daily DO. Results from this research can be implemented by water resource managers to assess conditions that lead to, and quantify the risk of low DO.

Data availability. The data used in this research may be obtained from the City of Calgary and Environment Canada.

Author contributions. U. T. Khan conducted all aspects of this research and wrote the manuscript. C. Valeo assisted in manuscript preparation and revisions.

Acknowledgements. The authors would like to thank S. Alvisi from the Università degli Studi di Ferrara for providing the MATLAB code for the original Fuzzy Neural Network model. The authors would like to thank: the Natural Sciences and Engineering Research Council of Canada; the Ministry of Advanced Education, Innovation and Technology – Government of British Columbia; and the University of Victoria, for funding this research. Lastly, the authors are grateful for the City of Calgary and Environment Canada for providing the data used in this research.

12355

References

- Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E., and Wilby, R. L.: Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting, *Prog. Phys. Geogr.*, 36, 480–513, doi:10.1177/0309133312444943, 2012.
- Adams, K. A., Barth, J. A., and Chan, F.: Temporal variability of near-bottom dissolved oxygen during upwelling off central Oregon, *J. Geophys. Res.-Oceans*, 118, 4839–4854, doi:10.1002/jgrc.20361, 2013.
- AENV – Alberta Environment: Alberta water quality guideline for the protection of freshwater aquatic life: Dissolved oxygen, Catalogue #: ENV-0.94-OP, Standards and Guidelines Branch, Alberta Environment, Edmonton, Alberta, Canada, 1997.
- Alvisi, S. and Franchini, M.: Fuzzy neural networks for water level and discharge forecasting with uncertainty, *Environ. Model. Softw.*, 26, 523–537, doi:10.1016/j.envsoft.2010.10.016, 2011.
- Alvisi, S. and Franchini, M.: Grey neural networks for river stage forecasting with uncertainty, *Phys. Chem. Earth*, 42, 108–118, doi:10.1016/j.pce.2011.04.002, 2012.
- Alvisi, S., Mascellani, G., Franchini, M., and Bárdossy, A.: Water level forecasting through fuzzy logic and artificial neural network approaches, *Hydrol. Earth Syst. Sci.*, 10, 1–17, doi:10.5194/hess-10-1-2006, 2006.
- Antanasijević, D., Pocajt, V., Perić-Grujić, A., and Ristić, M.: Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis, *J. Hydrol.*, 519, 1895–1907, doi:10.1016/j.jhydrol.2014.10.009, 2014.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology: Artificial neural networks in hydrology, II: Hydrologic applications, *J. Hydrol. Eng.*, 5, 124–137, doi:10.1061/(ASCE)1084-0699(2000)5:2(124), 2000.
- Ay, M. and Kisi, O.: Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado, *J. Environ. Eng.-ASCE*, 138, 654–662, doi:10.1061/(ASCE)EE.1943-7870.0000511, 2011.
- Bárdossy, A., Bogardi, I., and Duckstein, L.: Fuzzy regression in hydrology, *Water Resour. Res.*, 26, 1497–1508, doi:10.1029/WR026i007p01497, 1990.
- Bárdossy, A., Mascellani, G., and Franchini, M.: Fuzzy unit hydrograph, *Water Resour. Res.*, 42, W02401, doi:10.1029/2004WR003751, 2006.

12356

- Betrie, G. D., Sadiq, R., Morin, K. A., and Tesfamariam, S.: Uncertainty quantification and integration of machine learning techniques for predicting acid rock drainage chemistry: a probability bounds approach, *Sci. Total Environ.*, 490, 182–190, doi:10.1016/j.scitotenv.2014.04.125, 2014.
- 5 Bow River Basin Council: Profile of the Bow River Basin, available at: http://wsow.brbc.ab.ca/index.php?option=com_content&view=article&id=81&Itemid=83, last access: 23 November 2015.
- CCME – Canadian Council of Ministers of the Environment: Canadian Water Quality Guidelines for the Protection of Aquatic Life: Dissolved Oxygen (Freshwater), Canadian Council of Ministers of the Environment, Winnipeg, MB, Canada, 1999.
- 10 Chang, F. J., Tsai, Y. H., Chen, P. A., Coynel, A., and Vachaud, G.: Modeling water quality in an urban river using hydrological factors – data driven approaches, *J. Environ. Manage.*, 151, 87–96, doi:10.1016/j.jenvman.2014.12.014, 2015.
- Civanlar, M. R. and Trussell, H. J.: Constructing membership functions using statistical data, *Fuzzy Set. Syst.*, 18, 1–13, doi:10.1016/0165-0114(86)90024-2, 1986.
- 15 Deng, Y., Sadiq, R., Jiang, W., and Tesfamariam, S.: Risk analysis in a linguistic environment: a fuzzy evidential reasoning-based approach, *Expert Syst. Appl.*, 38, 15438–15446, doi:10.1016/j.eswa.2011.06.018, 2011.
- Dorfman, R. and Jacoby, H. D.: An illustrative model of a river basin pollution control, in: *Models for Managing Regional Water Quality*, edited by: Dorfman, R., Jacoby, H. D., and Thomas Jr., H. A., Harvard University Press, Cambridge, MA, USA, 84–141, doi:10.4159/harvard.9780674419216.c3, 1972.
- 20 Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimisation for conceptual rainfall–runoff models, *Water Resour. Res.*, 28, 1015–1031, doi:10.1029/91WR02985, 1992.
- 25 Dubois, D. and Prade, H.: Fuzzy sets and probability: misunderstandings, bridges and gaps, in: *Proceedings of the Second IEEE International Conference on Fuzzy Systems*, 28 March–1 April 1993, San Francisco, USA, 1059–1068, doi:10.1109/FUZZY.1993.327367, 1993.
- Dubois, D. and Prade, H.: Possibility theory and its applications: where do we stand?, in: *Springer Handbook of Computational Intelligence*, edited by: Kacprzyk, J. and Pedrycz, W., Springer-Verlag, Berlin, Heidelberg, 31–60, doi:10.1007/978-3-662-43505-2_3, 2015.
- 30

12357

- Dubois, D., Prade, H., and Sandri, S.: On possibility/probability transformations, in: *Fuzzy Logic*, edited by: Lowen, R. and Roubens, M., Springer Netherlands, Dordrecht, the Netherlands, 103–112, doi:10.1007/978-94-011-2014-2_10, 1993.
- Dubois, D., Foulloy, L., Mauris, G., and Prade, H.: Probability–possibility transformations, triangular fuzzy sets, and probabilistic inequalities, *Reliab. Comput.*, 10, 273–297, doi:10.1023/B:REOM.0000032115.22510.b5, 2004.
- 5 Elshorbagy, A., Corzo, G., Srinivasulu, S., and Solomatine, D. P.: Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology – Part 1: Concepts and methodology, *Hydrol. Earth Syst. Sci.*, 14, 1931–1941, doi:10.5194/hess-14-1931-2010, 2010.
- 10 Environment Canada: Wateroffice Hydrometric Station Meta Data, available at: https://wateroffice.ec.gc.ca/station_metadata/stationList_e.html, last access: 3 September 2015.
- Ferrero, A., Prioli, M., Salicone, S., and Vantaggi, B.: A 2-D metrology-sound probability–possibility transformation, *IEEE T. Instrum. Meas.*, 62, 982–990, doi:10.1109/TIM.2013.2246910, 2013.
- 15 Guyonnet, D., Bourguin, B., Dubois, D., Fargier, H., Come, B., and Chiles, J.–P.: Hybrid approach for addressing uncertainty in risk assessments, *J. Environ. Eng.-ASCE*, 129, 68–78, doi:10.1061/(ASCE)0733-9372(2003)129:1(68), 2003.
- Hall, M. J.: *Urban Hydrology*, Elsevier Applied Science, Essex, England, 1984.
- 20 Hauer, F. R. and Hill, W. R.: Temperature, light and oxygen, in: *Methods in Stream Ecology*, edited by: Hauer, F. R. and Lamberti, G. A., Academic Press, San Diego, CA, USA, 103–117, doi:10.1016/B978-012332908-0.50007-3, 2007.
- He, J. and Valeo, C.: Comparative study of ANNs versus parametric methods in rainfall frequency analysis, *J. Hydrol. Eng.*, 14, 172–184, doi:10.1061/(ASCE)1084-0699(2009)14:2(172), 2009.
- 25 He, J., Chu, A., Ryan, M. C., Valeo, C., and Zaitlin, B.: Abiotic influences on dissolved oxygen in a riverine environment, *Ecol. Eng.*, 37, 1804–1814, doi:10.1016/j.ecoleng.2011.06.022, 2011.
- 30 He, J., Ryan, M. C., and Valeo, C.: Changes in water quality characteristics and pollutant sources along a major river basin in Canada, in: *Environmental Management of River Basin Ecosystems*, edited by: Ramkumar, M., Kumaraswamy, K., and Mohanraj, R., Springer International Publishing, Cham, Switzerland, 525–548, doi:10.1007/978-3-319-13425-3_25, 2015.

12358

- Heddam, S.: Generalized regression neural network–based approach for modelling hourly dissolved oxygen concentration in the Upper Klamath River, Oregon, USA, *Environ. Technol.*, 35, 1650–1657, doi:10.1080/09593330.2013.878396, 2014.
- Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, *Neural Networks*, 2, 359–366, doi:10.1016/0893-6080(89)90020-8, 1989.
- 5 Huang, Y., Chen, X., Li, Y. P., Huang, G. H., and Liu, T.: A fuzzy-based simulation method for modelling hydrological processes under uncertainty, *Hydrol. Process.*, 24, 3718–3732, doi:10.1002/hyp.7790, 2010.
- Iwanyshyn, M., Ryan, M. C., and Chu, A.: Separation of physical loading from photosynthesis/respiration processes in rivers by mass balance, *Sci. Total Environ.*, 390, 205–214, doi:10.1016/j.scitotenv.2007.09.038, 2008.
- 10 Jacquin, A. P.: Possibilistic uncertainty analysis of a conceptual model of snowmelt runoff, *Hydrol. Earth Syst. Sci.*, 14, 1681–1695, doi:10.5194/hess-14-1681-2010, 2010.
- Kannel, P. R., Lee, S., Lee, Y. S., Kanel, S. R., and Khan, S. P.: Application of water quality indices and dissolved oxygen as indicators for river water classification and urban impact assessment, *Environ. Monit. Assess.*, 132, 93–110, doi:10.1007/s10661-006-9505-1, 2007.
- 15 Kasiviswanathan, K. S. and Sudheer, K. P.: Quantification of the predictive uncertainty of artificial neural network based river flow forecast models, *Stoch. Env. Res. Risk A.*, 27, 137–146, doi:10.1007/s00477-012-0600-2, 2013.
- 20 Kasiviswanathan, K. S., Cibin, R., Sudheer, K. P., and Chaubey, I.: Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations, *J. Hydrol.*, 499, 275–288, doi:10.1016/j.jhydrol.2013.06.043, 2013.
- Kaufmann, A. and Gupta, M. M.: *Introduction to Fuzzy Arithmetic: Theory and Applications*, Van Nostrand Reinhold Company, New York, NY, USA, 1985.
- 25 Khan, U. T. and Valeo, C.: Predicting dissolved oxygen concentration in urban watersheds: a comparison of fuzzy number based and Bayesian data-driven approaches, in: *Proceedings of the International Conference on Marine and Freshwater Environments*, 6–8 August 2014, St John's, Canada, 2014a.
- Khan, U. T. and Valeo, C.: Peak flow prediction using fuzzy linear regression: case study of the Bow River, in: *Proceedings of the International Conference on Marine and Freshwater Environments*, 6–8 August 2014, St Johns, NFLD, Canada, 2014b.
- 30 Khan, U. T. and Valeo, C.: A new fuzzy linear regression approach for dissolved oxygen prediction, *Hydrolog. Sci. J.*, 60, 1096–1119, doi:10.1080/02626667.2014.900558, 2015a.

12359

- Khan, U. T. and Valeo, C.: Short-term peak flow rate prediction and flood risk assessment using fuzzy linear regression, *J. Environ. Inform.*, in press, 2015b.
- Khan, U. T., Valeo, C., and He, J.: Non-linear fuzzy-set based uncertainty propagation for improved DO prediction using multiple-linear regression, *Stoch. Env. Res. Risk A.*, 27, 599–616, doi:10.1007/s00477-012-0626-5, 2013.
- 5 Klir, G. J. and Parviz, B.: Probability-possibility transformations: a comparison, *Int. J. Gen. Syst.*, 21, 291–310, doi:10.1080/03081079208945083, 1992.
- Lane, R. J.: *The Water Survey of Canada: Hydrometric Technician Career Development Program Lesson Package No. 10.1 – Principles of Discharge Measurement*, available at: <http://publications.gc.ca/site/eng/463990/publication.html> (last access: 23 November 2015), 1999.
- 10 Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. P.: Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions, *Environ. Model. Softw.*, 25, 891–909, doi:10.1016/j.envsoft.2010.02.003, 2010.
- 15 Martin-Clouaire, R., Cazemier, D. R., and Lagacherie, P.: Representing and processing uncertain soil information for mapping soil hydrological properties, *Comput. Electron. Agr.*, 29, 41–57, doi:10.1016/S0168-1699(00)00135-6, 2000.
- Mauris, G.: A review of relationships between possibility and probability representations of uncertainty in measurement, *IEEE T. Instrum. Meas.*, 62, 622–632, doi:10.1109/TIM.2012.2218057, 2013.
- 20 Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *T. ASABE*, 50, 885–900, doi:10.13031/2013.23153, 2007.
- 25 Mujumdar, P. P. and Ghosh, S.: Modeling GCM and scenario uncertainty using a possibilistic approach: application to the Mahanadi River, India, *Water Resour. Res.*, 44, W06407, doi:10.1029/2007WR006137, 2008.
- Napolitano, G., Serinaldi, F., and See, L.: Impact of EMD decomposition and random initialization of weights in ANN hindcasting of daily stream flow series: an empirical examination, *J. Hydrol.*, 406, 199–214, doi:10.1016/j.jhydrol.2011.06.015, 2011.
- 30 Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models: Part I. A discussion of principles, *J. Hydrol.*, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

12360

- Neupane, A., Wu, P., Ghanbarpour, R., and Martin, N.: Bow River Phosphorus Management Plan: Water Quality Modeling Scenarios, Alberta Environment and Sustainable Resource Development, Edmonton, AB, Canada, 2014.
- Oussalah, M.: On the probability/possibility transformations: a comparative analysis, *Int. J. Gen. Syst.*, 29, 671–718, doi:10.1080/03081070008960969, 2000.
- 5 Pogue, T. R. and Anderson, C. W.: Processes Controlling Dissolved Oxygen and pH in the Upper Willamette River and Major Tributaries, Oregon, 1994, Water Resources Investigations Report 95-4205, US Geological Survey, Portland, OR, USA, 1995.
- Robinson, K. L., Valeo, C., Ryan, M. C., Chu, A., and Iwanyshyn, M.: Modelling aquatic vegetation and dissolved oxygen after a flood event in the Bow River, Alberta, Canada, *Can. J. Civil Eng.*, 36, 492–503, doi:10.1139/L08-126, 2009.
- 10 Sadiq, R., Rodriguez, M. J., Imran, S. A., and Najjaran, H.: Communicating human health risks associated with disinfection by-products in drinking water supplies: a fuzzy-based approach, *Stoch. Env. Res. Risk A.*, 21, 341–353, doi:10.1007/s00477-006-0069-y, 2007.
- 15 Serrurier, M. and Prade, H.: An informational distance for estimating the faithfulness of a possibility distribution, viewed as a family of probability distributions, with respect to data, *Int. J. Approx. Reason.*, 54, 919–933, doi:10.1016/j.ijar.2013.01.011, 2013.
- Shimazaki, H. and Shinomoto, S.: A method for selecting the bin size of a time histogram, *Neural Comput.*, 19, 1503–1527, doi:10.1162/neco.2007.19.6.1503, 2007.
- 20 Shrestha, D. L. and Solomatine, D. P.: Data-driven approaches for estimating uncertainty in rainfall–runoff modelling, *Int. J. River Basin Manage.*, 6, 109–122, doi:10.1080/15715124.2008.9635341, 2008.
- Singh, K. P., Basant, A., Malik, A., and Jain, G.: Artificial neural network modelling of the river water quality – a case study, *Ecol. Model.*, 220, 888–895, doi:10.1016/j.ecolmodel.2009.01.004, 2009.
- 25 Solomatine, D. P. and Ostfeld, A.: Data–driven modelling: some past experiences and new approaches, *J. Hydroinform.*, 10, 3–22, doi:10.2166/hydro.2008.015, 2008.
- Solomatine, D. P., See, L. M., and Abrahart, R. J.: Data–driven modelling: concepts, approaches and experiences, in: *Practical Hydroinformatics*, edited by: Abrahart, R. J., See, L. M., and Solomatine, D. P., Springer, Berlin, Heidelberg, Germany, 17–30, doi:10.1007/978-3-540-79881-1_2, 2008.
- 30

12361

- Van Steenberg, N., Ronsyn, J., and Willems, P.: A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication, *Environ. Model. Softw.*, 33, 92–105, doi:10.1016/j.envsoft.2012.01.013, 2012.
- 5 Verhoest, N. E. C., De Baets, B., Mattia, F., Satalino, G., Lucau, C., and Defourny, P.: A possibilistic approach to soil moisture retrieval from ERS synthetic aperture radar backscattering under soil roughness uncertainty, *Water Resour. Res.*, 43, W07435, doi:10.1029/2006WR005295, 2007.
- Wen, X., Fang, J., Diao, M., and Zhang, C.: Artificial neural network modelling of dissolved oxygen in the Heihe River, northwestern China, *Environ. Monit. Assess.*, 185, 4361–4371, doi:10.1007/s10661-012-2874-8, 2013.
- 10 YSI Inc.: YSI 5200A Multiparameter Monitor and Control Specifications, available at: , last access: 23 November 2015.
- Zadeh, L. A.: Fuzzy sets, *Inform. Control*, 8, 338–353, doi:10.1016/S0019-9958(65)90241-X, 1965.
- 15 Zadeh, L. A.: Fuzzy sets as a basis for a theory of possibility, *Fuzzy Set. Syst.*, 1, 3–28, doi:10.1016/0165-0114(78)90029-5, 1978.
- Zhang, K.: Modeling uncertainty and variability in health risk assessment of contaminated sites, PhD thesis, Civil Engineering, University of Calgary, Calgary, Canada, 290 pp., 2009.
- Zhang, K. and Achari, G.: Correlations between uncertainty theories and their applications in uncertainty propagation, in: *Safety, Reliability and Risk of Structures, Infrastructures and Engineering Systems*, edited by: Furuta, H., Frangopol, D. M., and Shinozuka, M., Taylor and Francis Group, London, UK, 1337–1344, doi:10.1201/9781439847657-c20, 2010.
- 20 Zhang, K., Li, H., and Achari, G.: Fuzzy-stochastic characterisation of site uncertainty and variability in groundwater flow and contaminant transport through a heterogeneous aquifer, *J. Contam. Hydrol.*, 106, 73–82, doi:10.1016/j.jconhyd.2009.01.003, 2009.
- 25

12362

Table 1. A summary of low DO events in the Bow River between 2004 and 2012 and the corresponding minimum acceptable DO concentration guidelines.

Year	DO < 5 mgL ^{-1a}	DO < 6.5 mgL ^{-1b}	DO < 9.5 mgL ^{-1c}
2004	25	41	107
2005	1	26	133
2006	25	70	164
2007	0	27	182
2008	0	5	130
2009	0	15	85
2010	0	0	180
2011	0	0	122
2012	0	0	76
Total	51	184	1179

^a For the protection of aquatic life for 1-day (AENV, 1997).

^b For the protection of aquatic life in cold, freshwater for other-life (i.e. not early) stages (CCME, 1999).

^c For the protection of aquatic life in cold, freshwater for early-life stages (CCME, 1999).

12363

Table 2. Selected values for P_{Ci} for the FNN optimisation.

μ	P_{Ci} (%)
1.00	0.00
0.80	20.00
0.60	40.00
0.40	60.00
0.20	80.00
0.00	99.50

12364

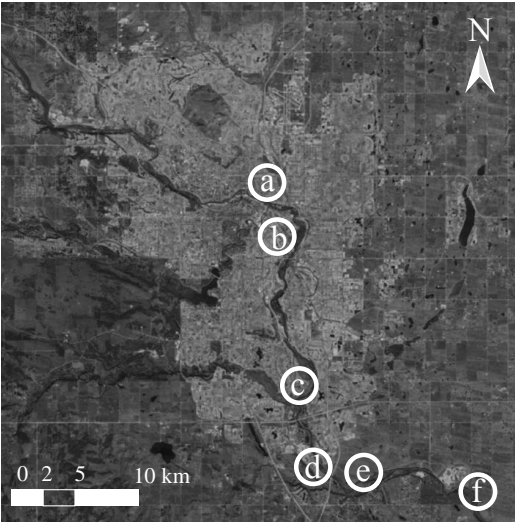


Figure 1. An aerial view of the City of Calgary, Canada showing the locations of @ the flow monitoring site Bow River at Calgary (Water Survey of Canada ID: 05BH004), three wastewater treatment plants at @ Bonnybrook, @ Fish Creek, and @ Pine Creek, and two water quality sampling sites @ Stier's Ranch and @ Highwood.

12367

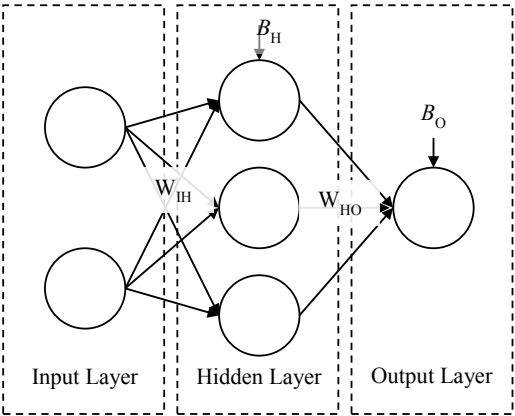


Figure 2. An example of a three-layer multilayer perceptron feed-forward ANN, with two input neurons, the hidden layer neurons, and one output neuron. W_{IH} are the weights between the input and hidden layer, W_{HO} are the weights between the hidden and output layer, B_H are the biases in the hidden layer, and B_O is the bias in the output layer.

12368

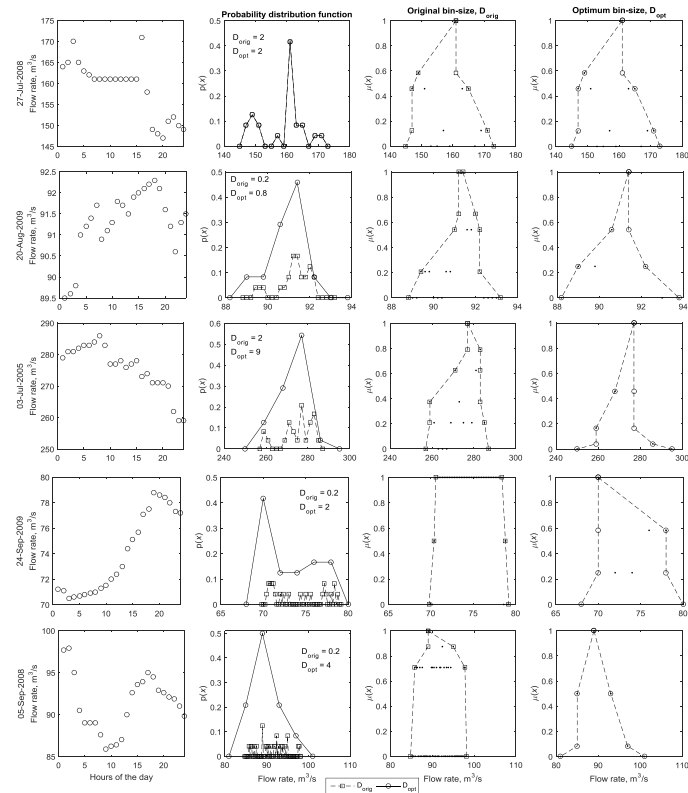


Figure 3. Sample results of probability-possibility transformation for flow rate, Q .

12369

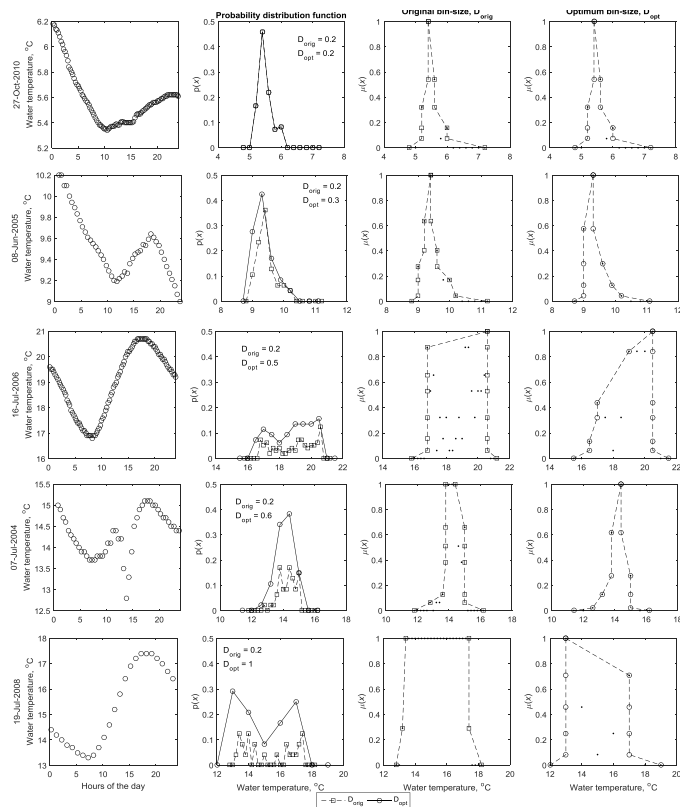


Figure 4. Sample results of probability-possibility transformation for water temperature, T .

12370

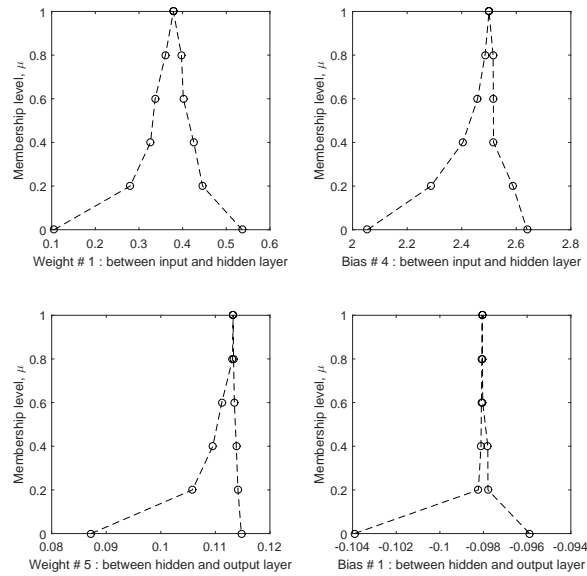


Figure 5. Sample plots of the produced membership functions for the weights and biases of the fuzzy neural network.

12371

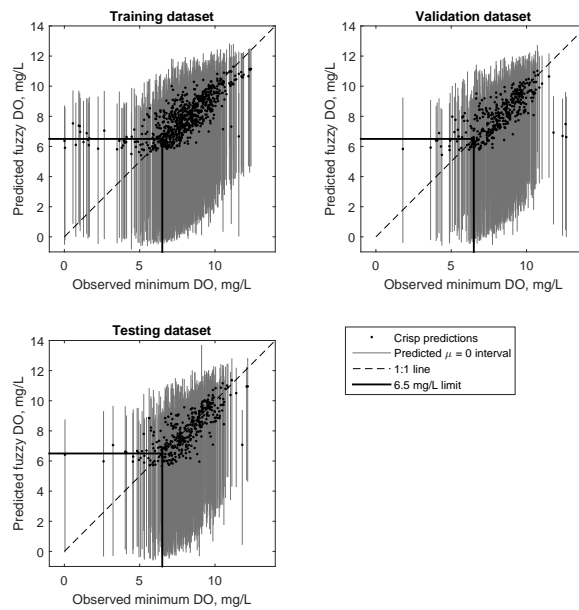


Figure 6. A comparison of the predicted and observed minimum DO at the $\mu = 0$ interval (black line) and at $\mu = 1$ (black dots).

12372

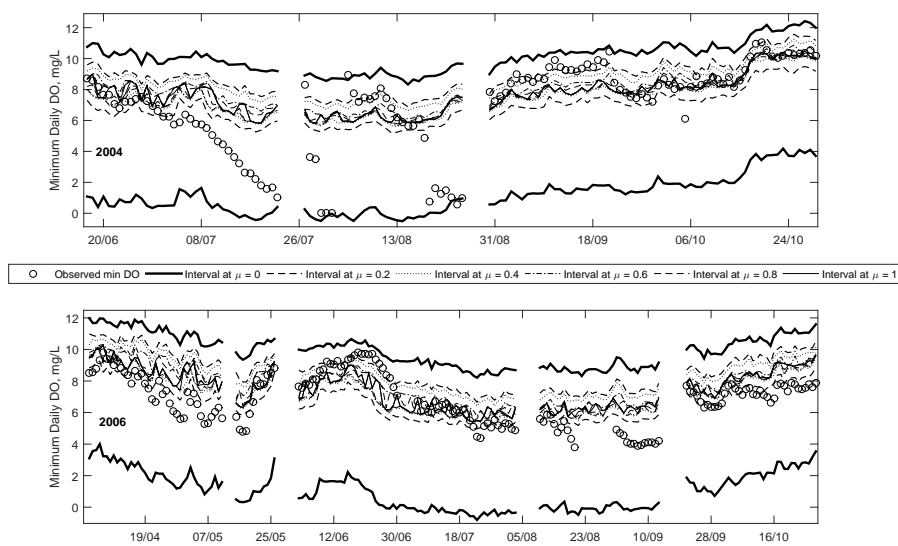


Figure 7. A comparison of the observed and predicted minimum DO trends for: 2004 (top panel), and 2006 (bottom panel).

12373

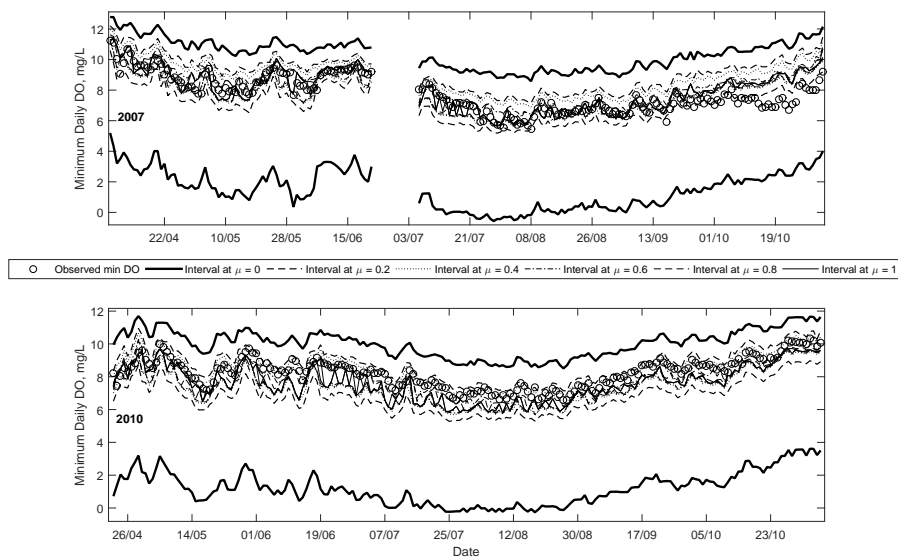


Figure 8. A comparison of the observed and predicted minimum DO trends for three sample years: 2007 (top panel) and 2010 (bottom panel).

12374

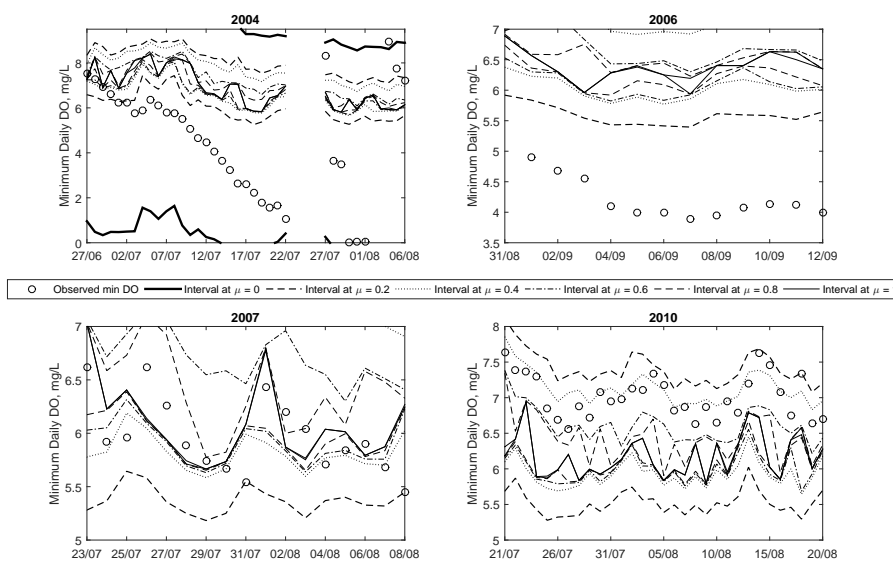


Figure 9. Zoomed in views of the trend plots for four sample year corresponding to important periods with low DO occurrences.

12375

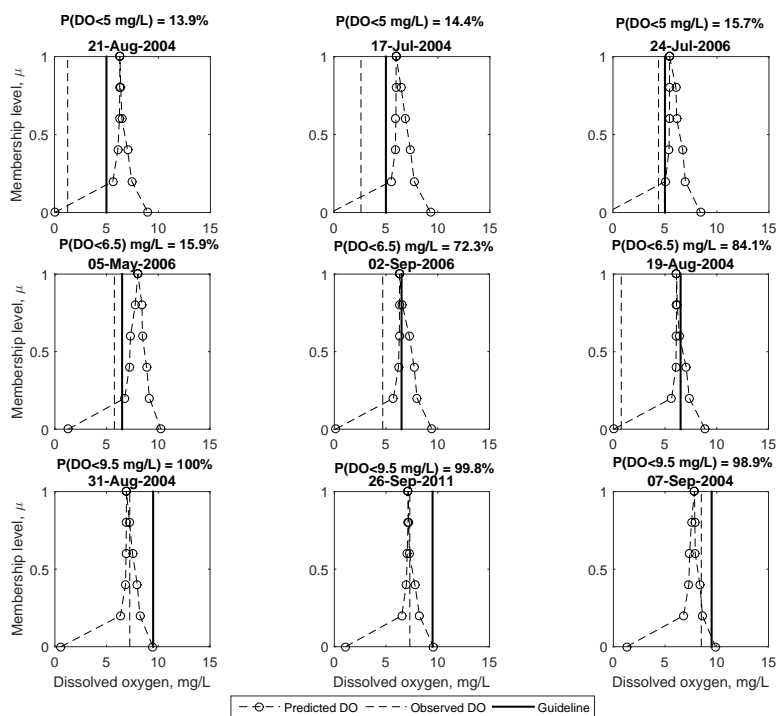


Figure 10. Sample plots of low DO events and the corresponding risk of low DO calculated using a possibility-probability transformation for the 5 mg L⁻¹ (top panels), 6.5 mg L⁻¹ (middle panels), and 9.5 mg L⁻¹ (bottom panels) guideline.

12376