**Response**

Title: Dissolved oxygen prediction using a possibility-theory based fuzzy neural network
Authors: U. T. Khan and C. Valeo
Journal: *Hydrology and Earth System Sciences*
Manuscript No.: hess-2015-446

**Note that all page and line numbers correspond to the track-change copy of the manuscript appended to this document.**

## Response to Anonymous Referee #1

> **RC 1:** "The authors present the application of the Fuzzy Neural Network (originally proposed by Alvisi and Franchini, 2011) for the prediction of the dissolved oxygen concentration in a river. The topic is of interest and within the scope of the journal. The manuscript is well written and technically sound, even though some sections could be shortened. As properly pointed out by the authors in the conclusions, *"the proposed model refines the exiting model by (i) using possibility theory based intervals to calibrate the neural network (rather than arbitrarily selecting confidence intervals), and (ii) using fuzzy number inputs rather than crisp inputs."* Indeed, the first aspect represents a valuable, but rather limited, step forward with respect to the existing model. As far as the second aspect concerns, I really appreciate both the idea of considering the inputs of the FNN as fuzzy numbers and the approach used to define these fuzzy inputs."

Thank you for these positive comments. We will endeavour to shorten the length of the final revised manuscript where possible.

> **RC 2.** "Unfortunately, the manuscript misses to point out the benefits of using the fuzzy inputs. A comparison of the performances of the prediction model featuring fuzzy inputs with respect to the prediction model using non-fuzzy inputs is completely missing. Does the application of fuzzy inputs allows for a more accurate prediction of the DO and, most important, for a reduction of the output uncertainty? Indeed, the discussion of the result is mainly focused on the benefits of using a FNN with respect to a traditional NN in which uncertainty is disregarded, but this should not be the main task of the manuscript, given that benefits of FNN have already been pointed out in other studies, whereas the attention should be focused on the application of Fuzzy inputs."

We apologise for not including more details of the specific advantages of the fuzzy number inputs in the FNN. We have included the following statement in the manuscript:

> Page 6, Line 26: "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."
>
> Page 21, Line 8: "The impact of this revision is that when there is known variance or uncertainty in the input dataset, it should be incorporated into the model structure. In Eqns. 13 and 14, this is done through the use of fuzzy rather than crisp inputs."

We have also included the advantages of the proposed method to create fuzzy inputs from observations:

> Page 12, Line 26: "This updated method requires no assumptions regarding the distribution of the underlying data or selection of an arbitrary bin-size, has the flexibility to create different shapes of fuzzy numbers depending on the distribution of the underlying data, and allows multiple elements to have equal $\mu = 1$."

A comparison between the existing FNN method with crisp inputs and the proposed FNN with fuzzy inputs has been included in the in the revised manuscript. Please see the extensive changes and additions in the Section 3.2, particularly page 36 (line 14) and the revision of Figs. 5 and 6, as well as the addition of Figs 10. And 11.

For clarification we would like to highlight a few issues related to this comparison:

1. Conceptually, an FNN model with crisp inputs and fuzzy inputs are completely different, making a direct comparison difficult (or at least not straightforward). It is not just a case of comparing error metrics, or percent of data captured within intervals (e.g. as shown in Tables 3 and 4 in the initial manuscript). This is because these two approaches are essentially modelling the system completely differently. In the crisp input case, the input uncertainty is completely ignored even though this data is available. This is essentially making a complex problem less complex by limiting the amount of data that is used in the model. Also, there is no definitive answer to what the crisp input should be: is it the mean daily value? The median value? Or the corresponding value from the fuzzy number at $\mu = 1$? (n.b. this final option is what we have selected for our comparison to allow for the closest approximation between the two approaches). On the other hand, the fuzzy number based input use all the available information (i.e. hourly observations), and condenses it into one fuzzy number. Arguably, in this approach the complexity of the system is not being ignored (by reducing the highly variable/uncertain inputs into crisp, single-values inputs). Thus, any analysis of the performance of these two methods should highlight that the proposed method accomplishes something that the existing method cannot.

2. Currently, they are no suitable performance metrics to compare fuzzy number based models with each other (or for that matter even with other crisp models). While the Nash-Sutcliffe Efficiency (or other similar metrics) can be calculated on an $\alpha$-cut interval basis, these values do not represent the overall model performance nor does calculating the amount of extreme values within the fuzzy interval. A suitable alternative may be to use the training method of the FNN: to see if the percentage of data captured within each interval is similar (see Table 4 in the manuscript). However, given that this is an optimisation problem, and both methods will have the same tolerances, the result is expected to be the same for both (at least for the

training dataset), although the computation time will differ. Thus, the computation time may be used as an effective metric, but even this does not account for the fact that in the proposed method the uncertainty in the inputs is being included in the model, and hence, the extra computational cost is acceptable if the input uncertainty is high (as in the case in this research where flowrate and water temperature are used as inputs).

3. With respect to “more accurate predictions”: both the proposed and existing FNN method are designed to capture the same amount of data ($P_{CI}$) in a given α-cut interval, whilst searching for the minimum width of that interval. An analysis comparing the widths of the intervals has been included in the revised version of the manuscript (Page 37, Line 7). However, it is worthwhile to point out that in the proposed FNN method more data (due to the construction of fuzzy numbers using hourly observation) was used to train the model, and represents the real uncertainty in the inputs. Whereas, selecting a modal value for the crisp FNN method (selected as the \mu = 1 value) for this research, but other values (e.g. daily mean or median may also be selected) do not capture the input data uncertainty. Thus, the proposed method gives similar results (i.e. the training criteria) but using the input uncertainty as well.

4. With respect to a “reduction in output uncertainty”: the uncertainty is reduced because by using fuzzy inputs all the observed data is used to calibrate the model, and this gives a full spectrum of *possible* outcomes. In other words, the uncertainty is lower because all possibilities of the output value have been mapped out using all available input data. In the crisp input case, the uncertainty is by definition higher, since the variability or uncertainty in the input data has not been accounted for (only one crisp value is used), as discussed in Point 3 above. Using this definition of uncertainty means that there is a reduction in output uncertainty (since more information is used), while not necessarily meaning that the predicted intervals are smaller or more accurate.

5. Based on our experience using fuzzy number based data-driven methods for hydrological and environmental applications, there is still a need and demand to compare new methods (like the one proposed in this research) with existing non-fuzzy methods to provide a baseline reference with other literature (see a brief discussion on page 18, line 3). Thus, we think it is important to include this comparison in this manuscript so that readers who may be unfamiliar with fuzzy number and possibility theory based methods may be able to directly compare results from this research to other NN based results. Secondly, while we agree that the benefits of the FNN method has been highlighted in previous studies, we would like to highlight that the FNN used in this research uses a different training criteria (i.e. the selection of $P_{CI}$ shown in Table 2 in the manuscript) compared to previous work, so there is a benefit in showing the results for this comparison.

**RC 3.** “Furthermore, I have some concerns also on the benefits of using the FNN with respect to a deterministic NN. Indeed, the authors state that (page 12351) “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." Capability of the FNN of identifying very low DO values is certainly appreciable, but on the other hand, by looking at figure 6, it seems that most of the predicted fuzzy DO numbers features a support which in some way intersects very low (i.e. <5 mg/l) DO values. In other words, according to the criteria proposed by the authors how many events would be flagged? And, how many of these flagged events were low (i.e. <5 mg/l) observed DO events and how many would have been false alarms?"

First, we would like to highlight that the data has been filtered to include only data from the April to October period for each year (to remove the ice-free period in the river). This means that the entire analysis has been conducted on the time period that is most susceptible to low DO (due to high water temperature). In other words, we are focusing on the most critical time period already. Thus, it is expected that the majority of the days will have some *possibility* of low DO (this is clarified on (Page 39, Line 18). This phenomenon is correctly reproduced in Figure 6 that shows that indeed there is a possibility (though typically at low membership levels) to predict "very low DO" (< 5mg/L) values.

Second, it is worth noting that using possibility theory means that "something should be possible before it is probable", i.e. Zadeh's consistency principle. Thus, the fact the FNN model predicts a possibility of very low DO does not necessarily mean that there will be a significant or high probability of this event to occur (this is discussed in detail on Page 10, Line 25, and Page 39, Line 22). In fact, this can be seen in the trend plots (Figs. 7 and 8 in the manuscript) that show that the produced fuzzy number membership functions are highly skewed (see Page 34, Line 12 and more examples in Figure 12), i.e. the predictions at the lower limit of the α-cut at $\mu = 0$ are much lower than the rest of the membership function. In the possibility-probability framework adopted in this research (see Sections 2.2 and 2.4 in the manuscript), this means that the highly skewed membership functions translate into very low probability events (based on Equation 20 in the manuscript).

Third, we have identified (Page 34, Line 22) that the 2004 data has contributed to the wide intervals in the predictions. The rapid decrease in DO in the 2004 data (which are likely due to instrument error), along with the optimisation constraints that requires 99.5% of the data to be included in the predicted interval at $\mu = 0$, means that the produced output will include these outliers at the expense of creating wider intervals. However, we have noted that as more data is available and include in the model, the 0.5% of data points that will no longer be captured within the $\mu = 0$ interval are likely to be these outliers, resulting in narrower predicted intervals (see Page 35, Line 1).

Finally, as requested by the referee, the revised manuscript includes a summary of the flagged days using our criteria for low DO as well as those incorrectly identified at the given threshold, i.e. "false alarms" (Page 39, Line 28).

We hope that this will satisfy all of the Referee's comments.

**RC 1**. What is the novelty of this study? What do the authors expect international readers (who are not interested in the study region) to learn from reading this paper.

There are three novel contributions presented in this research (which are listed in the manuscript). In addition to this, the proposed approach for hydrological prediction, uncertainty and risk analysis can be extended to many other applications. In more specific detail:

- a new method to construct fuzzy numbers from observed environmental and hydrological data is presented. Many fuzzy number based applications suffer from the fact that there is no widely accepted, consistent and objective method to construct fuzzy numbers from observations. We have attempted to address this issue by introducing a new two-step procedure where we first estimate the underlying, unknown probability mass function using a bin-size optimisation procedure, and then use a probability-to-possibility transformation to convert this to fuzzy number membership function. A number of different examples are used to demonstrate the advantage and suitability of this method.
- An existing fuzzy neural network (FNN) method is improved in this paper by proposing the use of possibility theory-based intervals for training the neural network. This replaces a somewhat arbitrary training criteria with a more objective criterion. Specifically, the original FNN uses pre-selected confidence intervals to define the amount of data captured within each fuzzy interval (i.e. $\alpha$-cut), for example 100% at $\mu = 0$, 99% at $\mu = 0.25$. We use a relationship proposed by Serrurier & Prade (2013) to define the amount of data captured within $\alpha$-cut. In doing so, the full spectrum of possible values are included in these calculations. This is so that modellers and end-users who are interested in events not included in the original, pre-determined criteria can use an objective (i.e. based on possibility theory) method to design their FNN.
- The existing FNN is further refined by allowing the use of fuzzy inputs, along with the fuzzy weights, biases and outputs. Current methods only allowed crisp (i.e. non-fuzzy inputs) in the FNN. This has significant advantages over current methods, namely that the uncertainty in the input data is also accounted for in predicting DO concentration. In other words, the model output has accounted for the total uncertainty, in the weights and biases, as well as the inputs.
- The approach used in this study (data-driven modelling with fuzzy numbers when the underlying physical system is complex and poorly understood) can be extended to many other applications dealing with water quality in rivers, in flood risk predictions, or hydrological and environmental applications that suffer from similar issues, namely a complex system with many source of uncertainty. International readers will benefit from potentially applying this technique in their own watersheds to improve water quality prediction, and the associated risk analysis presented in this research. As mentioned above, this paper also presents a new method to construct fuzzy numbers that relies on minimal assumptions of the underlying data. This directly addresses a major need in the hydrological community. Lastly, readers will benefit from seeing the refinements to an existing FNN model; these refinements create a more transparent model structure (i.e. objective criteria for training) and include the use of fuzzy inputs (which is necessary in many hydrological cases where input uncertainty is present).

**RC 2**. The authors didn' t define statistical parameters of input and output variables. The study will make more sense in interpretation of statistical parameters.

We apologize for these omissions. The revised manuscript includes:

Page 9, Line 4 : "The mean annual water temperature ranged between 9.23 and 13.2◦C, the annual mean flow rate was between 75 and 146 m3s−1, and the mean annual minimum daily DO was between 6.89 and 9.54 mgL-1, for the selected period."

**RC 3**. How many datas are used in this study? The authors didn't define to use training datas and test datas this study.

We apologize for this omission as well. A total of 9 years of data was used for this research (from 2004 to 2012); the data were filtered to include data only from the ice-free period (April to October of each year). The total amount of daily data was 1639 days (a yearly breakdown is shown in Table R1 and is now included in the revised manuscript in the revised Table 1).

**Table R1: Summary of amount of data used from each year**

| Year | Number of days |
|---|---|
| 2012 | 206 |
| 2011 | 204 |
| 2010 | 207 |
| 2009 | 96 |
| 2008 | 163 |
| 2007 | 211 |
| 2006 | 209 |
| 2005 | 208 |
| 2004 | 135 |
| **Total** | **1639** |

The amount of data used for training, validation and testing followed a 50–25–25% Split (randomly divided into each section). This is outlined in Section 2.3.3 (Page 23, Line 3) of the manuscript.

**RC 4**. The authors didn't write key board. What are key board for the manuscript?

The key words associated with this manuscript are listed below, and we will include these in the final version

Page 1, Line: "**Keywords:** dissolved oxygen; water quality; artificial neural networks; fuzzy numbers; fuzzy neural networks; risk analysis; uncertainty"

**RC 5**. Why is not continuous in Figüre 7, 8, and 9.

There are a number of missing data throughout the dataset due to numerous reasons, ranging from sampler error or no data recorded (as received from the data providers Environment Canada or the City of Calgary), or due to the data filters used for reasons highlighted in Section 2.1 (only ice-free period was considered, see Page 7, Line 23). We ignored all missing data from our analysis. The data that we used was thus for days where error-free data existed for each of the three parameters (flowrate, temperature and DO). Thus, Figures 7, 8 and 9 show some gaps in the trends for days when no data was collected, and hence no subsequent prediction was made.

**RC 6**. Fuzzy neural networks method is too large, it should be less the part.
**RC 7**. Results and discussion is too large, The authors should reduce the part.

Thank you for these suggestions, we will endeavour to reduce the length of the final manuscript.

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

**U. T. Khan[1] and C. Valeo[2]**

[1] PhD Candidate, Mechanical Engineering, University of Victoria.

PO Box 1700, Stn. CSC, Victoria, BC, V8W 2Y2, Canada.

[2] Professor, Mechanical Engineering, University of Victoria.

PO Box 1700, Stn. CSC, Victoria, BC, V8W 2Y2, Canada.

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

## 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 performance is compared with a fuzzy neural network with crisp inputs, as well as with a traditional neural network. 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.

**Keywords:** dissolved oxygen; water quality; artificial neural networks; fuzzy numbers; fuzzy neural networks; risk analysis; uncertainty

# 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 m³ 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 & Valeo, 2015a; 2015b). 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 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 & 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 studies, 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 if 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 impacts.

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~~ the interaction of numerous factors~~,~~ over a relatively small area and across different temporal scales means that DO trends and variability in urban areas are more difficult to ~~predict~~ model 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 understood, 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 (Soloma~~n~~tine & 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 & Achari 2010) – in other words when epistemic uncertainty 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 & 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. Heddam, 2014 and Ay & 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 types of 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 & 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 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 & Franchini, 2011). Khan et al. (2013) and Khan & 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 & 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 (Alvisi & Franchini, 2011; Antanasijević et al., 2014). A fuzzy neural network method proposed by Alvisi & 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 & 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. The impact of using fuzzy inputs and the new training criteria is evaluated by comparing results to the existing FNN method (by Alvisi & Franchini, 2011) as well as with a traditional, crisp ANN.

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 $m^3 s^{-1}$, an average width and depth of 100 m and 1.5 m, respectively (Khan & 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 & 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 minutes (from 2004 to 2005), and every 15 minutes (from 2006 to 2007). The station was then moved to Stier's Ranch and sampled data every hour (in 2008) and every 15 minutes (2009 to 2011). The monitoring site was moved further downstream to its current location (at Highwood) in 2012 where it samples every 15 minutes. 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.

Note that even though daily minimum DO was observed to be below 5 mg $L^{-1}$ on several occasions in 2004 and 2006 (in Table 1), the minimum DO was below 9.5 mg $L^{-1}$ only 107 and 164 days, respectively, for those two years. In contrast, in 2007 and 2010, no observations below 5 mg $L^{-1}$ are seen yet 182 and 180 days, respectively, below the 9.5 mg $L^{-1}$ guideline were seen for those years. The total amount of days below 9.5 mg $L^{-1}$ constitute approximately 90% of all observations for those years. This highlights that despite no DO events below 5 mg $L^{-1}$, 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 m$^3$ s$^{-1}$,-and the mean annual minimum daily DO was between 6.89 and 9.54 mgL$^{-1}$, 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 & 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, $\mu$, between 0 and 1 in the fuzzy set *A*. If $\mu$ equals 0, then *x* does not belong in 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 & 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 mg L$^{-1}$ ($\mu$ = 1), with a support about the modal value between ($\mu$ = $0^L$) and 12 mg L$^{-1}$ ($\mu$ = $0^R$). 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 & Prade, 2015; Dubois et al., 2004).

However, recent research (Khan et al., 2013; Khan & Valeo, 2014a; 2014b; 2015a; 2015b) 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 12] mg $L^{-1}$, has a uniform probability distribution bounded between 8 and 12 mg $L^{-1}$ with a mean value of 10 mg $L^{-1}$, suggesting that values between the support are equally likely to occur. It 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 & Parvais (1992), Oussalah (2000), Jaquin (2010) Mauris (2013) and Dubois & 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 & Prade, 1993; Civanlar & 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~~In this framework, 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:

if $p(x_1) > p(x_2) > \ldots > p(x_n)$,

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

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

The transformation is given by:

For $p(x_1) > p(x_2) > \ldots > p(x_n)$:

$\pi(x_1) = 1$

$$f(x) = \begin{cases} \sum_{j=i}^{n} p_j\,, & if\ p_{i-1} > p_i \\ \pi(x_{i-1}), & else \end{cases} \qquad (1)$$

where the $x_i$ are elements of a fuzzy number $A$, $\pi(x_i)$ is the possibility of element $x_i$, and $p(x_i)$ is the probability of element $x_i$. The concept of this transformation may be more illustrative when viewed in the continuous case: for any interval [$a$, $b$], the membership level $\mu$ (where $\pi(a) = \pi(b) = \mu$) is equal to the sum of the areas under the probability density function curve between ($-\infty$, $a$) and ($b$, $\infty$) (Zhang et al., 2009). It is important to highlight that this particular transformation has an inverse transformation associated with, where a probability distribution can be estimated from the possibility distribution.

However, a major drawback of this transformation is that it theoretically requires a full description of the probability density function, or in the finite case, the probability associated with each element of the fuzzy number, the probability mass function. For many hydrological~~y~~ ~~and time series based~~ applications this might not be possible because the hourly time series data ~~(that is typically collected)~~ may not adequately fit the mould of a known class of probability density functions, or one distribution amongst many alternatives may have to be selected based on best-fit. This best-fit function may not be universal, e.g. data from one ~~24 hour~~24-hour period may be best described by one class or family of probability density function, while the next day by a completely different class of density function. This means working with multiple classes of distribution functions for one application, which can be cumbersome. Also, given that each day may only have 24 data points (or fewer on days with missed samples) it is difficult to select one particular function.

In previous research by Khan & Valeo (2015a), a new approach to create a fuzzy number based on observations was developed. This process used a histogram-based approach to estimate the probability mass function of the observations, and then Eq. 1 was used to estimate the membership function of the fuzzy number. To create the histogram, the bin-size was selected based on the extrema observations for a given day and the number of the observations. A linear interpolation scheme was then used to calculate ~~the values of~~ the fuzzy number at five predefined membership levels. This method has a few short-comings, namely: the bin-size selection was arbitrarily selected based on the magnitude and number of observations which does not necessarily result in the optimum bin-size. This lack of optimality means that the resulting histogram may either be too smooth so as not to capture the variability between membership levels, or too rough and uneven so that the underlying shape of the membership function is difficult to discern. This is a common issue with histogram selection in many applications (Shimazaki & Shinomoto, 2007). Secondly, the aforementioned transformation used by Khan and Valeo (2015a) only allows one element to have $\mu = 1$ when $p(x)$ is maximum. However, there are a number of cases (e.g. bimodal distributions, or arrays when all elements are equal) where multiple elements have joint-equal maximum $p(x)$~~)~~, and hence multiple elements with $\mu = 1$. This means that all elements *within* the $\alpha$-cut interval $[a\ b]_{\mu=1}$ (where $a$ and $b$ are the minimum and maximum elements with $\mu = 1$) must by definition also have a membership level equal to 1. Thus, a method is necessary to be flexible enough to accommodate these types of issues.

In this research, a two-step procedure is proposed to create fuzzy numbers on the inputs (i.e. $\boldsymbol{Q}$ and $\boldsymbol{T}$) using hourly (or sub-hourly) observations. First, a bin-size optimisation method is used (an extension of an algorithm proposed by Shimazaki & Shinomoto, 2007) to create histograms to represent the estimate of discretised probability density functions of the observations. This estimate of the probability distribution is then transformed to the membership function of the fuzzy number using a new numerical procedure and the transformation principles described in Eq. 1. This updated method requires no assumptions regarding the distribution of the underlying data or selection of an arbitrary bin-size, has the flexibility to create different shapes of fuzzy numbers depending on the distribution of the underlying data, and allows multiple elements to have equal $\mu = 1$. The proposed algorithm is described in the proceeding section.

### 2.2.1 A new algorithm to create fuzzy numbers

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

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

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[·] is the expectation. In practice, $E_{\text{MISE}}$ cannot be directly calculated since the underlying distribution is unknown and thus, an estimate of the $E_{\text{MISE}}$ 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 $i^{\text{th}}$ bin of the histogram follow a Poisson point process. This means that the events in two disjoint bins (e.g., the $i^{\text{th}}$ and $i+1^{\text{th}}$ 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 & Shinomoto, 2007).

Using this property, the optimum bin-size can be found as follows. Let $\boldsymbol{X}$ be the input data vector for the observation period ($P$), e.g., a [24×1] vector corresponding to hourly samples for a given day. The elements in $\boldsymbol{X}$ are binned into $N$ bins of equal bin-size $D$. The number of events $k_i$ in each $i^{\text{th}}$ 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 $E_{\text{MISE}}$ listed in Eq. (2) and is derived by removing the terms from $E_{\text{MISE}}$ 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 & Shinomoto, 2007). The objective then is to search for $D_{\text{opt}}$: 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 bin-size $D$ is kept constant, but the first and last points are perturbed by a small value $\delta$: $[a–D/2+\delta, a+D/2+\delta]$, where $\delta$ ranges incrementally between 0 and $D$. Using these different values of $\delta$ 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, $\boldsymbol{Q}$ and $\boldsymbol{T}$. The array of daily data, $\boldsymbol{X}$, (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~~nine-year period. The bin-size was calculated for each day as follows:

$$D = \frac{x_{\max} - x_{min}}{N} \tag{6}$$

where the $x_{max}$ and $x_{min}$ are the maximum and minimum sampled values for $X$, 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 equal to ~~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 $\boldsymbol{Q}$ or $\boldsymbol{T}$, set at twice the accuracy ($r$) of the device. The rational 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 $\boldsymbol{T}$ is listed as ±0.1 °C, and thus, the resolution ($2r$) is 0.2°C (YSI Inc., 2015). For $\boldsymbol{Q}$ all measurements below 99 $m^3$ $s^{-1}$ have an accuracy of ±0.1 $m^3$ $s^{-1}$ and thus, a resolution of 0.2 $m^3$ $s^{-1}$, while measurements above 99 $m^3$ $s^{-1}$ have an accuracy of ±1 $m^3$ $s^{-1}$, and thus a resolution of 2 $m^3$ $s^{-1}$. 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 $\delta$ 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 $\boldsymbol{X}$), the $p(x)$ is calculated by dividing the number of events in each bin by the total number of elements in $\boldsymbol{X}$. 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 $\alpha$-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 $\alpha$-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 $\alpha$-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 $\boldsymbol{X}$, a *discretised* empirical membership function of the fuzzy number $\boldsymbol{X}$ can be constructed using the calculated $\alpha$-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 & 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 $\alpha$-cut intervals (Kaufmann & Gupta, 1985). Thus, linear interpolation is used to define each fuzzy number at a pre-set $\alpha$-cut interval using the empirical $\pi(x)$ calculated using the transformation. To select the pre-set $\alpha$-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 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 $\alpha$-cut levels intervals) to conduct interpolation, leading to equal interpolated values at multiple $\alpha$-cut levels. For this research, results (discussed in the following section) of the bin-size optimisation showed that most daily observations for $\boldsymbol{T}$ and $\boldsymbol{Q}$ resulted in 2 to 10 unique $p(x)$ values. Based on this, six pre-set $\alpha$-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 & Franchini, 2011; Antanasijević et al., 2014). This has resulted in significant use of~~ ANN models have been widely used 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~~A trained MLP network can be used as a universal approximator~~s~~ with only ~~a single~~one hidden layer (Hornik et al., 1989). This means that models are relatively 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 & Valeo 2009; Napolitano et al., 2011) and thus, form a reference for the basis of comparing ANN performance (Alvisi & 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 & Valeo, 2009).

The number of neurons in the input ($n_{\mathrm{I}}$) and output ($n_{\mathrm{O}}$) layers corresponds to the number of variables used as the input and the output, respectively and the number of neurons in the hidden layer ($n_{\mathrm{H}}$) are selected based on the relative complexity of the system (Elshorbagy et al., 2010). A typical MLP is expressed mathematically as follows:

$$\boldsymbol{y}_{\mathbf{i}} = \boldsymbol{f}_{\mathbf{HID}}(\mathbf{W}_{\mathbf{IH}}\boldsymbol{x}_{\mathrm{i}} + \boldsymbol{B}_{\mathrm{H}}) \qquad (8)$$

$$\boldsymbol{z}_{\mathbf{i}} = \boldsymbol{f}_{\mathbf{OUT}}(\mathbf{W}_{\mathbf{HO}}\boldsymbol{y}_{\mathbf{i}} + \boldsymbol{B}_{\mathrm{O}}) \qquad (9)$$

where $\boldsymbol{x_i}$ is the $i^{\mathrm{th}}$ observation (an $n_{\mathrm{I}}$ x 1 vector) from of a total of $\boldsymbol{n}$ observations, $\mathbf{W_{IH}}$ is a $n_{\mathrm{H}}$ x $n_{\mathrm{I}}$ matrix of weights between the input and hidden-layer, $\boldsymbol{B}_{\mathbf{H}}$ is a vector ($n_{\mathrm{H}}$ x 1) of biases in the hidden-layer, and $y_i$ is the $i^{\mathrm{th}}$ output (an $n_{\mathrm{H}}$ x 1 vector) of the input signal through the hidden-layer transfer function, $\boldsymbol{f}_{\mathbf{HID}}$. Similarly, $\mathbf{W_{HO}}$ is an $n_{\mathrm{O}}$ x $n_{\mathrm{H}}$ matrix of weights between the hidden and output-layers, $\boldsymbol{B}_{\mathbf{O}}$ is an $n_{\mathrm{O}}$ x 1 vector of biases in the output-layer, and $\boldsymbol{f}_{\mathbf{OUT}}$ the final transfer function to generate the $i^{\mathrm{th}}$ modelled output $\boldsymbol{z_i}$ (an $n_{\mathrm{O}}$ x 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 ($E_{\mathrm{MSE}}$) (He & Valeo, 2009) – between the modelled output and the target data (i.e. observations). ~~A number of training algorithms can be used, and one of the most common methods is the~~The Levenberg–Marquardt algorithm (LMA) is one of the most common training algorithms (Alvisi et al., 2006). In ~~this method~~LMA, 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 t~~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 & Franchini, 2012; Kasiviswanathan & 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 & 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 & 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 & 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 & 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~~ Most 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 & Franchini, 2011). Thus, the advantage of fuzzy outputs (as developed by Alvisi & Franchini, 2011) is that it provides the uncertainty of the predictions ~~as well, while the fuzzy parameters reflect the uncertainty in the model structure~~in addition to the uncertainty of the parameters. 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 $\alpha$-cut interval (at a defined membership level $\mu$). This is repeated for several $\alpha$-cut levels, thus building a discretised fuzzy number at a number of membership levels. This is done by using a stepwise, constrained optimisation approach:

$$\left[\boldsymbol{y}_{\mathbf{i}}^{\mathbf{L}}\ \boldsymbol{y}_{\mathbf{i}}^{\mathbf{U}}\right] = \boldsymbol{f}_{\mathbf{HID}}\left(\left[\mathbf{W}_{\mathbf{IH}}^{\mathbf{L}}\ \mathbf{W}_{\mathbf{IH}}^{\mathbf{U}}\right]\boldsymbol{x}_{\mathbf{i}} + \left[\boldsymbol{B}_{\mathbf{H}}^{\mathbf{L}}\ \boldsymbol{B}_{\mathbf{H}}^{\mathbf{U}}\right]\right) \tag{10}$$

$$\left[\boldsymbol{z}_{\mathbf{i}}^{\mathbf{L}}\ \boldsymbol{z}_{\mathbf{i}}^{\mathbf{U}}\right] = \boldsymbol{f}_{\mathbf{OUT}}\left(\left[\mathbf{W}_{\mathbf{HO}}^{\mathbf{L}}\ \mathbf{W}_{\mathbf{HO}}^{\mathbf{U}}\right] \times \left[\boldsymbol{y}_{\mathbf{i}}^{\mathbf{L}}\ \boldsymbol{y}_{\mathbf{i}}^{\mathbf{U}}\right] + \left[\boldsymbol{B}_{\mathbf{O}}^{\mathbf{L}}\ \boldsymbol{B}_{\mathbf{O}}^{\mathbf{U}}\right]\right) \tag{11}$$

where all the variables are as described as before, and the superscripts U and L represent the upper and lower limits of the $\alpha$-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:

$$\min\left(\sum_{i=1}^{n}\left(\boldsymbol{z}_{\mathbf{i}}^{\mathbf{L}} - \boldsymbol{z}_{\mathbf{i}}^{\mathbf{U}}\right)\right)$$

$$\frac{1}{n}\sum_{i=1}^{n}(\delta_{\mathrm{i}}) \geq \boldsymbol{P}_{\mathbf{CI}} \tag{12}$$

$$\delta_i = \begin{cases} 1, & if\ \boldsymbol{z}_{\mathbf{i}}^{\mathbf{L}} < \boldsymbol{t}_i < \boldsymbol{z}_{\mathbf{i}}^{\mathbf{U}} \\ 0, & otherwise \end{cases}$$

where $t$ is that target (observed data) and $P_{CI}$ is a predefined percentage of data. Alvisi & Franchini (2011) defined $P_{CI}$~~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 $\alpha$-cut intervals of the FNN output quantify the uncertainty around the crisp prediction, within which is expected to contain $P_{\mathrm{CI}}$ 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:

$$[\boldsymbol{y}_{\mathrm{i}}^{\mathrm{L}}\ \boldsymbol{y}_{\mathrm{i}}^{\mathrm{U}}] = \boldsymbol{f}_{\mathbf{HID}}\left([\mathbf{W}_{\mathbf{IH}}^{\mathbf{L}}\ \mathbf{W}_{\mathbf{IH}}^{\mathbf{U}}] \times [\boldsymbol{x}_{\mathrm{i}}^{\mathrm{L}}\ \boldsymbol{x}_{\mathrm{i}}^{\mathrm{U}}] + [\boldsymbol{B}_{\mathrm{H}}^{\mathrm{L}}\ \boldsymbol{B}_{\mathrm{H}}^{\mathrm{U}}]\right) \tag{13}$$

$$[\boldsymbol{z}_{\mathrm{i}}^{\mathrm{L}}\ \boldsymbol{z}_{\mathrm{i}}^{\mathrm{U}}] = \boldsymbol{f}_{\mathbf{OUT}}\left([\mathbf{W}_{\mathbf{HO}}^{\mathbf{L}}\ \mathbf{W}_{\mathbf{HO}}^{\mathbf{U}}] \times [\boldsymbol{y}_{\mathrm{i}}^{\mathrm{L}}\ \boldsymbol{y}_{\mathrm{i}}^{\mathrm{U}}] + [\boldsymbol{B}_{\mathrm{O}}^{\mathrm{L}}\ \boldsymbol{B}_{\mathrm{O}}^{\mathrm{U}}]\right) \tag{14}$$

Note that now the input vector is represented by its upper and lower limits. The impact of this revision is that when there is known variance or uncertainty in the input dataset, it should be incorporated into the model structure. In Eqns. 13 and 14, this is done through the use of fuzzy rather than crisp inputs. The major impact on this is that the training algorithm for the FNN needs to accommodate this fuzzy $\alpha$-cut interval, which requires the implementation of fuzzy arithmetic principles (Kaufmann & Gupta, 1985). The cost function for the optimisation remains unchanged.

The second modification of the original algorithm is related to the selection of the percent of data included in the predicted interval ($P_{\mathrm{CI}}$). 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 $P_{\mathrm{CI}}$. Thus, a full spectrum of possible values for a given prediction is required. Thus, the Alvisi & 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 $\alpha$-cut at a $\mu = \alpha$ implies that:

$$[p(x < a) + p(x > b)] = \alpha \tag{15}$$

This can be used to calculate the probability:

$$[p(a < x < b)] = (1 - \alpha) \tag{16}$$

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~~words, the $\alpha$-cuts of a possibility distribution (at any $\mu$) 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 $P_{CI}$~~P~~ for the optimisation constraints rather than the predetermined $P_{\mathrm{CI}}$ selected by Alvisi & Franchini (2011) These are shown in Table 2.

Note that for practical purposes, $P_{\mathrm{CI}}$ 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 $\mu$ 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.

### 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 ($\boldsymbol{Q}$) and fuzzified daily water temperature ($\boldsymbol{T}$) 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 & 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 $\boldsymbol{f_{HID}}$, and a pure linear function for $\boldsymbol{f_{OUT}}$. Both function selections follow Alvisi & Franchini (2011), Wen et al., (2012) and Elshorbagy et al., (2010), and are described as follows:~~.~~

$$\boldsymbol{f}_{\mathbf{HID}} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{17}$$

$$\boldsymbol{f}_{\mathbf{OUT}} = \boldsymbol{x} \tag{18}$$

The LMA method was used to train the network, minimising $E_{\mathrm{MSE}}$. 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 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. The same process and network architecture was used to run the original FNN method (proposed by Alvisi & Franchini, 2011) using crisp inputs for comparison purposes. For this case, further refinement of the optimised solution was conducted using the built-in MATLAB minimisation function, *fmincon*. Note that for the crisp inputs, values of fuzzified daily flowrate (***Q***) and fuzzified daily water temperature (***T***) at $\mu = 1$ were used to enable direct comparison. This option allows for the closest comparison between the two approaches that have completely distinct applications. Other options for the crisp inputs (e.g. mean daily value, or maximum daily value) may also be selected for the existing FNN case.

## 2.4 Risk analysis using defuzzification

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 Steenbergen 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 Steenb~~g~~ergen et al., 2012) have a probability-based meaning – in this case “most likely” is a measure of *likelihood*.

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 & Valeo, 2014a, 2015b): for any

$x$ in the support (defined as the $\alpha$-cut interval at $\mu$ = 0) of a fuzzy number [$a$ $b$] we have the corresponding $\mu$ and the paired value $x$' which shares the same membership level. The value $\mu$ is the sum of the cumulative probability between [$a$, $x$] and [$x$', $b$], labelled $P_L$ and $P_R$, respectively:

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

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:

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

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 mg $L^{-1}$, 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 mg $L^{-1}$.

# 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 $\boldsymbol{Q}$ and $\boldsymbol{T}$ data for the ~~nine year~~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 $\boldsymbol{Q}$ and $\boldsymbol{T}$ as a means of illustrating the advantages of using the proposed approach.

Fig. 3 shows sample results of converting hourly $\boldsymbol{Q}$ 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 hours. The resulting histogram-based probability functions are shown for both the optimised

($D_{opt}$, illustrated with circles) and original ($D_{orig}$=2$r$; see Sect. 2.2.1 for the definition, illustrated with squares) bin-sizes in the second column. The third and fourth columns in 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 ($D_{orig}$ = $D_{opt}$ = 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 $D_{orig}$ = 2$r$, and since this is not physically feasible (measurable) the algorithm did not consider any bin-sizes below 2$r$. 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 $m^3$ $s^{-1}$). Thus, the $\alpha$-level cut at this level is a simple singleton: $[161]_{\mu=1}$. The next membership level was calculated as 0.58; again the resulting $\alpha$-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 $\alpha$-cuts at higher membership levels define the upper and lower limits of α-cuts at lower levels. Thus, using the information from the $\alpha$-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 $\alpha$-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 $\alpha$-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 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 $\alpha$-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 are 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 ($D_{\mathrm{opt}}$ = 0.8 vs. $D_{\mathrm{orig}}$ = 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 $\mathrm{m}^3\ \mathrm{s}^{-1}$. 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 $D_{\mathrm{opt}}$ 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 $D_{\mathrm{orig}}$. Thus, in comparing columns 3 and 4 for the second row, the smoother membership function representing $D_{\mathrm{opt}}$ 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, ($D_{\mathrm{opt}}$ = 9 vs. $D_{\mathrm{orig}}$ = 2). Again, the original histogram is extremely uneven, whereas the post-optimisation histogram is considerably smoother with a definitive modal value at $Q$ = 277 $\mathrm{m}^3\ \mathrm{s}^{-1}$. 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 m$^3$ 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 membership 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.

Fig. 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 interpolated fuzzy number will all have equal intervals (due to the trapezoidal shape), transferring no useful information to the final fuzzy number.

Overall, the above examples illustrate the advantages of using the couple method of bin-size optimisation and probability-possibility transformation to create fuzzy numbers for the FNN application. The applicability of this method is not necessarily restricted to this application and can be applied whenever there is a need to construct fuzzy numbers from observed data. The utility of the first component, bin-size optimisation to estimate the density function, is that in cases where either not enough information is available to define a probability distribution, or if the data do not follow the mould of a known density function, or if assumptions on the class of distribution cannot be made, the optimum bin-size can be calculated to define an empirical distribution for the probability-possibility transformation. The advantage of the second component, the algorithm to construct the possibility distribution (i.e. the membership function of the fuzzy number) is that it provides a consistent, transparent and objective method to convert observations (e.g. time series data) into fuzzy numbers - which has been cited as a major hurdle in implementing fuzzy number based applications in the literature (Abrahart et al., 2010; Dubois & Prade, 1993; Civanlar & Trussel, 1986). A noteworthy component of this algorithm is that the fuzzy numbers do not reduce to the simple, triangular shaped functions that are widely used, but rather the functions better represent the information from the observations.

### 3.2 Training the fuzzy neural network

Once the observations of the abiotic input parameters ($\boldsymbol{Q}$ and $\boldsymbol{T}$) were converted to fuzzy numbers, the FNN training algorithm was run using five neurons in the hidden layer, to predict daily minimum DO in the Bow River. First, the values of the fuzzy numbers at $\mu = 1$ was used to train the crisp network. This was done to have initial estimates of the 10 $\mathbf{W}_{\mathbf{IH}}$ (5 for each input), 5 $\boldsymbol{B}_{\mathbf{IH}}$, 5 $\mathbf{W}_{\mathbf{HO}}$, and 1 $\boldsymbol{B}_{\mathbf{O}}$. These initial estimates were used to provide the upper and lower limits of the constraints for the proceeding optimisation algorithm. Once these estimates were calculated, the optimisation algorithm was used to calculate the fuzzy weights and biases using fuzzy inputs, and was started from $\mu = 0$ and moving sequentially to higher membership levels until $\mu = 0.8$. The final level (at $\mu = 1$) was calculated using the midpoint of the intervals estimated at $\mu = 0.8$. The total optimisation time ~~(using the SCE-UA algorithm)~~ for the ~~selected architecture took~~proposed method was 13 hours, whereas the existing method with crisp inputs was 8 hours, using a 2.40 GHz Intel® *Xeon* microprocessor (with 4 GB RAM).

The $E_{\mathrm{MSE}}$ and the Nash-Sutcliffe model efficiency coefficient ($E_{\mathrm{NSE}}$; Nash & Sutcliffe, 1970) for the training, validation and testing scenarios for $\mu = 1$ for both methods are shown in Table 3. The $E_{\mathrm{MSE}}$ for each dataset are low, between 11% and 16% of the mean annual minimum DO seen in the Bow River for the study period. The $E_{\mathrm{NSE}}$ values are approximately equal to 0.5 for each subset, which is higher than $E_{\mathrm{NSE}}$ values in the literature for water quality parameters when modelled daily (see Moriasi et al., 2007 for a survey of results) and is considered to be "satisfactory" by their standards. In comparing the two methods, it is obvious that including additional information (in the form of fuzzy inputs) does not decrease performance, as the metrics are nearly identical for both methods. This shows that the proposed method has successfully incorporated input data uncertainty in the model architecture. These ~~two~~ model performance metrics highlight that in general, predicting minimum DO using abiotic inputs and a data-driven approach is an effective technique.

The results of the optimisation component of the algorithm are summarised in Table 4, which shows the percentage of data ($P_{CI}$) captured within the resulting $\alpha$-cut intervals for each of the three data subsets. The performance for each of the datasets (i.e., train, validation and test) for both methods is nearly identical (on an interval-by-interval basis): the exact amount of data captured within the intervals, as required by the constraints, except for the $\mu$ = 0.8 interval. At this interval, the~~-~~ amount of coverage decreases (i.e. lower performance) as the membership level increases, which is unavoidable when the width of the uncertainty bands decrease. ~~However, as~~ As required by Eq. 12, the amount of data within the interval has to be greater-than or equal-to the limit defined by $P_{CI}$ (as per Alvisi & Franchini, 2011) which is true for all training data. This means that a solution to satisfy the constraints with a lower amount of data (e.g. reducing the 29.91% for the $\mu$ = 0.8 interval for the proposed method) would either result in non-minimal intervals (though this is unlikely) or that the constraints on the values of the intervals could not be maintained. ~~(t~~This latter issue will be discussed in detail with Fig. 5 below~~)~~. ~~For the validation and testing datasets, similar performance is seen for both with near perfect $P_{CI}$ captured at $\mu$ = 0.6, 0.8 and 1, and more than $P_{CI}$ at $\mu$ = 0.2 and 0.4.~~ Lastly, as mentioned above, the performance of non-training datasets for both methods decrease as the interval get narrow: this can be seen best by the inability for both methods to capture the exact amount of data required at the $\mu$ = 0.8 interval for the validation and testing datasets. These results are similar to the testing dataset in Alvisi & Franchini (2011). This comparison again demonstrates the ability of the proposed method with fuzzy inputs to function in similar manner to the original algorithm that used crisp inputs.

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 & Franchini, 2011; Khan et al. 2013; Khan & 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 shapes of the fuzzy weights and biases for the proposed and existing method are generally the same for the input-hidden layer, however differences can be seen for the hidden-output layer plots.

Since the existing method uses crisp inputs, it requires the produced weights and biases to represent the uncertainty in the data, to produce output intervals wide enough to capture the set amount of observations. This is reflected in hidden-output layer plots where the lower limit of the membership function for Weight #5 is highly skewed, which enables this method to capture the low DO events. Similarly, Bias #1 in the hidden-output layer has been translated to a lower value, to produce fuzzy DO outputs that capture the low DO observations.

The figure demonstrates that enough $\alpha$-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 $\mathbf{W}_{\mathbf{IH}}$ number 5 and $\boldsymbol{B}_{\mathbf{O}}$ (for the proposed method)~~,~~ 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.

Connecting this back to the results in Table 4, these two particular weights and biases show why the percentage of data calculated at $\mu$ = 0.8~~0.2~~ (for training) cannot be improved by further optimisation. At some point, if the intervals at $\mu$ = 0.8~~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.

Table 3 and 4, and Fig. 5 demonstrates the overall success of the proposed approached to calibrate an FNN model as compared to a crisp ANN, as well as an FNN that uses crisp inputs. 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 versus 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, for the proposed method, with similar results for the crisp input FNN method) 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 mg $L^{-1}$ 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 $\alpha$-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. Thus, there is a need to develop an equivalent performance metric when comparing crisp observations to fuzzy number predictions.

Fig. 6 also demonstrates the benefit of the FNN approach as compared to the crisp ANN approach with respect to predicting low DO (i.e. when DO is less than 6.5 mg $L^{-1}$). Both the~~The~~ FNN methods predict~~s~~ 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. In comparing the two FNN methods, both methods give similar results: the average width of predicted low DO intervals for the nine-year period (at $\mu = 0$) is 8.84 mg $L^{-1}$ for the proposed method, and 8.60 mg $L^{-1}$ for the existing method. The impact of the width of the predicted intervals is discussed later.

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 results are shown only for the proposed method for clarity; difference between the existing method (using crisp inputs) and the proposed method (using fuzzy inputs) is discussed later. 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 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 $\alpha$-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 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 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 mg $L^{-1}$) occur from mid-July and continue occasionally until mid-September. Details of some of these low DO events are plotted in Fig. 9. Fig. 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 mg $L^{-1}$) events are captured at a membership level greater than $0.2^{L}$. Meaning that the low DO predictions in 2007 for the 6.5 mg $L^{-1}$ 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 $\alpha$-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 mg $L^{-1}$ in 2010. Also of note for this year is that *all* observations are less than 10 mg $L^{-1}$, and about 90% of all observations are below the 9.5 mg $L^{-1}$ 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 where DO falls from a high of 10 mg $L^{-1}$ to a low of 7 mg $L^{-1}$, 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.

Fig. 9 shows details of a low DO (<9.5 mg $L^{-1}$) 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 mg $L^{-1}$) 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 mg $L^{-1}$) while under predicting the DO < 9.5 mg $L^{-1}$ 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 $\boldsymbol{Q}$ and $\boldsymbol{T}$ 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.

Figs. 10 shows a comparison of the predicted minimum DO trends from both the proposed FNN method (solid black line) and the existing FNN method (dashed black line), along with the observed data (circles) for each membership level for the 2009 data. These figures show that despite the use of more data for the inputs (i.e. fuzzy numbers versus crisp numbers), both methods are optimised to show similar results (due to the optimisation algorithm requiring a specific amount of data being captured at each level). This shows that the optimisation algorithm developed in this manuscript for fuzzy inputs successfully mimics the original algorithm developed by Alvisi & Franchini (2011) that only used crisp inputs. Thus, when modelling a complex system, such as the minimum daily DO in the Bow River, the uncertainty in the inputs can also be quantified and propagated through the data-driven model, by using the proposed method. This is a major advantage over the original model (Alvisi & Franchini, 2011) that only allowed crisp inputs to be used. Note that as per Table 4, both methods approximately capture the same amount of data at each interval, however as Fig. 10 indicates this does not necessarily mean that the predicted intervals are exactly the same for both methods. Both methods predict unique intervals, with the overall result being that $P_{CI}$ amount of data is captured within each interval. Note that at the interval $\mu = 0$, the existing method using crisp inputs by design predicts this interval as a singleton (thus the interval width will always be zero), whereas the proposed method has an additional feature of predicting an interval for the $\mu = 1$ level. An instance of this can be seen near the end of September where an interval is predicted rather than a singleton.

Fig 11 compares with width of predicted intervals at four selected membership levels for both methods, generally showing mixed results. As discussed above the existing method does not predict an interval for $\mu = 1$, thus, a comparison cannot be made and is not included. At the $\mu = 0.8$ level, the average width of the intervals for the existing method is close to 0, whereas for the proposed method is 0.36 mg $L^{-1}$. This is consistent with the results shown in Table 4 and Fig. 5, that demonstrates the narrow interval at higher membership levels. Annual comparisons of the remaining intervals show that the intervals at $\mu = 0$ are larger for the proposed method compared to the existing method for all but one year (2009). However, at $\mu = 0.2$, 0.4 and 0.6, the width of the intervals is smaller for the proposed method for a majority of years, however the overall differences are not statistically significant ($p > 0.05$ using the two-sample Kolmogorov-Smirnov test). The results demonstrate that while both method can achieve the optimisation objectives whilst respecting the constraints (i.e. $P_{CI}$), it is reflected differently in the predicted fuzzy intervals. Generally, the predicted fuzzy numbers using the proposed method have a larger support (at $\mu = 0$) signifying that the increased uncertainty due to fuzzy inputs into the model are propagated through the model. Whereas, the lack of inclusion of uncertain input data in the existing method results in a slightly narrower average support. In essence, the proposed FNN model is modelling a more complex system (because of the inclusion of input uncertainty) whereas the existing method models the system by assuming lower complexity (by ignoring the input uncertainty).

### 3.3 Risk analysis for low DO events

The utility of the FNN method is illustrated through an analysis of the ability of the proposed 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 cable to capture 100% of all low DO events (i.e. below 5, or 6.5 or 9.5 mg $L^{-1}$) within the predicted intervals. In comparison, the crisp ANN network (i.e. at $\mu = 1$) did not predict DO to be less than 5 mg $L^{-1}$ on any of the 51 occasions. Similarly, it predicted DO to be less than the more conservative limit of 6.5 mg $L^{-1}$ in only 53% of the 184 occurrences. For the last case, the 9.5 mg $L^{-1}$ 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 (DO < 5 mg $L^{-1}$). 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 mg $L^{-1}$), 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~~12). Once these were identified, the defuzzification technique was used to predict the probability of low DO (e.g. $P(\{DO<5 \text{ mg L}^{-1}\})$).

For the first case, $P(\{DO<5 \text{ mg L}^{-1}\})$, 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 mg $L^{-1}$, the FNN results identified the possibility of low DO and the probability of DO to be below the 5 mg $L^{-1}$ 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. 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 mg $L^{-1}$, 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 mg $L^{-1}$ (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 were minimum daily DO is observed to be below the 9.5 mg $L^{-1}$ limit.

It is worth noting here that the proposed FNN model was designed to only include data from the April to October each year, corresponding to the ice-free period (as defined in Sect. 2.1). This implies that the analysis has been conducted on the time period that is most critical or susceptible to low DO. Thus, as the proposed FNN model predicts, there is possibility of low DO on most days (as shown in the trend plots in Figs. 7 and 8). However, the consistency principle (Zadeh, 1978) implies that an event must be possible before it is probable. Thus, a possibility to predict low DO does not imply that it will occur with a high probability. In fact, nearly all the possibility of low DO events occurs at low membership levels (i.e. $\mu < 0.2$) implying a low possibility – and the skewed nature of the results deem the probability to be low as well. For example, for the DO < 5 mg $L^{-1}$ case, the proposed FNN model predicted 1367 days where low DO was predicted but not observed, however, on average the probability of low DO for 98% of these events was much lower than the threshold criteria (14%) mentioned above. Thus, the number of "false alarms" predicted by the proposed method is very low. Similarly, for the DO < 6.5 mg $L^{-1}$ and 9.5 mg $L^{-1}$ cases, each had ~94% of the low DO cases to fall below their respective threshold criteria. This shows that while FNN model correctly predicts a possibility of low DO for the majority of the days (corresponding to the typical low DO conditions), the risk of predicting a "false alarm" is low. Lastly, it should be noted that the wide intervals predicted at $\mu = 0$ are a function of the rapidly decreasing low DO value seen in 2004 (discussed in Sect. 3.2) that are likely due to instrument error. With the inclusion of new data as it becomes available, and as the model parameters are updated, it is expected that these outliers will be part of the 0.5% of $P_{CI}$ not included in the predicted intervals, resulting in narrower bands of predictions.

The predicted membership functions of minimum DO for nine examples are shown Fig. ~~10~~12, 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 mg $L^{-1}$; 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 membership functions are not triangular shaped as assumed in many fuzzy number based applications. This is of significance because it shows that the amount of uncertainty (or credibility) of the model output does not change linearly with the magnitude of DO, which has important implications regarding the risk of low DO.

For the 5 mg $L^{-1}$ guideline, the intersection of the membership function and the guideline occurs at low possibility levels (between $\mu = 0^L$ and $0.2^L$), meaning that the corresponding probability will be low as well, as illustrated by the probability values shown in the figure. This again highlights that the risk of low DO (< 5 mg $L^{-1}$) is predicted to be low by the FNN mostly due to the fact that the observations are captured at low membership levels. Note that the crisp ANN results (at $\mu = 1$) always over predict low DO, as shown in these three examples. The observed value falls within the predicted interval for each case, also at low membership levels.

The examples for the 6.5 mg $L^{-1}$ guideline (second row in Fig. ~~10~~12) show that the intersection between the membership function and the guidelines occurs between $\mu = 0.4$ and 0.6 on 26 July 2006, between $\mu = 0.6$ and 0.8 on 8 August 2007, and at about $\mu = 0.6$ on 29 September 2004. This illustrates the broader trend with the 6.5 mg $L^{-1}$ guideline (which was discussed earlier and had a large range of risk predictions), which is that for the full dataset, the possibility of low DO (< 6.5 mg $L^{-1}$) occurs at every interval with the majority occurring at higher intervals. This is in contrast to the 5 mg $L^{-1}$ guideline where the possibility of low DO only occurs only between $\mu = 0$ and 0.2.

The last row in Fig. ~~10~~ 12 show sample low DO results for the 9.5 mg $L^{-1}$ guideline. As discussed above, more than 80% of these events had a high (more than 90%) risk of low DO. In the first example, on 23 September 2004, the guideline intersects the membership function at $\mu = {\sim}0.2^R$, corresponding to a ~97% risk of low DO. The 6 August 2008 has a low DO prediction of 100% – this is because the predicted fuzzy number is entirely below the guideline limit. A similar result can be seen for the last example. These examples also illustrate that had only a triangular membership function been used (i.e. the fuzzy numbers defined at two membership levels), the probability of low DO could not be quantified as specifically as it has been here. The slight changes in membership function shapes between intervals impact the final probability, and a linear function would have not captured these changes.

These examples ~~are meant to~~ illustrate the potential utility of the data-driven and abiotic input parameter DO model, that can be used to assess the risk of low DO. Given that it is a data-driven approach, the model can be continually updated as more data is available, further refining the predictions. Various combinations of input values can be used to predict fuzzy minimum DO and defuzzification technique can be used to *quantify* the risk of low DO given the input values. The utility of this method is that a water-resource manager can use forecasted water temperature data and expected flow rates to *quantify the risk* of low DO events in the Bow River, and can plan accordingly. For example, if the risk of low DO reaches a specific numerical threshold or trigger, different actions or strategies (e.g. increasing flow rate in the river by controlled release from the upstream dams) can be implemented. The quantification of the risk to specific probabilities means that the severity of the response can be tuned to the severity of the calculated risk.

# 4 Conclusions

A new method to predict DO concentration in an urbanised watershed is proposed. Given the lack of understanding of the physical system that governs DO concentration in the Bow River (in Calgary, Canada), a data-driven approach using fuzzy numbers is proposed to account for the uncertainty. Further, the model uses abiotic (non-living, physical and chemical attributes) factors as inputs to the model. Specifically, water temperature and flow rate were selected which are routinely monitored and thus, a large dataset is available.

The data-driven approach proposed is a modification of an existing fuzzy neural network method that quantifies the total uncertainty in the model by using fuzzy number weights and biases. The proposed model refines the exiting model by (i) using possibility theory based intervals to calibrate the neural network (rather than arbitrarily 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.

## Author contributions

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

## Data availability

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

## Acknowledgements

The authors would like to thank Dr S. Alvisi from the Università degli Studi di Ferrara for providing the MATLAB code for the original Fuzzy Neural Network model. The authors would also 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, t~~The authors are grateful for the City of Calgary and Environment Canada for providing the data used in this research. Lastly, the authors would like to acknowledge the comments of two anonymous reviewers, whose feedback greatly helped improve this manuscript.

## References


Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Asaad Y. Shamseldin, A. Y., Solomatine, D. P., Toth, E., & Wilby, R. L.: Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting, Prog. Phys. Geog., 36, 480-513, doi:10.1177/0309133312444943, 2012.

Adams, K. A., Barth, J. A., & 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.

Alberta Environment (AENV): 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. & Franchini, M.: Fuzzy neural networks for water level and discharge forecasting with uncertainty, Environ. Modell. Softw., 26, 523–537, doi:10.1016/j.envsoft.2010.10.016, 2011.

Alvisi, S., & 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., & 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., & 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., & 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., & Duckstein, L.: Fuzzy regression in hydrology, Water Resour. Res., 26, 1497-1508, doi:10.1029/WR026i007p01497, 1990.

Bárdossy, A., Mascellani, G., & Franchini, M.: Fuzzy unit hydrograph, Water Resour. Res., 42, W02401, doi: 10.1029/2004WR003751, 2006.

Betrie, G. D., Sadiq, R., Morin, K. A., & 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.

Bow River Basin Council - Profile of the Bow River Basin: http://wsow.brbc.ab.ca/index.php?option=com_content&view=article&id=259&Itemid=83 , last access: 3 September 2015.

Canadian Council of Ministers of the Environment (CCME): Canadian water quality guidelines for the protection of aquatic life: Dissolved oxygen (freshwater), Canadian Council of Ministers of the Environment, Winnipeg, MB, Canada, 1999.

Chang, F. J., Tsai, Y. H., Chen, P. A., Coynel, A., & Vachaud, G.: Modeling water quality in an urban river using hydrological factors–Data driven approaches, J. of Environ. Manage., 151, 87-96, doi:10.1016/j.jenvman.2014.12.014, 2015.

Civanlar, M. R., & Trussell, H. J.: Constructing membership functions using statistical data, Fuzzy Sets Syst., 18(1), 1-13, doi:10.1016/0165-0114(86)90024-2, 1986.

Deng, Y., Sadiq, R., Jiang, W., & Tesfamariam, S.: Risk analysis in a linguistic environment: a fuzzy evidential reasoning-based approach, Expert Syst. Appl., 38(12), 15438-15446, doi:10.1016/j.eswa.2011.06.018, 2011.

Dorfman, R., & Jacoby, H. D.: An illustrative model of a river basin pollution control, in: Models for Managing Regional Water quality, Dorfman, R., Jacoby, H. D., & Thomas Jr., H. A. (Eds.), Harvard University Press, Cambridge, MA, USA, 84-141, doi:10.4159/harvard.9780674419216.c3, 1972.

Duan, Q., Sorooshian, S., & Gupta, V.: Effective and efficient global optimisation for conceptual rainfall-runoff models, Water Resour. Res, 28(4), 1015-1031, doi:10.1029/91WR02985, 1992.

Dubois, D., & Prade, H.: Fuzzy sets and probability: misunderstandings, bridges and gaps, in: Proceedings of the Second IEEE International Conference on Fuzzy Systems, San Francisco, USA, 28 March – 1 April 1993, 1059-1068, doi:10.1109/FUZZY.1993.327367, 1993.

Dubois, D., & Prade, H.: Possibility Theory and its Applications: Where do we stand? in: Springer Handbook of Computational Intelligence, Kacprzyk, J., & Pedrycz, W. (Eds.), Springer-Verlag Berlin Heidelberg, 31-60, doi:10.1007/978-3-662-43505-2_3, 2015.

Dubois, D., Prade, H., & Sandri, S.: On possibility/probability transformations, in: Fuzzy logic, Lowen, R., & Roubens, M. (Eds.), Springer Netherlands, 103-112, doi:10.1007/978-94-011-2014-2_10, 1993.

Dubois, D., Foulloy, L., Mauris, G., & 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.

Elshorbagy, A., Corzo, G., Srinivasulu, S., & 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.

Environment Canada: Wateroffice Hydrometric Station Meta Data. https://wateroffice.ec.gc.ca/station_metadata/stationList_e.html , last access 3 September 2015.

Ferrero, A., Prioli, M., Salicone, S., & Vantaggi, B.: A 2-D Metrology-Sound Probability–Possibility Transformation, IEEE T. Instrum. Meas., 62, 982-990, doi:10.1109/TIM.2013.2246910, 2013.

Guyonnet, D., Bourgine, B., Dubois, D., Fargier, H., Come, B. & 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.

Hauer, F. R., & Hill, W. R.: Temperature, light and oxygen, in: Methods in stream ecology, Hauer, F. R. & Lamberti, G. A. (Eds.), Academic Press, San Diego, CA, USA, 103–117, doi:10.1016/B978-012332908-0.50007-3, 2007.

He, J., & 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.

He, J., Chu, A., Ryan, M. C., Valeo, C. & Zaitlin, B.: Abiotic influences on dissolved oxygen in a riverine environment. Ecological Engineering, 37, 1804–1814, doi:10.1016/j.ecoleng.2011.06.022, 2011.

He, J., Ryan, M. C., & Valeo, C.: Changes in Water Quality Characteristics and Pollutant Sources along a Major River Basin in Canada, in: Environmental Management of River Basin Ecosystems, Ramkumar, M., Kumaraswamy, K., & Mohanraj, R. (Eds.), Springer International Publishing, 525-548, doi:10.1007/978-3-319-13, 425-3_25, 2015.

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., & White, H.: Multilayer feedforward networks are universal approximators, Neural Networks, 2, 359-366, doi:10.1016/0893-6080(89)90020-8, 1989.

Huang, Y., Chen, X., Li, Y. P., Huang, G. H., & 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., & 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.

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., & 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.

Kasiviswanathan, K. S., & 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.

Kasiviswanathan, K. S., Cibin, R., Sudheer, K. P., & 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. & Gupta, M. M.: Introduction to Fuzzy Arithmetic: Theory and Applications. Van Nostrand Reinhold Company, New York, NY, USA, 1985.

Khan, U. T., Valeo, C., & 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.

Khan, U. T., & 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, St John's, Canada, 6-8 August 2014, 2014a.

Khan, U. T., & 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, St Johns, NFLD, Canada, 6-8 August 2014, 2014b.

Khan, U. T., & 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.

Khan, U. T., & Valeo, C.: Short-term peak flow rate prediction and flood risk assessment using fuzzy linear regression, J. Environ. Inform. (in press), 2015b.

Klir, G. J., & 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, 1999.

Maier, H. R., Jain, A., Dandy, G. C., & 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. Modell. Softw., 25, 891-909, doi:10.1016/j.envsoft.2010.02.003, 2010.

Martin-Clouaire, R., Cazemier, D. R., & 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.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & 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.

Mujumdar, P. P. & 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., & See, L.: Impact of EMD decomposition and random initialisation of weights in ANN hindcasting of daily stream flow series: an empirical examination, J. Hydrol., 406(3-4), 199-214, doi:10.1016/j.jhydrol.2011.06.015, 2011.

Nash, J. E., & Sutcliffe, J. V.: River flow forecasting through conceptual models: Part I. A discussion of principles, J. Hydrol., 10(3), 282-290. doi:10.1016/0022-1694(70)90255-6, 1970.

Neupane, A., Wu, P., Ghanbarpour, R. & 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.

Pogue, T. R., & 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, U. S. Geological Survey, Portland, OR, USA, 1995.

Robinson, K. L., Valeo, C., Ryan, M. C., Chu, A., & 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.

Sadiq, R., Rodriguez, M. J., Imran, S. A., & 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.

Serrurier, M., & 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., & 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.

Shrestha, D. L., & Solomatine, D. P.: Data-driven approaches for estimating uncertainty in rainfall-runoff modelling, JRBM, 6, 109–122, doi:10.1080/15715124.2008.9635341, 2008.

Singh, K. P., Basant, A., Malik, A., & 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.

Solomantine, D. P., & 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., & Abrahart, R. J.: Data–driven modelling: concepts, approaches and experiences, in: Practical Hydroinformatics, Abrahart, R. J., See, L. M., & Solomatine, D. P. (Eds.), Springer Berlin Heidelberg, Berlin, Germany, 17–30, doi:10.1007/978–3–540–79881–1_2, 2008.

Van Steenbergen, N., Ronsyn, J., & Willems, P.: A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication, Environ. Modell. Softw., 33, 92-105, doi:10.1016/j.envsoft.2012.01.013, 2012.

Verhoest, N. E. C., De Baets, B., Mattia, F., Satalino, G., Lucau, C., & 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., & 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.

YSI Inc. - YSI 5200A Multiparameter Monitor & Control Specifications: https://www.ysi.com/File%20Library/Documents/Specification%20Sheets/W45-01-5200A.pdf, last accessed September 3, 2015.

Zadeh, L. A.: Fuzzy sets. Inform. Control, 8, 338–353, doi:10.1016/S0019–9958(65)90241–X, 1965.

Zadeh, L. A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets 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. Ph. D. thesis, Civil Engineering, University of Calgary, Canada, 290 pp., 2009.

Zhang, K., Li, H., & 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.

Zhang, K. & Achari, G.: Correlations between uncertainty theories and their applications in uncertainty propagation, in: Safety, reliability and risk of structures, infrastructures and engineering systems, Furuta, H., Frangopol, D.M., & Shinozuka, M. (Eds.), Taylor & Francis Group, London, UK, 1337–1344, doi:10.1201/9781439847657–c20, 2010.

## Table Captions

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

Table 2: Selected values for PCI for the FNN optimisation

Table 3: The $E_{MSE}$ and $E_{NSE}$ for each subset of the fuzzy neural network using the method proposed (using fuzzy inputs) and using the original method (using crisp inputs)

Table 4: Percentage of data captured within each α-cut interval for the three subsets of data

# Tables

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 mg $L^{-1}$ [a] | DO < 6.5 mg $L^{-1}$ [b] | DO < 9.5 mg $L^{-1}$ [c] | Total number of samples |
|---|---|---|---|---|
| 2004 | 25 | 41 | 107 | 135 |
| 2005 | 1 | 26 | 133 | 208 |
| 2006 | 25 | 70 | 164 | 209 |
| 2007 | 0 | 27 | 182 | 211 |
| 2008 | 0 | 5 | 130 | 163 |
| 2009 | 0 | 15 | 85 | 96 |
| 2010 | 0 | 0 | 180 | 207 |
| 2011 | 0 | 0 | 122 | 204 |
| 2012 | 0 | 0 | 76 | 206 |
| Total | 51 | 184 | 1179 | 1639 |

[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)

Table 2: Selected values for $P_{CI}$ for the FNN optimisation

| $\mu$ | $P_{CI}$ (%) |
|---|---|
| 1.00 | 0.00 |
| 0.80 | 20.00 |
| 0.60 | 40.00 |
| 0.40 | 60.00 |
| 0.20 | 80.00 |
| 0.00 | 99.50 |

Table 3: The $E_{MSE}$ and $E_{NSE}$ for each subset of the fuzzy neural network using the method proposed (using fuzzy inputs) and using the original method (using crisp inputs)

| | $E_{MSE}$ (mg $L^{-1}$)$^2$ | | $E_{NSE}$ | |
|---|---|---|---|---|
| | Proposed | Original | Proposed | Original |
| Train | 1.52 | 1.55 | 0.52 | 0.51 |
| Validation | 1.19 | 1.18 | 0.49 | 0.49 |
| Test | 1.09 | 1.10 | 0.54 | 0.54 |

Table 4: Percentage of data captured within each $\alpha$-cut interval for the three subsets of data

| | Percent captured, $P_{CI}$ (%) | | | | | |
|---|---|---|---|---|---|---|
| | Proposed method | | | Existing method | | |
| μ | Train | Validation | Test | Train | Validation | Test |
| 1.00 | - | - | - | - | - | - |
| 0.80 | 29.91 | 28.54 | 28.78 | 20.02 | 14.39 | 18.05 |
| 0.60 | 39.93 | 40.98 | 40.24 | 40.05 | 35.85 | 40.49 |
| 0.40 | 59.95 | 66.10 | 64.15 | 60.07 | 60.73 | 61.95 |
| 0.20 | 79.98 | 80.49 | 82.93 | 80.10 | 79.51 | 82.44 |
| 0.00 | 99.39 | 98.78 | 99.02 | 99.51 | 98.54 | 99.02 |

## Figures Captions

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

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. WIH are the weights between the input and hidden layer, WHO are the weights between the hidden and output layer, BH are the biases in the hidden layer, and BO is the bias in the output layer.

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

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

Figure 5: Sample plots of the produced membership functions for the weights and biases of the fuzzy neural network for both the proposed and existing methods

Figure 6: A comparison of the predicted and observed minimum DO at the $\mu = 0$ interval (~~black~~ grey line) and at $\mu = 1$ (black dots) for the proposed (top row) and existing (bottom row) methods

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

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

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

Figure 10: Comparison of predicted trends of the proposed (solid black line) and existing (dashed black line) methods shown for 2009 for each membership level. Observations are shown as black circles

Figure 11: A comparison of average annual interval widths of predicted fuzzy numbers using the proposed and existing FNN methods for four selected membership levels

Figure ~~10~~12: Sample plots of low DO events and the corresponding risk of low DO calculated using a possibility-probability transformation for the (top) 5 mg $L^{-1}$, (middle) 6.5 mg $L^{-1}$, and (bottom) 9.5 mg $L^{-1}$ guideline

## Figures

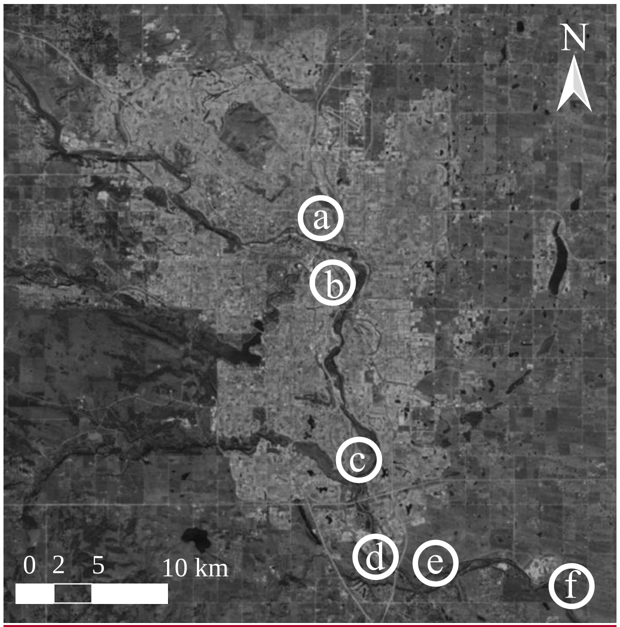



Figure 1: An aerial view of the City of Calgary, Canada showing the locations of (a) the flow monitoring site Bow River at Calgary (Water Survey of Canada ID: 05BH004), three wastewater treatment plants at (b) Bonnybrook, (c) Fish Creek, and (d) Pine Creek, and two water quality sampling sites (e) Stier’s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

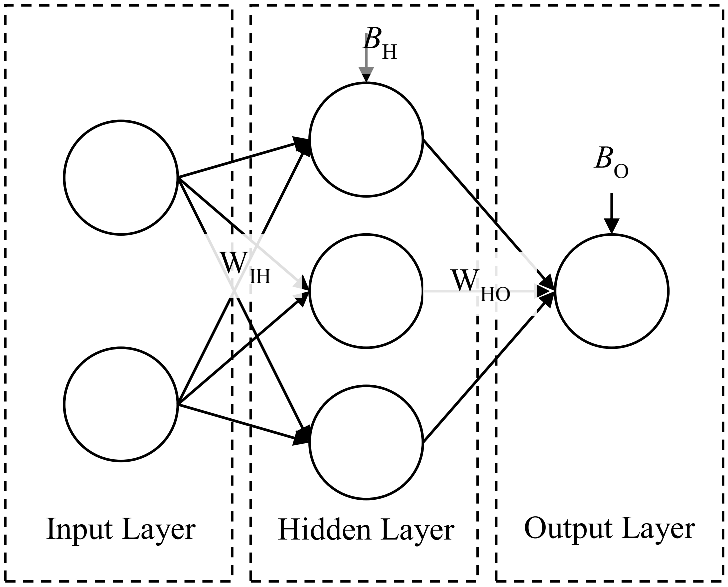
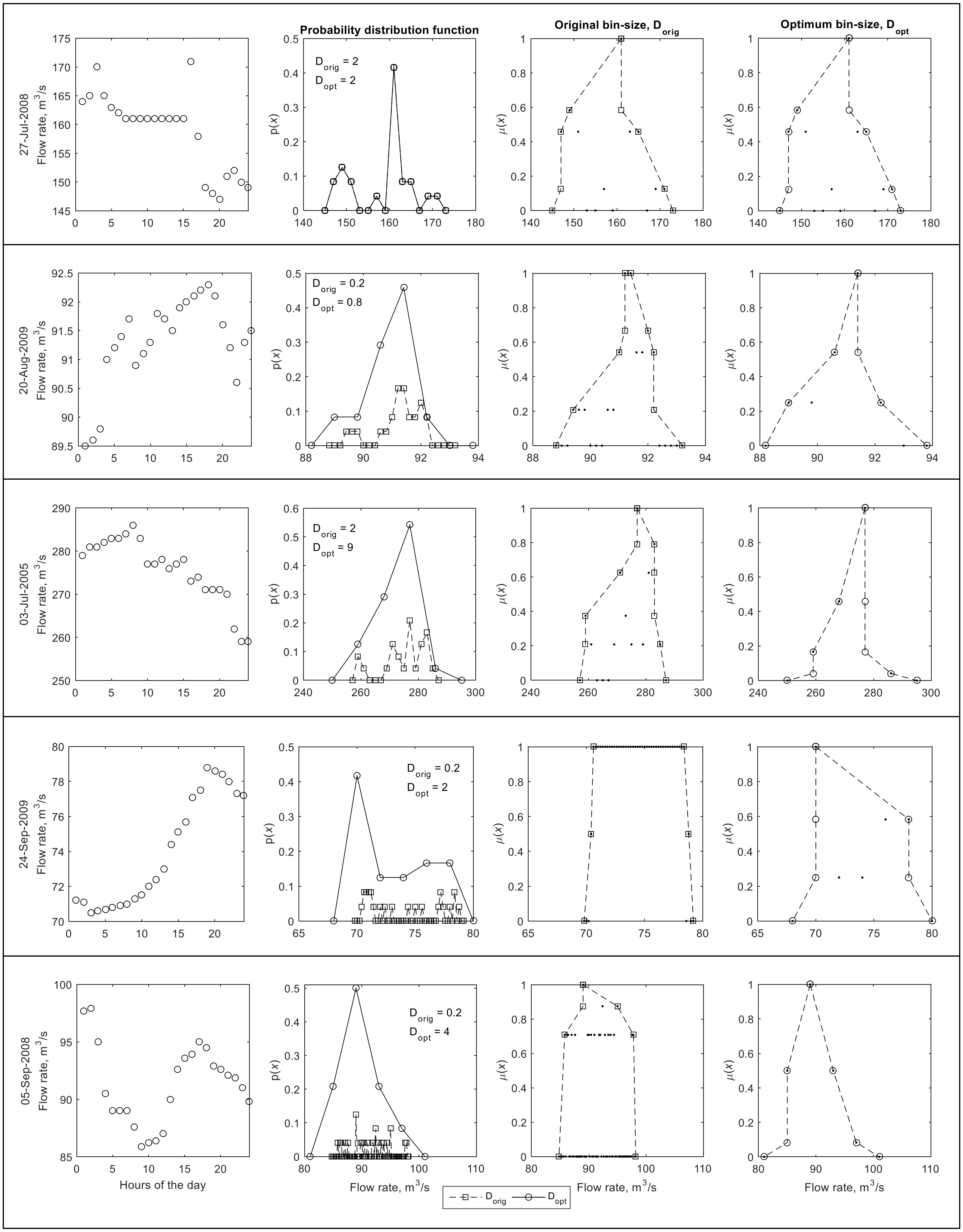



Figure 2: An example of a three-layer multilayer perceptron feed-forward ANN, with two input neurons, the hidden layer neurons, and one output neuron. WIH are the weights between the input and hidden layer, WHO are the weights between the hidden and output layer, BH are the biases in the hidden layer, and BO is the bias in the output layer.



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

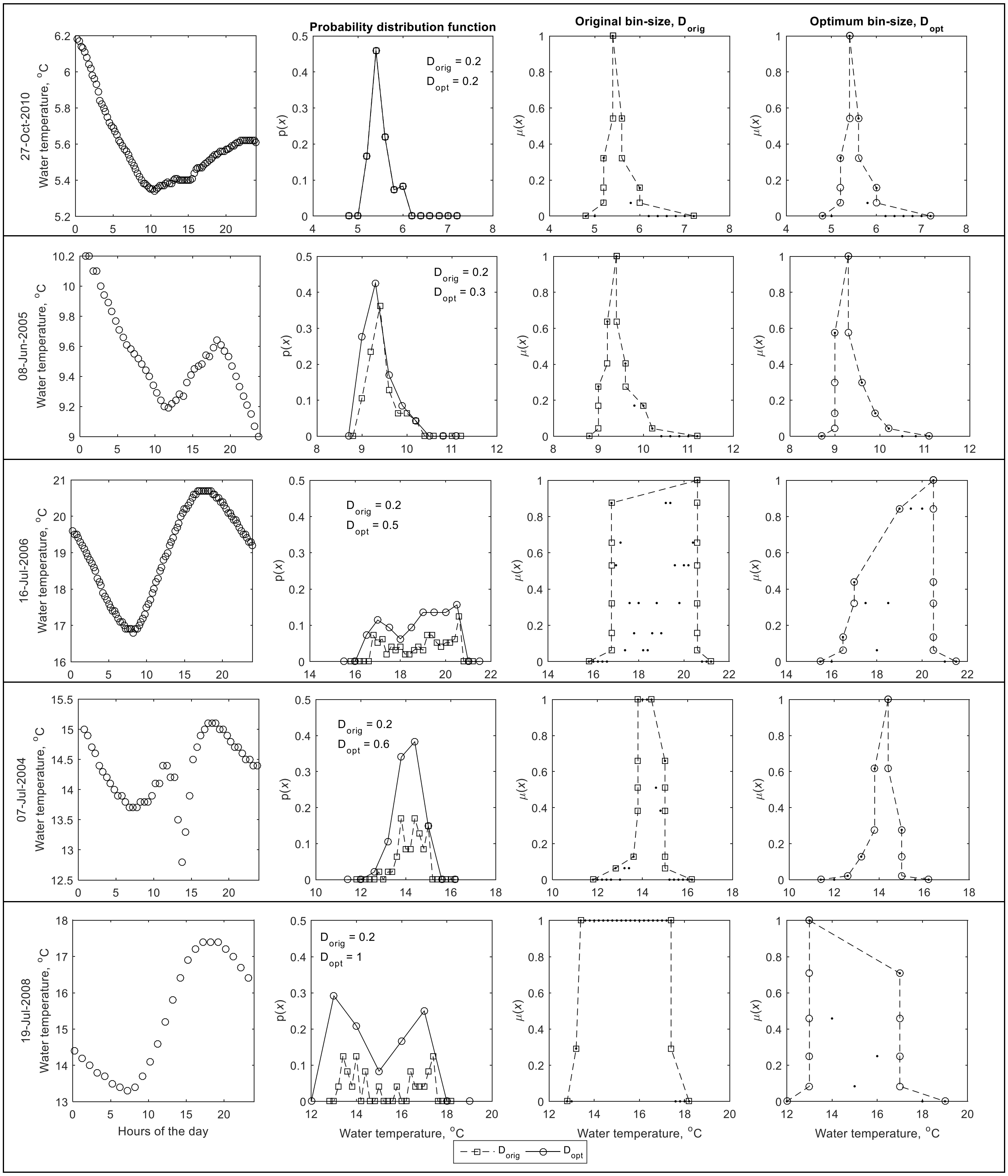



Figure 4: Sample results of probability-possibility transformation for water temperature, T

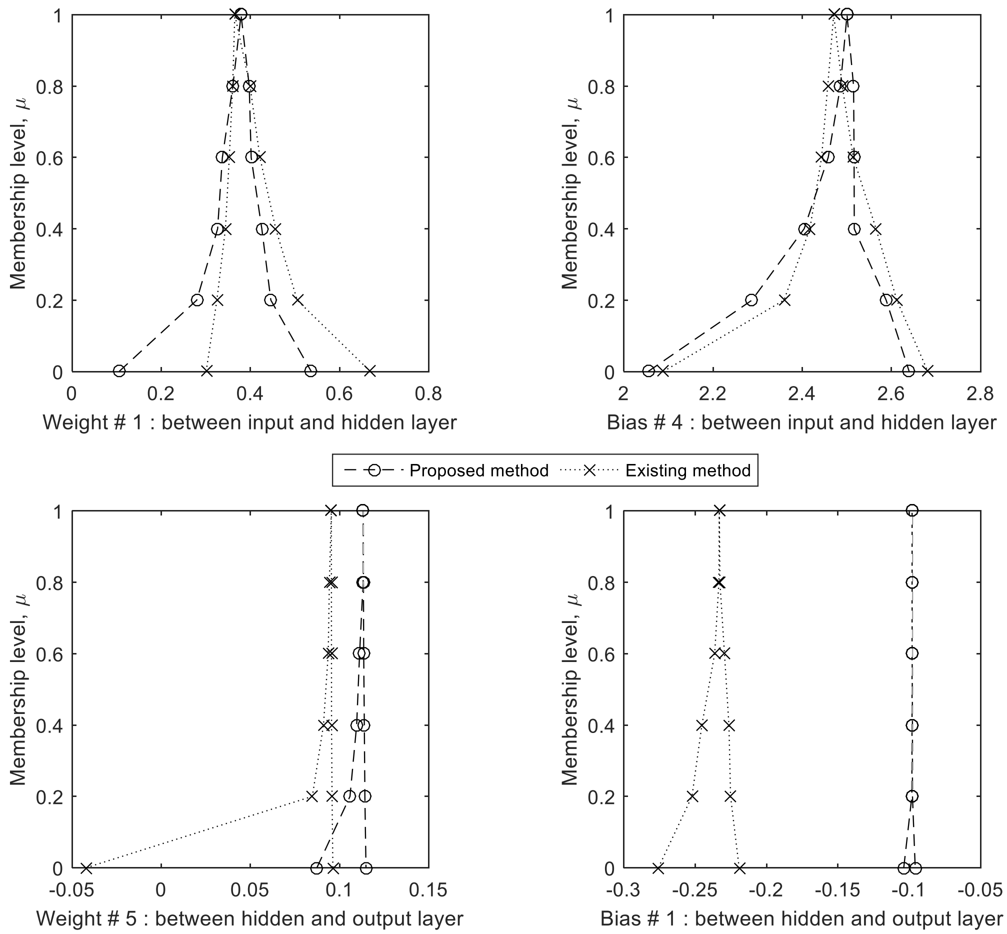



Figure 5: Sample plots of the produced membership functions for the weights and biases of the fuzzy neural network for both the proposed and existing methods

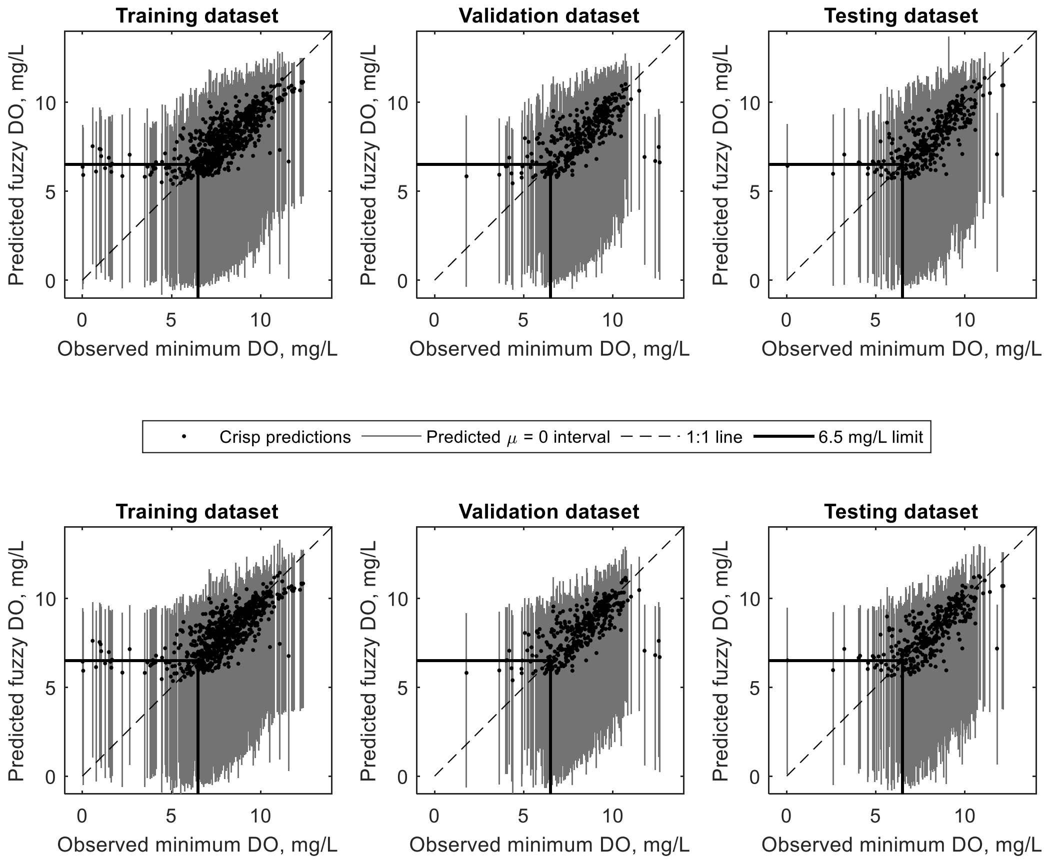



Figure 6: A comparison of the predicted and observed minimum DO at the μ = 0 interval (grey line) and at μ =1 (black dots) for the proposed (top row) and existing (bottom row) methods

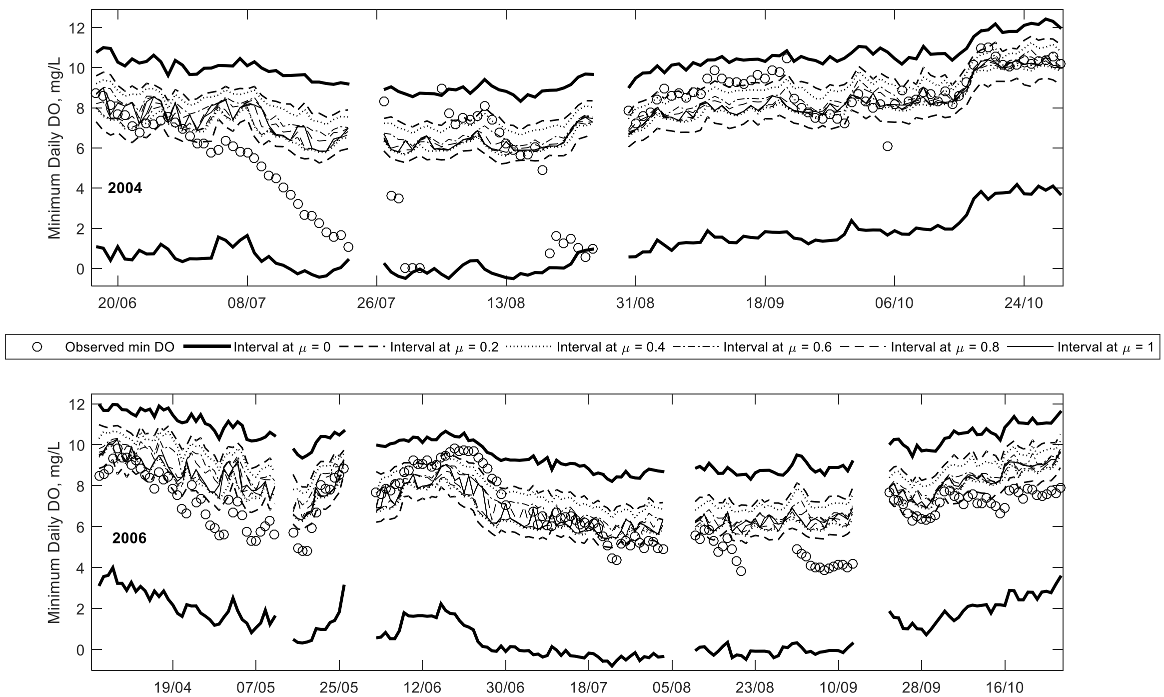



Figure 7: A comparison of the observed and predicted minimum DO trends for: (top) 2004, and (bottom) 2006

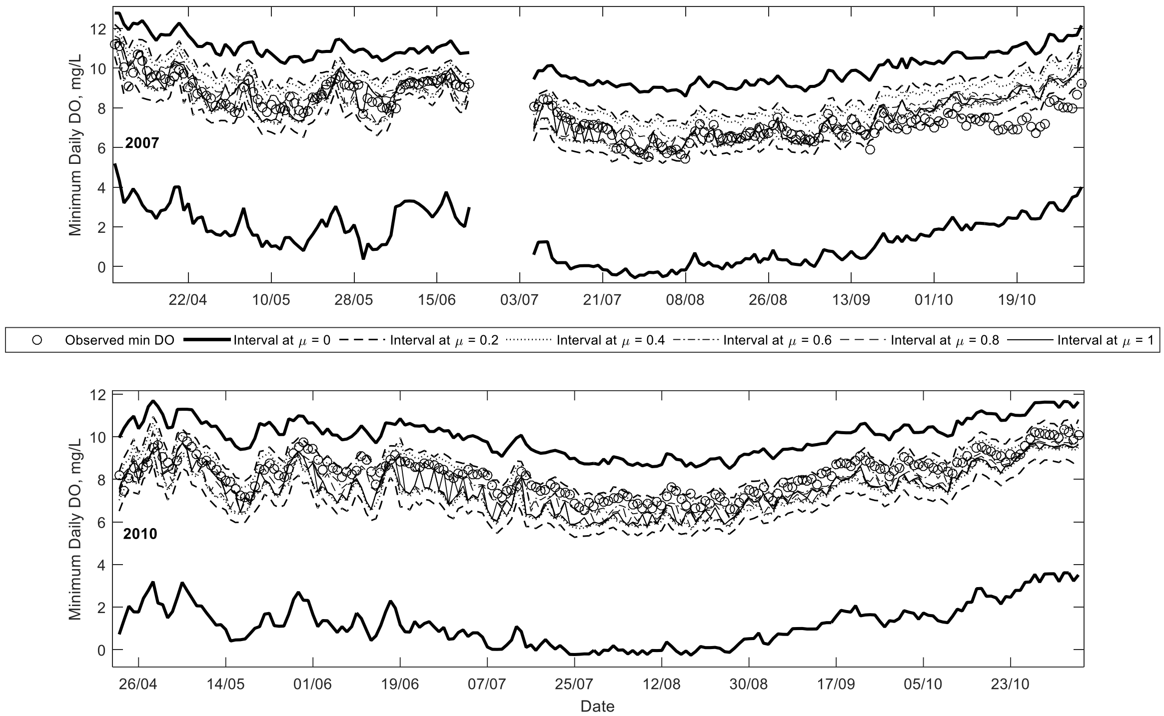



Figure 8: A comparison of the observed and predicted minimum DO trends for three sample years: (top) 2007 and (bottom) 2010

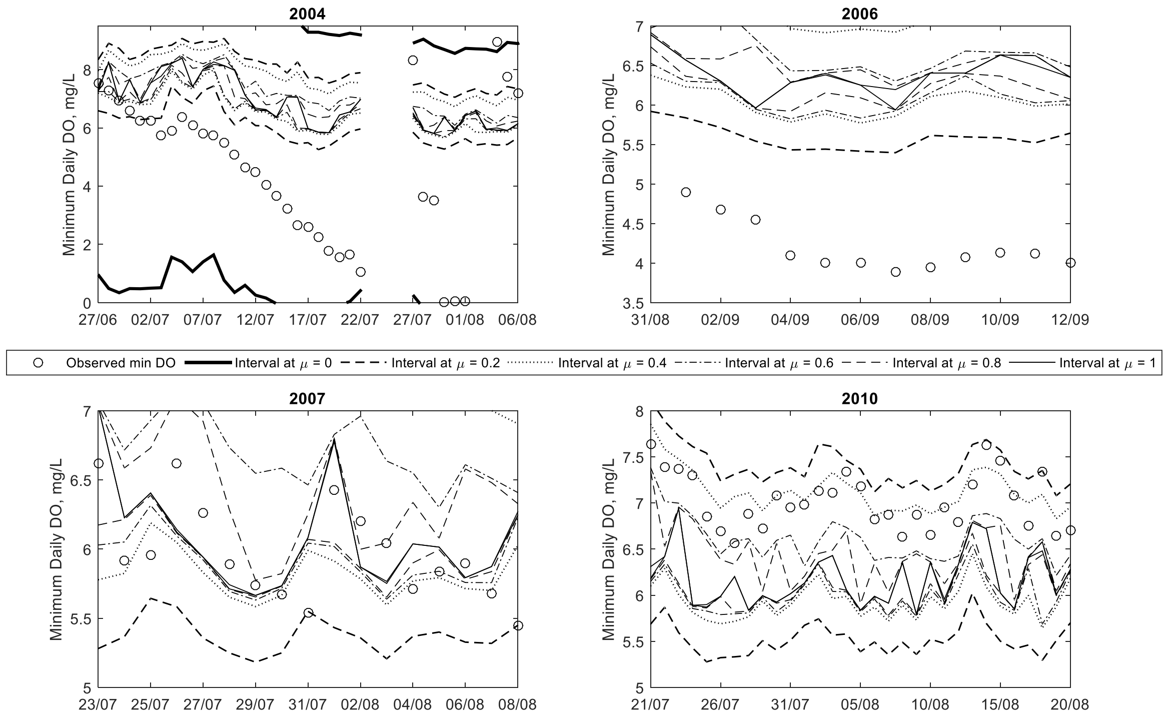



Figure 9: Zoomed in views of the trend plots for four sample year corresponding to important periods with low DO occurrences

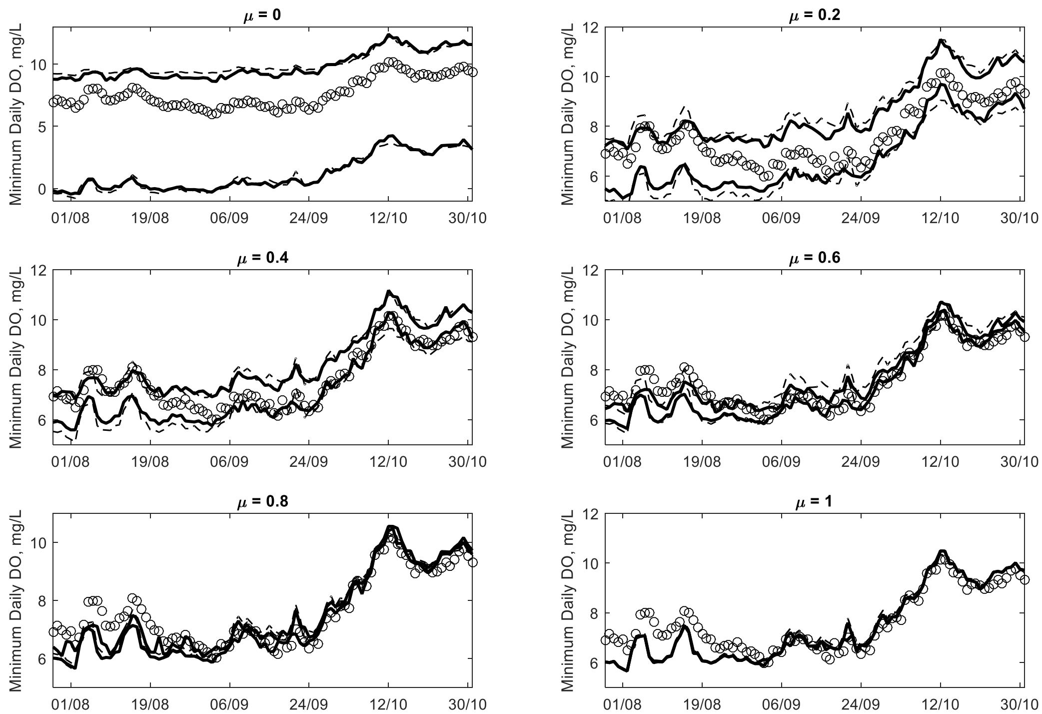



Figure 10: Comparison of predicted trends of the proposed (solid black line) and existing (dashed black line) methods shown for 2009 for each membership level. Observations are shown as black circles

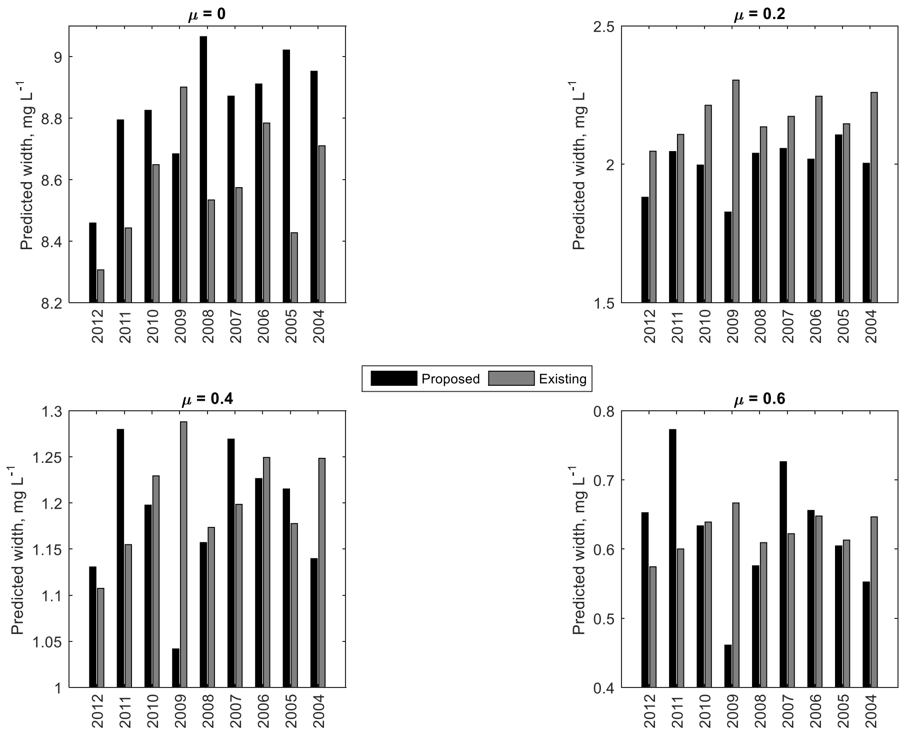



Figure 11: A comparison of average annual interval widths of predicted fuzzy numbers using the proposed and existing FNN methods for four selected membership levels

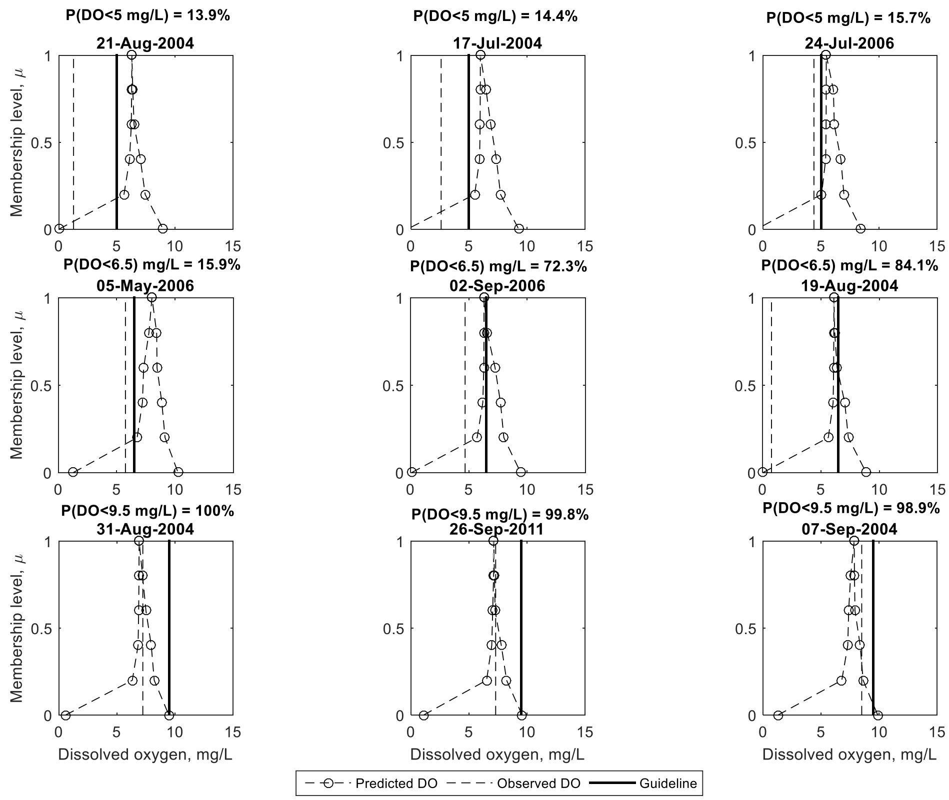



Figure 12: Sample plots of low DO events and the corresponding risk of low DO calculated using a possibility-probability transformation for the (top) 5 mg $L^{-1}$, (middle) 6.5 mg $L^{-1}$, and (bottom) 9.5 mg $L^{-1}$ guideline