River flood prediction using fuzzy neural networks: an investigation on automated network architecture

Usman T. Khan, Jianxun He and Caterina Valeo

ABSTRACT

Urban floods are one of the most devastating natural disasters globally and improved flood prediction is essential for better flood management. Today, high-resolution real-time datasets for flood-related variables are widely available. These data can be used to create data-driven models for improved real-time flood prediction. However, data-driven models have uncertainty stemming from a number of issues: the selection of input data, the optimisation of model architecture, estimation of model parameters, and model output. Addressing these sources of uncertainty will improve flood prediction. In this research, a fuzzy neural network is proposed to predict peak flow in an urban river. The network uses fuzzy numbers to account for the uncertainty in the output and model parameters. An algorithm that uses possibility theory is used to train the network. An adaptation of the automated neural pathway strength feature selection (ANPSFS) method is used to select the input features. A search and optimisation algorithm is used to select the network architecture. Data for the Bow River in Calgary, Canada are used to train and test the network.

Key words | flood prediction, machine learning, risk assessment, uncertainty analysis

Usman T. Khan (corresponding author)
Department of Civil Engineering,
York University,
4700 Keele Street, Toronto, ON,
Canada M3J 1P3
E-mail: *utkhan@yorku.ca*

Jianxun He
Department of Civil Engineering,
University of Calgary,
2500 University Drive NW, Calgary, AB,
Canada T2N 1N4

Caterina Valeo
Department of Mechanical Engineering,
University of Victoria,
P.O. Box 1700 STN CSC, Victoria, BC,
Canada V8W 2Y2

INTRODUCTION

The 2013 floods in southern Alberta (in western Canada) were one of the worst natural disasters in Canadian history. The floods were responsible for four deaths, caused approximately $6 billion in damage (Environment Canada 2013), and displaced more than 100,000 residents (Alberta Government 2014; Khan & Valeo 2016a). This flood event highlighted the need for better short-term (e.g. 1 to 3 day lead time) flood forecasting models which can be used as part of an early warning system (e.g. Alfieri *et al.* 2012) to implement flood protection and mitigation strategies. These emergency actions may have prevented some of the destruction and casualties that occurred in southern Alberta in 2013.

While the mechanisms behind extreme flood events in southern Alberta are generally understood and documented (Valeo *et al.* 2007), predicting floods accurately remains a challenge because of the uncertainty in the numerical models. The uncertainty stems from a number of different sources, for example, the physically-based models (rainfall-runoff models) rely on a simplified representation of highly complex, correlated and spatially distributed processes that occur in a watershed (Vrugt *et al.* 2005; Wijesekera *et al.* 2012). These models cannot fully capture the complexity of the physical system, nor account for their dynamic nature (i.e. they model a stationary system). Furthermore, the data used to calibrate the models may be fraught with errors, or not be representative of the extreme events that the model is used to predict. Often these models give deterministic results which give a false sense of confidence to the users on the accuracy of the model. However, flood forecast models play an integral role in flood protection, and improved models are critically needed to help protect against more frequent and intense floods expected in the future.

Data-driven models are an alternative to the physically-based models described. The types of models use generalised relationships between input and output datasets (Solomantine & Ostfeld 2008), define a system with limited assumptions, and have similar, if not better, performance than physically-based models. A simpler structure means that the uncertainty from different sources (e.g. model parameters) can be quantified and propagated through the

doi: 10.2166/wst.2018.107

model. These types of models (e.g. artificial neural networks, ANNs) have been widely used for predicting flow rate in rivers (Alvisi & Franchini 2011; Duncan *et al.* 2011; Li *et al.* 2015).

A major advantage of using a data-driven approach for flow rate modelling is that data collected from real-time, high-frequency, flow rate monitoring stations can be used to calibrate the model. This type of data is routinely collected in many jurisdictions globally (e.g. by Environment Canada in southern Alberta). This means that the site specific surveys required for physically-based methods are not necessary if these data can be used. Additionally, by continually collecting the most recent data, any changes in the watershed or system are implicitly represented in the data, hence, capturing the dynamic nature of the system. However, data-driven models, like ANNs, also have several types or sources of uncertainty associated with them: (i) the selection of input variables from a larger dataset which are to be used for model prediction (Bowden *et al.* 2005 has an in-depth analysis); (ii) the selection of the architecture of the data-driven model (e.g. Abrahart *et al.* 2012); and (iii) the model parameters (e.g. Khan & Valeo 2016b). Recent reviews of ANN applications in hydrology have indicated that the lack of uncertainty quantification is a major reason for the limited appeal of ANN by water resource managers (Abrahart *et al.* 2012). Most ANN applications have a deterministic structure that generate point predictions without a quantification of the intervals corresponding to these predictions (Kasiviswanathan & Sudheer 2013). This means that the end-users of these models may have excessive confidence in the forecasted values, and overestimate the applicability of the results (Alvisi & Franchini 2011). However, the characterisation of uncertainty in a model is essential for both research and operational purposes. Without this, model results have limited values (Kasiviswanathan & Sudheer 2013).

Thus, in this research, three approaches are proposed to reduce the uncertainty from the three sources identified above. First, an updated automated neural pathway strength feature selection (ANPSFS) method is used to select the optimum input data from a larger dataset. Second, a search algorithm is used to find the optimum network architecture (Khan & Valeo 2017). Third, a possibility-theory-based fuzzy neural network (FNN) algorithm is used (based on Khan & Valeo 2016b) to quantify the uncertainty of the model parameters and outputs by using fuzzy numbers rather than deterministic values.

MATERIAL AND METHODS

Data collection

The Bow River (located in southern Alberta, Canada as shown in Figure 1) originates in Bow Lake in the Rocky Mountains and flows south easterly through Calgary and into Hudson Bay. The River averages a 0.4% slope over its 645 km length and has a drainage area of approximately 25,123 km^2. The Bow River is supplied by the snowpack in the Rocky Mountains, rainfall, and discharge from shallow groundwater. Flow often peaks in the spring (June) while low flows occur in the winter.

The Bow River flows through the commercial centre of Calgary (the largest city in Alberta with a population of approximately 1.2 million). Thus, it is extremely important for accurate and timely peak flow prediction of this river for the safety of the residents, and for protection of assets in central Calgary. Therefore, this research selected the flow rate measuring station of 'Bow River at Calgary' (Station ID: 05BH004), which is located in central Calgary, to demonstrate the performance of the proposed approach.

Eleven years of hourly flow rate data were obtained for the period from January 1, 2000 to December 31, 2010. The annual median flow rate and the annual peak flow rate varied between 47–85 m^3/s and 172–787 m^3/s, respectively. The hourly flow rate collected at the station was filtered by removing dates where shift corrections (usually due to ice conditions in the river) were applied by Environment Canada. In addition to this, since the objective of this research is to focus on high peak flow rate, days with low flow rate (below the annual median ice-free flow rate) were removed. This reduced the data from an original, unfiltered set of 4,018 days of data to 1,860 days.

The hourly flow rate was used to calculate the daily peak flow rate (Q_P) and daily mean flow rate (Q_D). In addition to this, the daily minimum (T_{min}), mean (T_{mean}) and maximum (T_{max}) daily temperature and daily cumulative precipitation (P) was collected for the same period. The input candidates were lagged by 1 to 3 days (as example of short-term lead times), to create a total input array of 18 candidate variables. Table 1 below summarises the model output and inputs, along with the symbols used in this research.

Input variable selection method

ANNs are defined as a massively parallel distributed information processing system (Elshorbagy *et al.* 2010). A

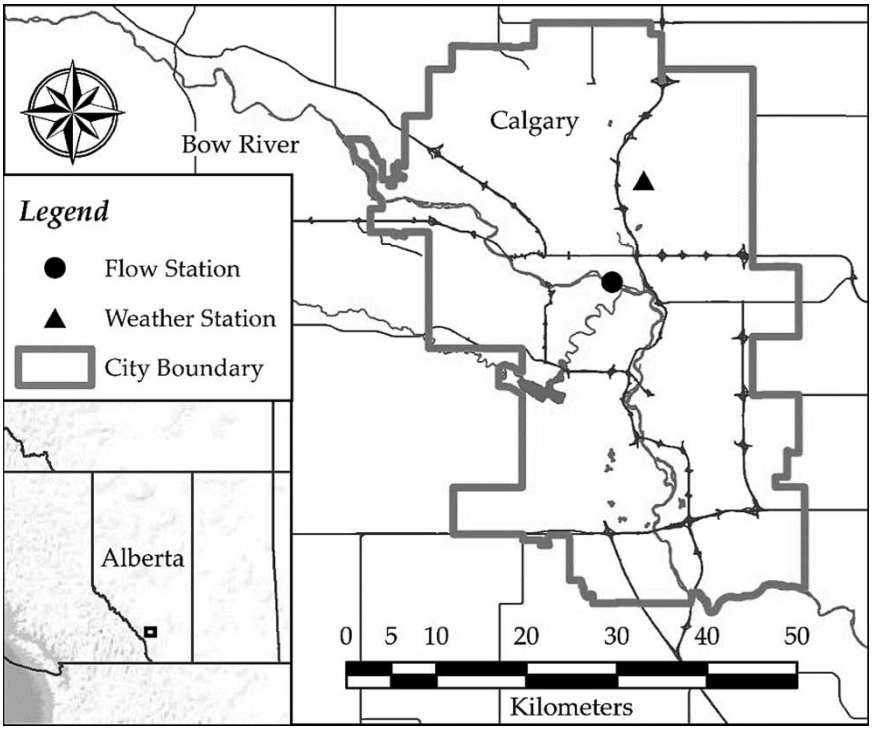

Figure 1 | A map of the City of Calgary (in Alberta, Canada shown in the inset in the bottom left) with the Bow River flowing from the northwest to the southeast of the city, through the city centre.

commonly used type of ANN is called a multi-layer perceptron model which consists of three layers: an input layer, hidden layers (often just one), and an output layer. This structure is one of the most common structures for ANNs (He *et al.* 2011) and, thus, was selected to be able to compare directly with previous studies. Each layer consists of a number of neurons that receives a signal (e.g. the input dataset), and emits an output based on the strength of that signal (quantified by the model coefficients called *weights* and *biases*). Thus, the final output layer is the synthesis and transformation of all the input signals from the input and hidden layers (Khan & Valeo 2016b). The number of neurons in the hidden layer is reflective of the complexity of the system being modelled; more neurons represent a more complex system (Elshorbagy *et al.* 2010).

Duncan (2014) introduced the ANPSFS method; it is designed to reduce the complexity of ANN models by reducing the number of inputs (i.e. the features) from a larger dataset to only those that are most relevant to the model at hand. While most data-driven models are typically considered black-box models, input feature selection methods can potentially provide additional insights into the system by identifying these prominent features. In the ANPSFS method, the strength of a particular pathway from a given

Table 1 | A summary of the input and output variables, symbols and definitions used for this research

Symbol	Definition	Use
$Q_P(t)$	Daily peak flow rate at time *t*	Output
$Q_P(t-1)$; $Q_P(t-2)$; $Q_P(t-3)$	Daily peak flow rate lagged 1, 2, and 3 days	Input
$Q_D(t-1)$; $Q_D(t-2)$; $Q_D(t-3)$	Daily mean flow rate lagged 1, 2 and 3 days	Input
$P(t-1)$; $P(t2)$; $P(t-3)$	Daily cumulative precipitation lagged 1, 2 and 3 days	Input
$T_{min}(t-1)$; $T_{min}(t-2)$; $T_{min}(t-3)$	Daily minimum air temperature lagged 1, 2 and 3 days	Input
$T_{mean}(t-1)$; $T_{mean}(t-2)$; $T_{mean}(t-3)$	Daily mean air temperature lagged 1, 2 and 3 days	Input
$T_{max}(t-1)$; $T_{max}(t-2)$; $T_{max}(t-3)$	Daily maximum air temperature lagged 1, 2 and 3 days	Input

Note that six primary variables are considered, each of which are lagged three times.

input to the output is calculated using Equation (1):

$$W_{IO} = W_{IH}.W_{HO} \quad (1)$$

where W_{IH} are the weights between the input and hidden layer, and W_{HO} are the weights between the hidden layer and output later. The larger the value of the W_{IO}, the more significant the associated input. In the present research, W_{IO} is calculated using all the input data for each year individually rather than collectively for the entire dataset. By doing so, the method will highlight which inputs are consistently significant and also demonstrate the dynamic features of the system. The annual W_{IO} are collected in an array, which is then used to calculate the ensemble interquartile range (EQR) for each of the inputs as described in Equation (2):

$$EQR = \min(|Q_1|, |Q_3|)/\max(|Q_1|, |Q_3|).\,\mathrm{sign}(Q_1).\,\mathrm{sign}(Q_3) \quad (2)$$

where Q_1 and Q_3 are the first and third quartile of the W_{IO}. The EQR are ranked in descending order, and the positively ranked inputs are selected as the most significant features. Thus, the ANPSFS method uses W_{IO} and EQR to select the most important features the output (Q_P) given a set of input data.

Network architecture selection

The three layer architecture selected for this research requires two transfer functions: one between the input- and hidden-layer (the hyperbolic tangent sigmoid function), and one between the hidden-layer and output-layer (a linear function) (following Alvisi & Franchini 2011). Network training was completed using the Levenberg-Marquardt method, to minimise mean squared error (MSE). The input and output data were normalised so that input and output data fell within the interval [−1 1]. However, there is no consistent method for selecting two important components of ANN architecture: the number of neurons in the hidden-layer (n_H), and the amount of the total dataset used for model training, validation and testing (known as data-division) (Abrahart *et al.* 2012). Typically, an ad hoc or trial-and-error method is used to select the number of neurons (Maier *et al.* 2010; Alvisi & Franchini 2011; He *et al.* 2011). If too many neurons are selected, it will increase the complexity and hence the processing speed of the ANN, while reducing the transparency of the model. Insufficient number of neurons risk reducing model performance and forgoing the ability of modelling non-linear systems. Similarly, the issue of data-division, which can have significant impacts on final model structure, is also determined in an ad hoc basis (Maier *et al.* 2010).

In this research, a coupled method to select the optimum n_H and data-division for the ANN model is proposed. The smallest number of neurons and the least amount of data for training is targeted. The first is to reduce computational effort. The second is to prevent the risk of over-fitting to the training data, thus, having a larger dataset for robust statistical inference of the training dataset. Full technical details of this method can be found in (Khan & Valeo 2017; Khan 2017). The algorithm search for n_H between 1 and 20, and training data for data-division is varied between 50% and 75%. Using these criteria, a number of different circumstances where the highest performance can be determined, and the best combination can be objectively selected.

Fuzzy neural network

Khan & Valeo (2016b, 2017) modified an FNN original proposed by Alvisi & Franchini (2011) where the ANN coefficients and the output of the network are fuzzy rather than deterministic numbers. The method uses a stepwise, constrained optimisation algorithm to define the fuzzy number coefficients. These coefficients can be viewed as a series of nested intervals defined at discrete membership levels. Each of the intervals is defined by an upper and lower bound which translates into an α-cut interval, i.e. an interval which has a membership level of α within the fuzzy number. The constraints are defined so that the network captures a predefined amount of observations within each interval, while minimising the width of the given interval. Theoretical details of this approach are expounded in Khan & Valeo (2016b).

Model implementation

The FNN method was implemented in MATLAB (version 2017a). The built-in MATLAB Neural Network Toolbox was used with a two-step optimisation approach: first, the Shuffled Complex Evolution algorithm (Duan *et al.* 1992) was used to find an initial solution to the minimisation problem. Further refinement of the solution was conducted using the MATLAB function *fmincon*. A pseudo-real-time algorithm was used for training, validation and testing. That is, once the most significant inputs and network

architecture were selected based on the ANPSFS method, the final model was trained sequentially. The first year of data (from 2000) were used to train and validate the model, and data from 2001 were used to test the data. Following this, the model coefficients were updated by using data from 2000 and 2001, and testing with data from 2002, and so on. The Nash-Sutcliffe efficiency (R^2) and the root mean square error (RMSE) were used to assess the performance of the network using Equations (3) and (4):

$$R^2 = 1 - \sum(y_m - y_o)^2 / \sum(y_o - y)^2 \quad (3)$$

$$RMSE = \text{sqrt}\left[\sum (y_o - y_m)^2\right] \quad (4)$$

where y_m is the model output, y_o is the observed value, and y is the mean of the observed data.

RESULTS AND DISCUSSION

Input selection

Figure 2 shows the results of calculating the strength (i.e. the W_{IO}) using Equation (1) for each year for all 18 candidate inputs. The figure highlights the relative strength of each input parameter and how it changes on an annual basis. If W_{IO} is negative, the input parameter is not significant in predicting Q_P, i.e. it acts as an inhibitor. However, if W_{IO} is positive, it suggests that the parameter is an important input parameter to predict $Q_P(t)$. Figure 2 demonstrates that the strength of each input parameter varies each year, fluctuating from positive to negative values. This highlights the dynamic nature of the system, i.e. any given year the impact, or strength, of a given input to the output changes. This is one of the major strengths of the adapted ANPSFS approach adopted in the present research: the method is able to identify the dynamic nature of the system by applying the input variable selection on an annual basis rather than over the entire dataset. The figure also highlights that there is no single parameter or set of parameters that are consistently, or continually, above zero, and therefore not significant input parameters over the study period. Thus, there is no clear input parameter that is consistently ranked high, indicating the varying important of each input parameter. Additionally, several candidate variables have an average (for eleven years) value of the strength, W_{IO}, values that are negative, indicating that their potential as an input is low and therefore should not be considered as input parameters to predict Q_P.

Figure 3 further explores the strength of each input variable and shows the ranked values of *EQR*, which was calculated using Equation (2). The *EQR* value takes into account the annual variability (i.e. the ensemble) in the strength of each input variable (i.e. W_{IO}), and thus provides a more useful metric for identifying those variables that have a significant impact in predicting the model output. The *EQR* values in Figure 3 show that only three inputs have a positive (and thus, non-inhibiting) effect: $Q_D(t-1)$: 1-day lagged mean flow, $P(t-1)$: 1-day lagged precipitation, and $T_{min}(t-2)$:2-day lagged minimum air temperature. The other 15 candidate inputs have an inhibiting effect over the ensemble dataset (i.e. the 11 years of data) and, thus, should no longer be considered as inputs into the data driven model. Thus, for this dataset, the three aforementioned

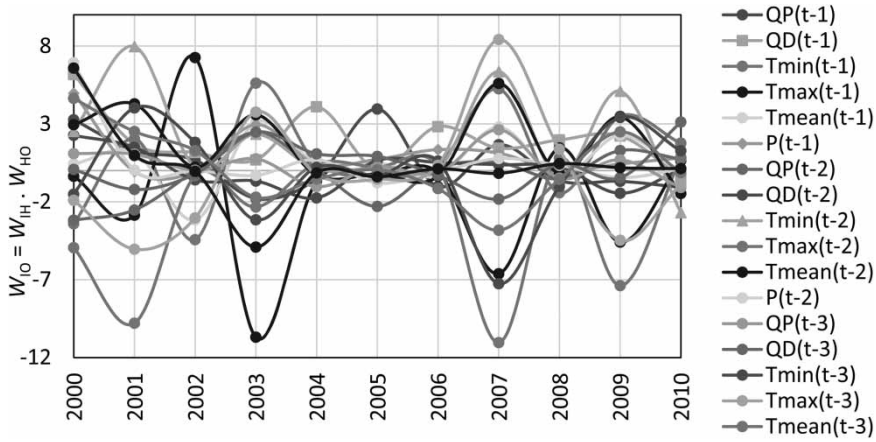

Figure 2 | Strength of each candidate input over the eleven 11-year period.

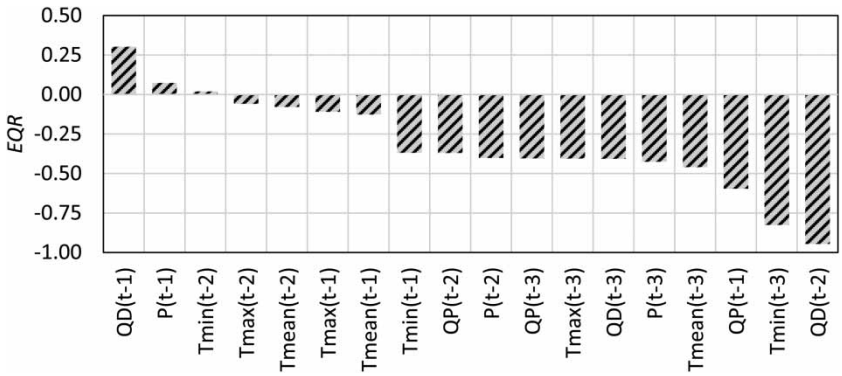

Figure 3 | Ranked values of EQR for the entire dataset.

inputs are selected as the final inputs to the model to predict Q_P.

The significance of these results is that the method can be used to reduce the complexity of the model which has some important implications. First, it allows for the optimisation of model structure by only including the inputs that have a non-inhibiting impact on model performance. A smaller input size means that the computational effort (and thus, time) required to calibrate the model is lower. This means that the model can be used in real-time prediction scenarios that may require frequent re-calibration as more input data is collected. Additionally, since using the ANPSFS method helps identify the most important model input parameters, this information can be used to enhance existing monitoring and data collection networks. For example, if the *EQR* values suggest that a certain parameter provide little benefit in predicting the model output, the monitoring network can be optimised to collect the data (and thus, the inputs) that are most useful in predicting the output. Last, the ANPSFS approach for input variable selection guarantees that the selected inputs are those that provide the best model performance as compared to other combination of input variables. Unlike some other input variable selection methods (e.g. principal component analysis, partial mutual information), the ANPSFS method helps select inputs based on the neural network performance. This approach would be widely applicable in other areas of water resources modelling, and engineering in general. As identified in the Introduction, the selection of input variables is one of the sources of uncertainty in using data-driven models. By using the proposed ANPSFS approach, the uncertainty is reduced, since an objective method is provided to select the model inputs. This approach can be used in other areas, for example in predicting water quality parameters using data-driven models (see Khan & Valeo 2016b) where the selection of input variables is also a necessary step in designing the optimum data-driven model.

Architecture selection

The selection of neural network architecture is the second major source of uncertainty considered in this research. After the three candidate input variables ($Q_D(t-1)$, $P(t-1)$, and $T_{min}(t-2)$) were selected using the adapted ANPSFS method, these inputs were used along with the proposed coupled method to select the optimum network architecture. Figures 4 and 5 show sample results of this proposed method to select the number of neurons in the hidden layer (n_H) and the amount of data used for training, validation and testing.

Figure 4 shows the change in the MSE and epochs (a measure of the duration of training) of the test dataset for the 60%–20%–20% data-division scenario, with n_H varying between 1 to 20 neurons. This data-division scenario means that 60% of the data was used to train the network, 20% was used for validation, and the remaining 20% was used for testing the network. The figure demonstrates that while there is significant reduction in computational effort by increasing the number of neurons to five, the impact on model performance (i.e. MSE) is less clear. Specifically, Figure 4 shows that at lower n_H values (i.e. at 3 and below), the MSE of the dataset is more variable and typically higher. A similar trend is seen at higher n_H values (i.e. 10 or greater). This suggests that there is a window of optimum performance (with respect to MSE) between n_H values of 3 and 10. Looking more closely at the epochs, it is quite clear that the smallest and least complex models (where n_H is equal to 4 or lower), take longer to calibrate. This is not surprising since the number of inputs is three for this model – typically if n_H is lower than the number of inputs,

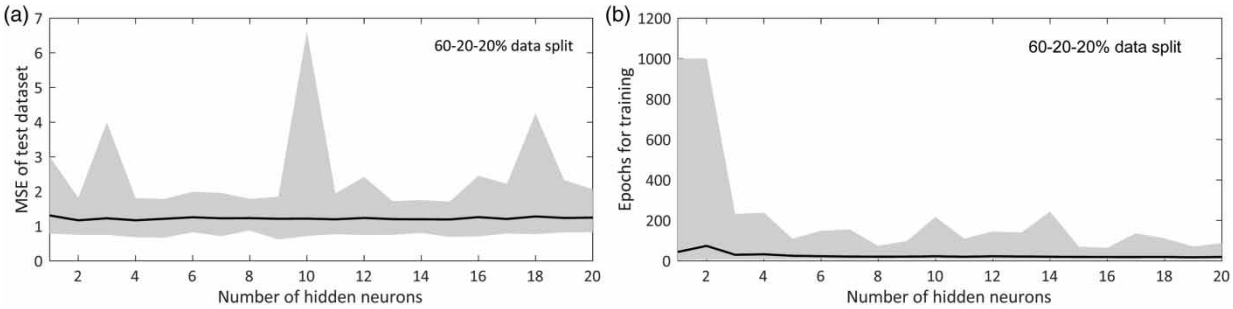

Figure 4 | A comparison of the performance (mean squared error (MSE) and epochs) of the ANN model with increasing model complexity (i.e. higher n_H).

it creates additional constraints in training the network. The epochs reduce drastically at n_H values of 5 or higher, suggesting a wide range of n_H. This suggests that optimum network structure (with respect to epochs) would be where nH is greater than 5. This information, combined with the information related to the MSE performance, suggests that the optimum structure for the network will lie in the range of $n_H > 5$ and $n_H > 10$. This result is significant because it shows that using a systematic approach rather than the typical ad hoc approach can help identify network architecture that is suboptimal, i.e. where performance is highly variable and generally lower. The proposed approach helped identify the conditions where model performance (with respect to MSE and training time, epochs) was best for the given inputs. This approach can be extended to other cases, for example those using different inputs or for other applications and will help reduce a major source of uncertainty in neural network applications.

The second component in analysing the uncertainty due to the network architecture was to test the performance of the network (evaluated similarly using MSE and epochs) as the amount of data used for training was increased from 50% to 75%. This analysis was generally inconclusive as shown in Figure 5 for one case, when $n_H = 5$, though similar results were seen for other values of n_H. The two subplots in Figure 5 do not show a significant trend, or change in variance of the performance of the network as the amount of data for training was increased. In both cases (MSE and epochs), increasing the amount of data did not result in noticeably improved performance. Thus, using the least amount of data for training (i.e. 50%) and therefore having a higher fraction available for testing (25%) would be ideal. This is because all model inferences should be done on the independent test dataset. Having a relatively larger dataset means more robust statistical inference is possible.

Based on these results of comparing network performance whilst changing the number of hidden neurons and the amount of data used for training, the optimum network architecture was selected as $n_H = 5$ with a 50%–20%–20% data-division for this research. The overall outcome of this component of this research was that the proposed method systematically provides objective results instead of using the typical trial-and-error based approach to selecting neural network architecture parameters. The implication for this research, as well as for other applications of data-driven modelling, is that the proposed method helps reduce the uncertainty related to model structure, or architecture. Given the large amount of possibilities of network options, the approach allows end-users to select the

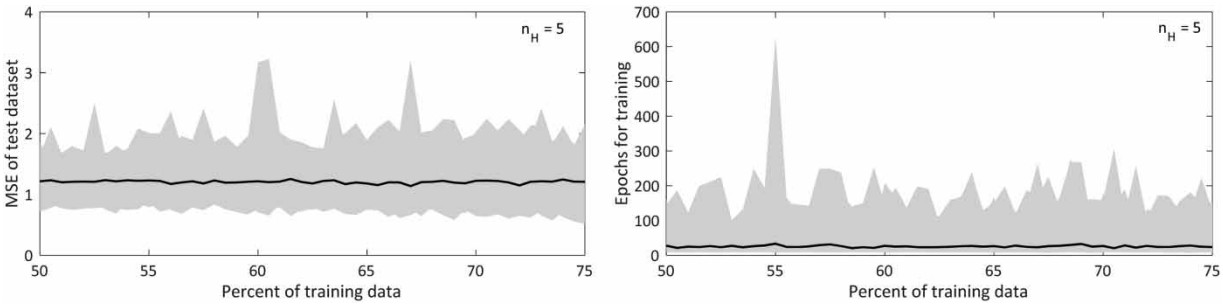

Figure 5 | A comparison of the performance (mean squared error (MSE) and epochs) of the ANN model with increasing proportion of data used for model training.

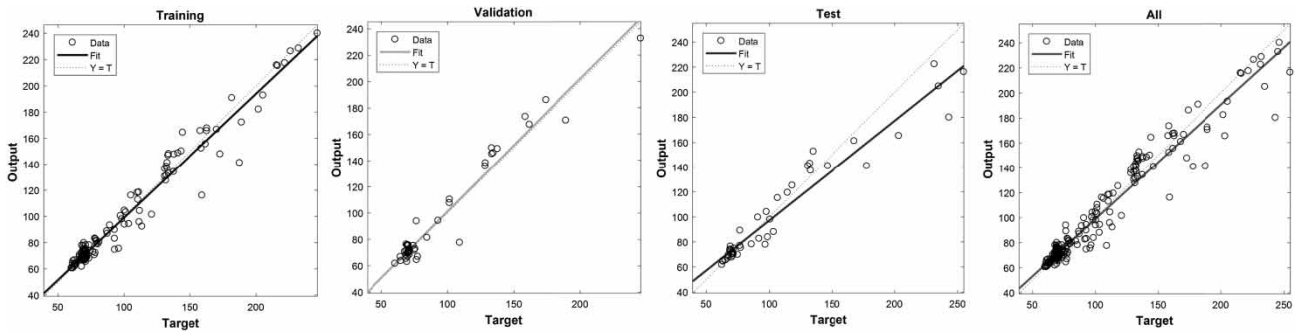

Figure 6 | Sample results for the training, validation, test and complete dataset for the 2004 data; similar results were seen for other years.

parameters that provide the maximum benefit (i.e. model performance) for a given dataset.

Network output

An FNN model was trained using the optimum architecture and three inputs identified in the previous section. Figure 6 shows the comparison of observed (i.e. the target) and predicted (i.e. the output) peak flow rate for each dataset (training, validation, testing and complete data) for 2004 (shown as a sample). Note that the Q_P values for the output are only shown for a membership level of 1 (the highest membership level). These plots demonstrate that the model is able to predict peak flow rate (Q_P) quite accurately with a 1-day lead-time. A notable feature of the modelling approach is demonstrated in each subplot in Figure 6, which is that the model can predict low, medium and high ranges of Q_P equally well. The use of the constrained optimisation ensures that all observations fall within the upper and lower bounds of the predicted intervals. In comparing the different datasets, as expected, the test dataset has the lowest performance (since the model is calibrated on the training dataset). The validation dataset performance is also high – it is used to prevent over-fitting through the application of an early-stopping procedure, as described in the Methods.

Figure 7 includes more details of the FNN performance: it shows the R^2 and *RMSE* values for the entire study period for the independent test dataset only (the performance for the training and validation datasets is typically higher than the test dataset). The R^2 values were above 0.9 for all years except for 2005 (where it was lower, owing to the 2005 flood, at 0.74). The *RMSE* values ranged between 7 and 17 m^3/s for all years except the 2005 flood year

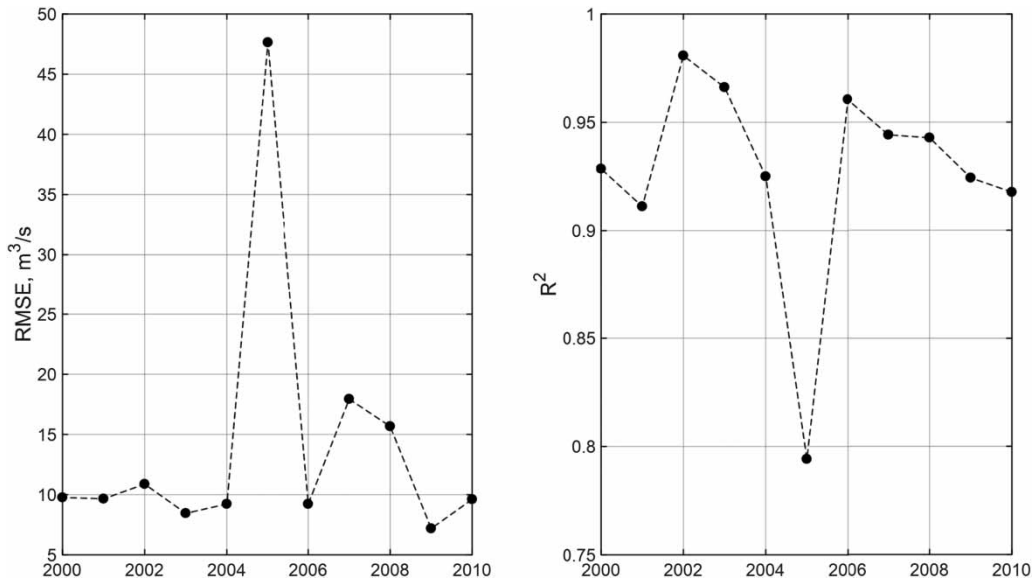

Figure 7 | Model performance metrics, R^2 and *RMSE*, for the test dataset for the 11-year period.

(where it was higher at 47 m^3/s) which are small compared to the median flow rate in the Bow River. Generally speaking, this shows that the FNN model was quite adept at predicting Q_P and that a data-driven approach is suitable to model short-term Q_P, that is those with a short lead-time of 1, 2 or 3 days. A significant reason for the high performance of prediction peak flow rate through the proposed approach is the adoption of a systematic and objective method of selecting important network parameters, such as the number and types of inputs, the n_H, the data division, and the constrained optimisation approach to train the FNN. Each of these components is an important feature to help reduce the overall uncertainty of using data-driven models (particularly neural network based models). Given the model performance metrics shown in Figure 7, the metrics clearly demonstrate the success of the proposed approach, owing to the high performance (high R^2 and low *RMSE*).

CONCLUSIONS

In this research, three different methods were combined to reduce the overall uncertainty in an FNN model used to predict peak flow rate in the Bow River in Calgary, Canada. First, an ANPSFS method was used to determine the most influential inputs from a larger dataset. Second, the network architecture was selected using a search algorithm rather than in an ad hoc manner and, last, a constrained algorithm was used to ensure that the model coefficients and outputs were fuzzy rather than a deterministic value to capture the uncertainty. The impact of this research is two-fold: (i) it demonstrates that data-driven modelling is an appropriate approach for flood modelling in urban watersheds; and (ii) the uncertainty in the model can be accounted for by following the three processes described in this research and, thus, addressing one of the major reasons that prevent the widespread adoption of ANNs by water resource managers.

ACKNOWLEDGEMENT

The authors would like thank: NSERC and York University for funding; Environment Canada for the data; Dr A. P. Duncan (Centre for Water Systems, Exeter) for discussions on ANN feature selection; Ms R. Shakir (York University) for assisting with preliminary data analysis; Mr E. Snieder (York University) for assisting with editing the figures. Lastly, the authors are grateful for the feedback from two reviewers who have helped improve the quality of 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. 2012 Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Prog. Phys. Geog.* **36**, 480–513.

Alberta Government 2014 *Respecting Our Rivers: Alberta's Approach to Flood Mitigation*. Available from: https://pabappsuat.alberta.ca/albertacode/images/respecting-our-rivers.pdf (19 June 2017).

Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F. & Thielen, J. 2012 Operational early warning systems for water-related hazards in Europe. *Environmental Science & Policy* **21**, 35–49.

Alvisi, S. & Franchini, M. 2011 Fuzzy neural networks for water level and discharge forecasting with uncertainty. *Environmental Modelling & Software* **26** (4), 523–537.

Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 Input determination for neural network models in water resources applications. Part 1 – background and methodology. *Journal of Hydrology* **301** (1), 75–92.

Duan, Q., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimisation for conceptual rainfall-runoff models. *Water Resour. Res* **28** (4), 1015–1031.

Duncan, A. P. 2014 *The Analysis and Application of Artificial Neural Networks for Early Warning Systems in Hydrology and the Environment*. PhD Thesis, University of Exeter, Exeter, UK.

Duncan, A., Chen, A. S., Keedwell, E., Djordjevic, S. & Savic, D. 2011 Urban flood prediction in real-time from weather radar and rainfall data using artificial neural networks. In: *Weather Radar and Hydrology*. IAHS Red Book series no. 351. IAHS, Exeter, UK.

Elshorbagy, A., Corzo, G., Srinivasulu, S. & Solomatine, D. P. 2010 Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology-Part 1: concepts and methodology. *Hydrology & Earth System Sciences* **14** (10), 1931–1941.

Environment Canada 2013 *Alberta's Flood of Floods on Canada's Top Ten Weather Stories for 2013*. http://ec.gc.ca/meteo-weather/default.asp?lang=En&n=5BA5EAFC-1&offset=2&toc=show (19 June 2017).

He, J., Chu, A., Ryan, M. C., Valeo, C. & Zaitlin, B. 2011 Abiotic influences on dissolved oxygen in a riverine environment. *Ecological Engineering* **37**, 1804–1814.

Kasiviswanathan, K. S. & Sudheer, K. P. 2013 Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. *Stoch. Env. Res. Risk A.* **27**, 137–146.

Khan, U. T. 2017 Automated feature selection for fuzzy neural networks: an application for urban flood protection. In:

Proceedings of the 2017 CSCE Annual Conference, Vancouver, Canada.

Khan, U. T. & Valeo, C. 2016a Short-term peak flow rate prediction and flood risk assessment using fuzzy linear regression. *Journal of Environmental Informatics* **28** (2), 71–89.

Khan, U. T. & Valeo, C. 2016b Dissolved oxygen prediction using a possibility theory based fuzzy neural network. *Hydrology & Earth System Sciences* **20**, 2267–2293.

Khan, U. T. & Valeo, C. 2017 Optimising fuzzy neural network architecture for dissolved oxygen prediction and risk analysis. *Water* **9**, 381.

Li, Z., Huang, G., Han, J., Wang, X., Fan, Y., Cheng, G., Zhang, H. & Huang, W. 2015 Development of a stepwise-clustered hydrological inference model. *Journal of Hydrologic Engineering* **20** (10), 04015008.

Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. 2010 Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental Modelling & Software* **25**, 891–909.

Solomantine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics* **10** (1), 3–22.

Valeo, C., Xiang, Z., Bouchart, F. J.-C., Yeung, P. & Ryan, M. C. 2007 Climate change impacts in the Elbow River watershed. *Canadian Water Resources Journal*. **32** (4), 285–302.

Vrugt, J. A., Diks, C. G., Gupta, H. V., Bouten, W. & Verstraten, J. M. 2005 Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. *Water Resources Research* **41**, W01017.

Wijesekara, G. N., Gupta, A., Valeo, C., Hasbani, J. G., Qiao, Y., Delaney, P. & Marceau, D. J. 2012 Assessing the impact of future land-use changes on hydrological processes in the Elbow River watershed in southern Alberta, Canada. *Journal of Hydrology* **412**, 220–232.

First received 24 October 2017; accepted in revised form 26 February 2018. Available online 8 March 2018