

## Predicting Dissolved Oxygen Concentration in Urban Watersheds: A Comparison of Fuzzy Number Based and Bayesian Data-Driven Approaches

Usman T Khan<sup>1\*</sup>, and Caterina Valeo<sup>1</sup>

<sup>1</sup>Mechanical Engineering, University of Victoria, BC, Canada

\*Corresponding author's e-mail: [utkhan@uvic.ca](mailto:utkhan@uvic.ca)

### Abstract

Dissolved Oxygen (DO) concentration is significantly and adversely impacted by urbanisation. However, the processes that govern DO concentration in riverine environments are complex, difficult to understand and to model. In the Bow River in Calgary, AB, Canada there is a need to predict DO concentration as part of the overall goal to improve river health for downstream users. In this research three data-driven models are used to predict DO concentration in the Bow River are compared. Fuzzy linear regression performed better than both simple linear regression and Bayesian linear regression. The fuzzy number based model was better able to capture the daily variability in DO in the Bow River and substantially improved low DO predictions.

**Keywords:** Bayesian Linear Regression, Dissolved Oxygen, Fuzzy Linear Regression, Simple Linear Regression, Uncertainty analysis

### Introduction

The impacts of urbanisation on water quality in urban watersheds, and on aquatic ecosystems in general, are significant (Hall, 1984). Dissolved Oxygen (DO) is an important water quality parameter in this aspect, as it is typically used as an indicator of overall aquatic ecosystem health (Khan & Valeo, 2014). This is because urbanisation affects a number of abiotic and biotic factors in aquatic ecosystems, such as water temperature, nutrient concentration, pH level, and discharge rates (He et al, 2011). These factors can directly or indirectly impact DO concentration. Thus, the DO concentration in a system can be seen as a response from multiple factors and is used as an indicator of the collective impacts of urbanisation.

However, this means that modelling and predicting DO in urban areas can be difficult. The relationships between the aforementioned abiotic and biotic factors and DO are poorly understood and are temporally and spatially dependent. The collective and individual impacts of these factors on DO are difficult to quantify. This means that the physically-based simplifications that are typically used to model DO may not adequately represent the complexity or the uncertainty of the system. Thus, there is a need to develop data-driven modelling techniques to predict DO concentration in urban areas that do not have the limitations of physically-based methods, and can incorporate the dynamic and uncertain nature of abiotic or biotic effects.

Unlike physically-based models, data-driven models are based on generalized relationships, links or connections between input and output datasets (Solomantine & Ostfeld, 2008). While physically-based models require very specific (and often difficult to obtain) data to be parameterized, data-driven models can characterize a system with limited assumptions. An advantage of a data-driven approach to predict DO concentration in complex systems is that by

using specific combinations of input and output data, all the different factors affecting DO may be captured indirectly, improving the accuracy of predictions.

Apart from these advantages, data-driven models are also useful when using data from real-time or continuous monitoring systems. These monitoring systems are routinely used in many jurisdictions as part of on-going, long-term data collection efforts for various applications (Environment Canada, 2014). Often the data collected are of high resolution, making it extremely suitable for data-driven models. If these datasets can be effectively utilised, *e.g.* for DO concentration prediction, it further negates the necessity of characterizing physically-based models, which can be expensive, time consuming, and specific to a particular location. On the other hand, the real-time systems are easy to implement in different locations (*e.g.* to get better spatial coverage) making them more amenable for data-driven techniques.

This approach, and the nature of data-driven modelling in general, has intrinsic uncertainties associated with it, which must be identified and propagated through the model. Uncertainty in data-driven modelling techniques is not of a purely random or probabilistic in nature (Dubois & Prade, 1997; Ozbek & Pinder, 2006). Although there are many methods of dealing with uncertainty, such as Bayesian inference based algorithms, this particular application, *i.e.* to predict DO using real-time data and data-driven models is well suited for the use of fuzzy number analysis and possibility theory.

Fuzzy numbers use fuzzy set theory and possibility theory to describe uncertain or imprecise information. A fuzzy number is a specific type of quantity that expresses uncertain or imprecise quantities, measurements or observations (Bárdossy et al, 1990; Guyonnet, et al., 2003; Zhang & Achari, 2010; Huang, et al, 2010). An advantage of using fuzzy numbers is that some of the strict assumptions associated with probability theory can be relaxed (Bárdossy et al 1990; Kim et al, 1996; Kahraman et al, 2006). This is especially important in environmental engineering applications, where often only limited data or data from a number of different sources is available.

Fuzzy numbers have been widely used in hydrology and environmental engineering to represent uncertainty in the parameters of numerical models. The literature clearly demonstrates the utility and advantage of using fuzzy numbers; a summary of these applications can be found in Khan and Valeo (2014). However, there is a need to adapt data-driven methods to be able to use fuzzy numbers as inputs, outputs and model coefficients. In this paper, a fuzzy linear regression (FLR) method, proposed in Khan and Valeo (2014) is described. The utility of this new method is illustrated by a direct comparison with simple linear regression (SLR) and Bayesian linear regression (BLR) to predict daily DO in the Bow River in Calgary, Alberta, Canada. The BLR model is included for comparison because a Bayesian approach for uncertainty analysis is extremely popular in the field (Han et al 2014), and provides a good baseline to illustrate the usefulness of the fuzzy approach.

## Dissolved Oxygen in the Bow River

The Bow River is a river that flows through Calgary (see Figure 1 for a map and location of the river within the city). It has an average annual discharge of 90 m<sup>3</sup>/s and provides the majority of the potable water for the city (Robinson et al, 2009). It is also widely used for recreation, and supports an aquatic ecosystem used for its fish resources. Water from the river is diverted from within the City limits to an irrigation district east of the city. The City of Calgary has three

## Joint Conferences:

The 2014 Annual Conference of the International Society for Environmental Information Sciences (ISEIS)

The 2014 Atlantic Symposium of the Canadian Association on Water Quality (CAWQ)

The 2014 Annual General Meeting and 30th Anniversary Celebration of the Canadian Society for Civil Engineering Newfoundland and Labrador Section (CSCE-NL)

The 2nd International Conference of Coastal Biotechnology (ICCB) of the Chinese Society of Marine Biotechnology and Chinese Academy of Sciences (CAS)

wastewater treatment plants (WWTP) as well as several stormwater outfalls that discharge effluent into the Bow River (He et al, 2011). In addition to this, downstream users who rely on the river for potable water, for irrigation and to support local fisheries.



**Figure 1.** An aerial image of the City of Calgary (a) Bonnybrook WWTP (b) Fish Creek WWTP, (c) Pine Creek WWTP, (d) Upstream of Highwood sampling location, (e) Bow River at Bearspaw Dam

Calgary has been one of the fastest growing cities in Canada over the last decade, with a population that has grown to exceed 1 million people. The impacts of rapid and extensive urbanisation can be seen on the health of the Bow River, with occasional occurrences of low DO concentration, *i.e.* below the provincial chronic or acute guidelines (He et al, 2011). The city is mandated by Alberta Environment to reduce the total daily maximum loadings of nutrients and sediment into the River, as an attempt to improve the health of the river, and to reduce the occurrences of low DO concentration (He et al, 2011). The city has taken a number of steps to reduce the impacts of urbanisation and to ensure compliance with provincial guidelines. One component of this strategy has been to model various water quality parameters, including DO, in the river, using a physically-based model (He et al, 2011). The intention of this model is to be able to quantify the impacts of proposed changes within the watershed that impact water quality to ensure that DO concentration meet provincial guidelines.

## Objectives

The objective of this research is to model daily DO concentration in the Bow River, at a site downstream of the three WWTPs in Calgary. Three data-driven models will be used: SLR, BLR and FLR. Model structure is limited to first-order autoregressive models; this structure allows the prediction of future DO concentration using current DO measurements. The performance of each model will be assessed using a suite of metrics. The ability of the models to capture the observed uncertainty in DO concentration, especially low DO, will be quantified.

## Methods

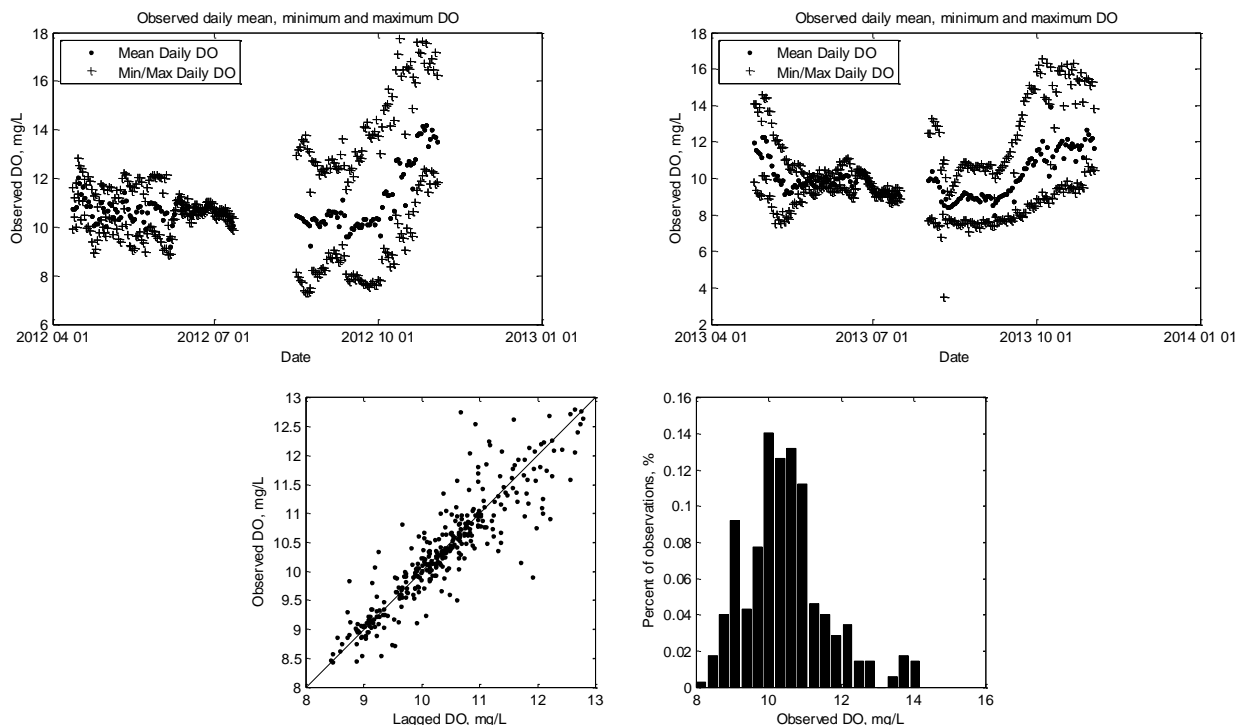
### Data Collection

For this research, real-time (*i.e.* hourly) water quality data for the years 2012 and 2013 were obtained from the City of Calgary. The monitoring site is “Bow River Upstream of Highwood” is

located downstream of the city limits (see Figure 1). The data was collected using a YSI sonde that measured water temperature, specific conductivity, DO concentration, pH and turbidity. Only data from the ice-free period (roughly April through November every year) is used in this research. Initial data analysis suggested that DO is serially dependent on the 1-day lagged DO time series:

$$DO(t) = f(DO(t - 1 \text{ day})) \quad [1]$$

where  $DO(t)$  is the mean concentration of DO on day  $t$ ,  $DO(t - 1)$  is the mean concentration of DO on day  $t - 1$ . This relationship is exclusively explored in this paper as it exhibited the strongest correlation between independent and dependent variables compared to the other listed parameters. The 2012 dataset was used for model construction, while the 2013 dataset is used for model validation.



**Figure 2.** A trend plot of observed mean daily DO for 2012 and 2013 at Upstream of Highwood, a scatted plot of observed DO and 1-day lagged DO, and a histogram of observed daily mean DO

### Simple Linear Regression

Simple Linear Regression (SLR) is an ordinary least squares approach for estimating model parameters in a linear regression model, and is possibly the most widely used model in hydrology (Haan 1977). In this approach, the model is of the form:

$$y = A + Bx \quad [2]$$

where  $x$  are the independent variable, in this case lagged mean daily DO, and  $y$  is the dependent variable, mean daily DO. For the SLR model, the 24 hourly data points for each collected at the “Bow River Upstream of Highwood” sampling location were averaged to determine  $y$  and  $x$ . The model parameters  $A$  and  $B$  can be estimate using the standard approach and is not described here. In SLR, it is assumed that the residuals between the observed and predicted values of  $y$  follow a

Normal distribution with a mean of 0, and variance  $\sigma^2$ . An unbiased estimate of  $\sigma^2$  can be calculated by:

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \quad [3]$$

where  $\hat{y}_i$  is a realization of  $y$  estimated using the regression equation. This estimate of the variance can be used to calculate the confidence interval (CI) of the regression line and of the individual predictions as follows:

$$\overline{\hat{y}}_{kL}^U = \hat{y}_k \pm s \left( \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^{1/2} t_{\left(\frac{1-\alpha}{2}, n-2\right)} \quad [4]$$

$$\hat{y}_{kL}^U = \hat{y}_k \pm s \left( 1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^{1/2} t_{\left(\frac{1-\alpha}{2}, n-2\right)} \quad [5]$$

where  $\overline{\hat{y}}_{kL}^U$  represents the mean upper and lower response,  $\hat{y}_k$  is the estimated value of  $y$  at time  $k$ ,  $x_k$  is the value of interest, and  $t_{\left(\frac{1-\alpha}{2}, n-2\right)}$  is the value of the  $t$ -distribution where  $\alpha$  is the probability of the interval (95% CI in this case), with  $n - 2$  degrees of freedom (Haan 1977).

### Bayesian Linear Regression

A Bayesian Linear Regression (BLR) is an approach to linear regression within the context of a Bayesian framework. In general, BLR consists of a Bayesian posterior inference of model parameters based on a non-informative prior distribution of the typical Normal linear model:

$$y|[A; B], \sigma^2 \sim N(A + Bx, \sigma^2) \quad [6]$$

The standard, non-informative prior distribution of the model parameters is uniform:

$$p([A; B], \sigma^2 | x) \propto \frac{1}{\sigma^2} \quad [7]$$

The conditional posterior distribution of the regression coefficients given  $\sigma^2$  is Normal:

$$[A; B] | \sigma^2, y \sim N([A_{SLR}; B_{SLR}], V_{A,B,SLR} \sigma^2) \quad [8]$$

where  $A_{SLR}$  and  $B_{SLR}$  are the SLR estimates of  $A$  and  $B$ , respectively, and  $V_{A,B,SLR} \sigma^2$  is the variance. Using this and Bayes theorem, the marginal posterior distribution of  $\sigma^2$  is defined as:

$$p(\sigma^2 | y) = \frac{p([A; B], \sigma^2 | y)}{p([A; B] | \sigma^2, y)} \quad [9]$$

This distribution is an inverse scaled chi-squared distribution (note that in equation 10  $s^2$  is the same as in the SLR case):

$$\sigma^2 | y \sim Inv \chi^2(n - 2, s^2) \quad [10]$$

Sampling from the posterior distribution requires that first  $A_{SLR}$ ,  $B_{SLR}$  and  $V_{A,B,SLR}$  are estimated using SLR. These values are used to calculate  $s^2$ . With this,  $\sigma^2$  is estimated by randomly sampling from the distribution (equation 10). The values of  $A$  and  $B$  are then determined by sampling from their respective posterior distributions (equation 8). With these estimates,  $\hat{y}$  can be sampled from

$$\hat{y} \sim N(A + Bx_k, \sigma^2) \quad [11]$$

where the values of  $A$ ,  $B$  and  $\sigma^2$  are values sampled from their respective distributions, *i.e.* for each sampled value of  $A$ ,  $B$  and  $\sigma^2$ ,  $\hat{y}$  is sampled from the Normal distribution, creating a vector

of results, for each  $x$ . For the simple BLR regression case discussed here,  $\hat{y}$  may also be estimated via as a multivariate  $t$ :

$$\hat{y} \sim t_{n-2}(A_{SLR} + B_{SLR}x_k, s^2(1 + x_k V_{A,B,SLR} x_k^T)) \quad [12]$$

Thus to implement BLR, results from the SLR for the coefficients and variance are used to sample the predicted daily mean values of DO. Along with the mean values, the minimum and maximum sampled values and the 95% CI are also collected for comparison with observed data and the SLR CI results, respectively. MATLAB was used to program BLR; 10,000 random samples were taken to estimate the values of  $\sigma^2$ ,  $A$ ,  $B$  and  $\hat{y}$  (Gelman et al, 2004).

### Fuzzy Linear Regression

Fuzzy linear regression is an attempt to extend linear regression for applications involving fuzzy numbers (Khan and Valeo, 2014). It provides an alternative method when SLR may not be possible, whether it is due to the inability to meet strict assumptions or if there is obvious fuzziness in the underlying data or process. In crisp regression, the deviations from the true data can be caused by measurement and modelling error. FLR tries to capture the vagueness, and the non-random, or fuzzy error in the model structure: it is assumed that deviations are due to system fuzziness, *i.e.* the fuzziness of the regression coefficients (Chang and Ayyub, 2001).

In this research a fuzzy linear regression is proposed by Khan and Valeo (2014) is implemented to predict daily DO. This FLR method is unique in that fuzzy number inputs, outputs and regression coefficients are used, whereas other FLR techniques do not typically use fuzzy numbers for each of these parameters (Khan and Valeo, 2014). In addition to this, the proposed method uses non-linear membership functions to define fuzzy numbers. This is more suitable for the required application. A probability-possibly transformation is used to convert the observed hourly DO data to daily fuzzy numbers. The background for this transformation can be found in in Dubois et al (1993) and Dubois et al (2004), and is not discussed further here.

The objective of FLR method is similar to the SLR method: instead of minimizing the residual between an observed and regressed value, the distance between two fuzzy numbers is minimized. Given a set of fuzzy observations  $\tilde{x}_i$  and  $\tilde{y}_i$ , and their corresponding membership functions,  $\mu(\tilde{x}_i)$  and  $\mu(\tilde{y}_i)$ , for ( $i = 1, 2, \dots, n$ ) an FLR model is defined as:

$$\tilde{y} = \tilde{A} + \tilde{B}\tilde{x} \quad [13]$$

where the coefficients  $\tilde{A}$  and  $\tilde{B}$  are fuzzy numbers. The objective is to solve the following least-squares problem:

$$\min r(\tilde{A}, \tilde{B}) = \sum_{i=1}^n d^2(\tilde{y}_i, \tilde{A} + \tilde{B}\tilde{x}_i) \quad [14]$$

where  $d^2(\tilde{y}, \tilde{A} + \tilde{B}\tilde{x}) = \cup [\tilde{y}_i - \tilde{A} - \tilde{B}\tilde{x}_i]_{\mu}$  for  $i = 1, 2, \dots, n$  and  $\mu = 0$  to 1. The metric  $d$  measures the sum of the squared-deviations of the observed ( $\tilde{y}_i$ ) and predicted ( $\tilde{A} + \tilde{B}\tilde{x}_i$ ) intervals  $[\dots]_{\mu}$ , for all alpha-cuts between  $\mu = 0$  and  $\mu = 1$ . Using fuzzy arithmetic ensures that the coefficients  $\tilde{A}$  and  $\tilde{B}$  are normal and convex, a requirement of fuzzy numbers. The results from predictions of DO using FLR will be compared to the minimum, mean and maximum observed data, the predicted daily mean DO using SLR and the corresponding CI, and with the BLR predictions, the corresponding CI as well as the minimum and maximum sampled data.

### Model Performance and Error Analysis

The coefficient of determination ( $R^2$ ), the Nash-Sutcliffe model efficiency (NSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Akaike Information Criterion (AIC) were used to measure the performance of each model. For the FLR results, these methods cannot be directly calculated, thus these values were calculated at five different membership levels between  $\mu = 0$  and  $\mu = 1$ . In addition to this, the percent of observed data following outside the lower limits of the 95% CI for the SLR method, outside the minimum sampled data for the BLR and outside the crisp element defined by alpha cut at  $\mu = 0^L$  for the FLR method was quantified. The distance between these points was also quantified by calculating the mean squared error (MSE). These metrics are used to show the amount of observed variability captured by each model.

### Results and Discussion

Data from 171 days between 12 April and 3 November 2012 was used to calibrate the models. The estimates for the regression coefficients using SLR were 0.918 and 0.892 for  $A$  and  $B$ , respectively. An analysis of the residuals using SLR is shown in Figure 3; it illustrates that the assumptions of SLR have not been violated when modelling DO as a function of 1-day lagged DO. The residuals can be assumed to follow a Normal distribution, there is no significant correlation between the observed values and the residuals, the variance of the residuals is not a function of time, and the residuals do not exhibit serial correlation. Thus, based on this, the model is well specified.

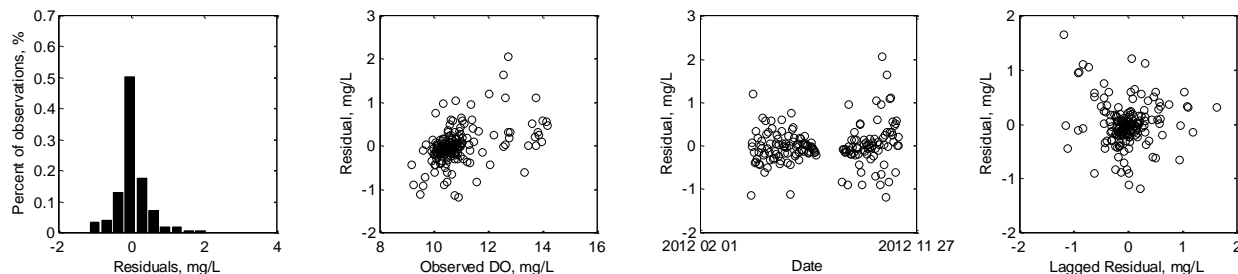
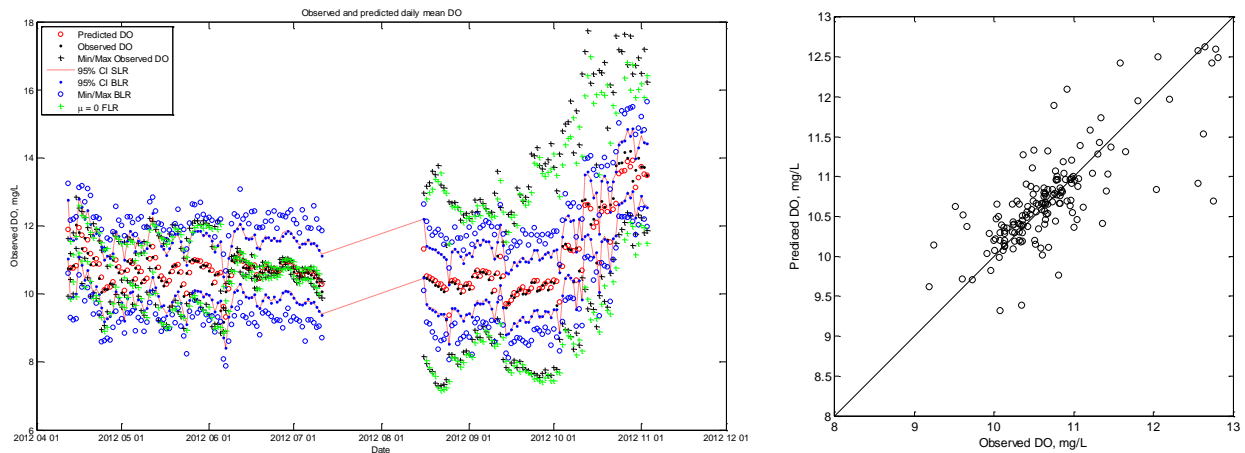


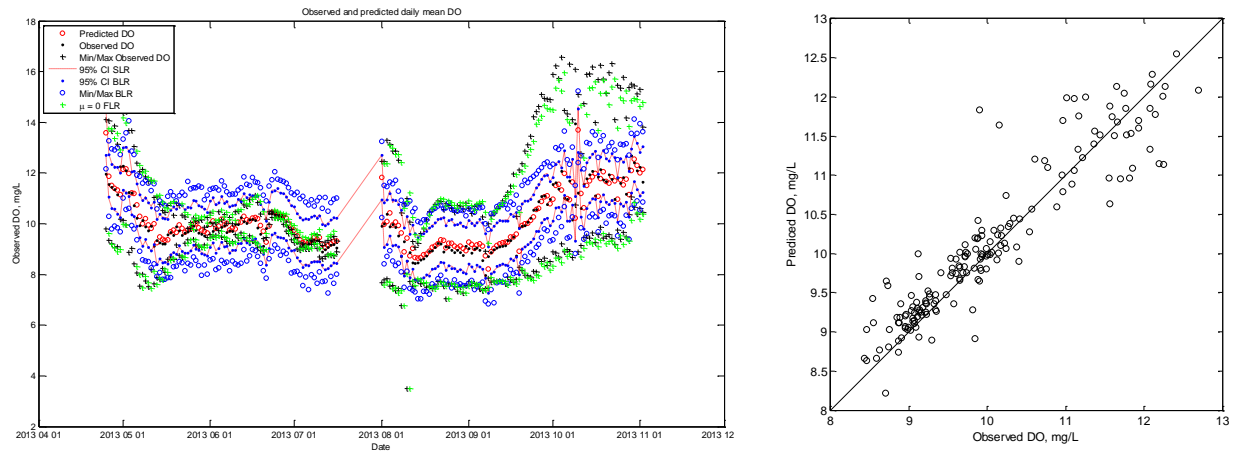
Figure 3. Residual analysis of the SLR method

Figure 4 shows a plot of observed and predicted daily mean DO versus time from each model (including the daily maximum and minimum observed DO) and observed versus predicted daily mean DO. The error terms (95% CI) for SLR and BLR are included, as are the minimum and maximum realization of  $\hat{y}$  using BLR sampling. Results for FLR are only shown at  $\mu = 0^{L,R}$ . The figure shows that the SLR model can accurately reproduce the observed mean daily DO trend, but it cannot capture the observed variability. Most of the minimum and maximum data fall outside the range of the 95% CI. The BLR model performs similarly; however, the minimum and maximum sampled  $\hat{y}$  show greater variability than SLR. The FLR model fully captures the observed variability in mean daily DO. Apart from the temporal trend of mean daily DO, the minimum and maximum trend is also captured (as seen by the green crosses in the Figure 4). Also, the FLR results are more sensitive to changes in observed variability, whereas the BLR method provides a range that is mostly constant.



**Figure 4.** Predictions from each model for 2012 (L); predicted vs. observed daily DO concentration using SLR (R)

Similar results can be seen for the validation dataset (which included 177 daily mean DO values from 25 April to 2 November 2013); these are illustrated in Figure 5 below. Again the FLR model can capture the mean daily DO trend and also better reproduce the observed variability of daily DO. The model is more sensitive to changes in DO variation as well as to seasonal trends seen in DO. In both datasets, the variability of DO is low in between April and July and higher July onwards. Predictions from the BLR method do not reproduce this seasonal trend, as its upper and lower limits of predictions are constant throughout the study period.



**Figure 5.** Predictions from each model for 2013 (L); predicted vs. observed daily DO concentration using SLR (R)

This point is further illustrated in Figure 6 which shows a plot of the independent variable (lagged mean daily DO) and dependent variable (mean daily DO) for both datasets, from each model. While most of the observed mean daily DO falls within the 95% CI for both the SLR and BLR cases, the minimum and maximum DO is not. This is particularly important when low DO is being modelled – these models suggest a risk of over-predicting DO. On the other hand, the FLR model results show that most of the observed low DO is captured by the lower bounds (*i.e.*  $\mu = 0^L$ ). These differences are further quantified in the next section.



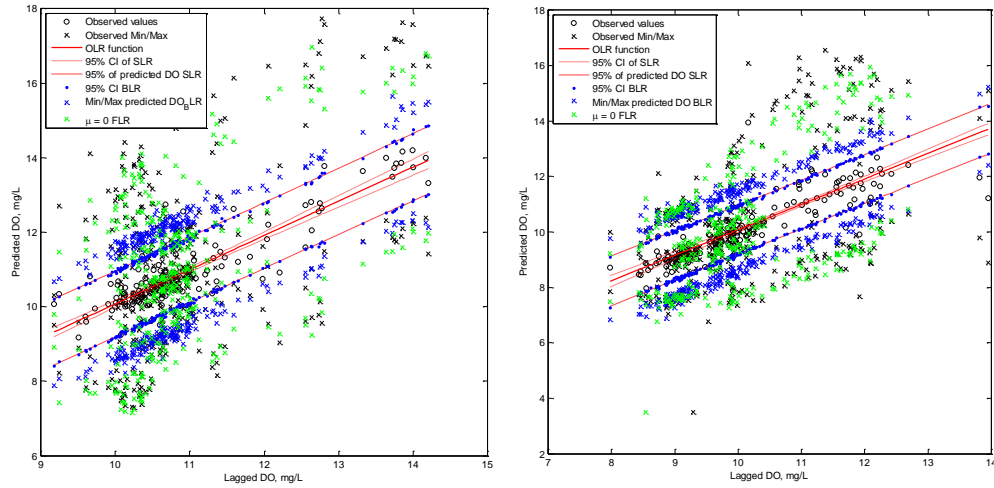
Joint Conferences:

The 2014 Annual Conference of the International Society for Environmental Information Sciences (ISEIS)

The 2014 Atlantic Symposium of the Canadian Association on Water Quality (CAWQ)

The 2014 Annual General Meeting and 30th Anniversary Celebration of the Canadian Society for Civil Engineering Newfoundland and Labrador Section (CSCE-NL)

The 2nd International Conference of Coastal Biotechnology (ICCB) of the Chinese Society of Marine Biotechnology and Chinese Academy of Sciences (CAS)



**Figure 6.** A comparison of model predictions for the 2012 dataset (L) and 2013 dataset (R)

## Error analysis

The table below summarizes the performance of each model, for both datasets. The SLR and BLR model performance is essentially identical to two significant digits. The FLR model outperforms the other two models for  $R^2$ , NSE, and AIC, while having larger RMSE and MAE.

**Table 1.** Summary of error statistics

	Calibration			Validation		
	SLR	BLR	FLR	SLR	BLR	FLR
$R^2$	0.81	0.81	0.85	0.78	0.78	0.79
NSE	0.81	0.81	0.84	0.77	0.77	0.77
RMSE	0.44	0.44	0.52	0.54	0.54	0.56
MAE	0.29	0.29	0.36	0.32	0.32	0.37
AIC	-137	-137	-1008	-107	-107	-846

This analysis was conducted again after using only half the validation data, *i.e.* instead of using 24 data points (representing 24 hours) to calculate the mean daily DO to input into the calibrated model, only 12 data points were used. This was done to simulate model performance at lower data resolution. These results showed that  $R^2$  and NSE decreased for each model; however the reduction in the FLR model performance was only half compared to the reduction in performance of the other two models. This is an important result as it illustrates that beyond the direct benefits of using FLR, it is also more suitable when less data is available. Another similar analysis at quarter-resolution show further gains by using FLR versus the other two regression methods.

Lastly, the number of data points where the observed minima fell outside the lower bound of predicted intervals (e.g. the 95% CI for the SLR case) was calculated. This gives an indication of the amount of variability captured by each model, and more importantly, whether or not low DO is being correctly identified. For SLR, approximately 30% of the data-points over the two year period are within the predicted range. This means that for up to 70% of the study period, minimum observed DO was below the predicted range. Thus cases of regulatory infraction may be under-predicted with the SLR model. For the BLR method, 56% of data-points are within the predicted range. And lastly, using the FLR method, 60% of the minimum observed DO fell within the predicted range (at  $\mu = 0$ ). This is a clear improvement over the SLR case.

A further improvement of the FLR method is apparent in the MSE between the observed minima and the lower bounds of the predicted intervals. For the FLR method, the MSE is approximately 6 times lower than the SLR method, and 5 times lower than the BLR method.

## Conclusions

Three data-driven models were used to predict daily DO concentration in the Bow River, in Calgary, AB, Canada. The structure of each model was a first-order autoregressive model, *i.e.* a one-day lag value of mean daily DO was used to predict mean daily DO. The fuzzy linear regression model outperformed both the Bayesian and simple linear regression models. This is based on the models ability to reproduce the observed variability over the study period. The FLR model vastly improved the ability to predict low DO.

## Acknowledgements

The authors would like to acknowledge NSERC, the BC Ministry of Education and the University of Victoria for funding this research, and for the City of Calgary for providing the datasets.

## References

- Bárdossy, A., Bogardi, I., & Duckstein, L. (1990). Fuzzy regression in hydrology. *Water Resource Research*, 26(7), 1497-1508.
- Chang, Y.-H. O. & M Ayyub, B. (2001). Fuzzy regression methods—a comparative assessment. *Fuzzy sets and systems*, 119(2), 187-203, doi: 10.1016/S0165-0114(99)00091-3.
- Dubois, D., & Prade, H. (1997). Bayesian conditioning in possibility theory. *Fuzzy Sets and Systems*, 92(2), 223-240. doi:10.1016/S0165-0114(97)00172-3
- Dubois, D., Foulloy, L., Mauris, G. & Prade, H. (2004). Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10, 273-297, doi:10.1023/B:REOM.0000032115.22510.b5.
- Dubois, D., Prade, H. & Sandri, S. (1993). On possibility/probability transformations. In R. Lowen and M. Roubens, eds. *Fuzzy Logic*, Dordrecht, Netherlands: Kluwer Academic Publishers, 103-112.
- Environment Canada (2014). Real-time Hydrometric Data on the Wateroffice website, [http://www.wateroffice.ec.gc.ca/index\\_e.html](http://www.wateroffice.ec.gc.ca/index_e.html), visited 1 July 2014.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian data analysis. CRC press, Boca Raton, FL.
- Guyonnet, D., Bourguine, B., Dubois, D., Fargier, H., Come, B., & Chiles, J. -P. (2003). Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, 129(1), 68-78.
- Haan, C. T. (1977). Statistical Methods in Hydrology. The Iowa State University Press. Ames, IA.
- Han, J. C., Huang, G. H., Zhang, H., Li, Z., & Li, Y. P. (2014). Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of SLURP model applied to the Xiangxi River watershed, China. *Stochastic Environmental Research and Risk Assessment*, 28(4), 973-989.
- Hall, M. J. (1984). *Urban Hydrology*. Essex, England: Elsevier Applied Science.
- He, J., Chu, A., Ryan, M. C., Valeo, C., & Zaitlin, B. (2011). Abiotic influences on dissolved oxygen in a riverine environment. *Ecological Engineering*, 37(11), 1804– 1814. doi:10.1016/j.ecoleng.2011.06.022
- Huang, Y., Chen, X., Li, Y. P., Huang, G. H., & Liu, T. (2010). A fuzzy-based simulation method for modelling hydrological processes under uncertainty. *Hydrological Process*, 24(25), 3718-3732.
- Kahraman, C., Beşkese, A., & Bozbura, F. .. (2006). Fuzzy regression approaches and applications. In C. Kahraman (Ed.), *Fuzzy Applications in Industrial Engineering* (pp. 589-615). Berlin, Germany: Springer Berlin Heidelberg.
- Khan, U. T., & Valeo, C. (2014). A new fuzzy linear regression approach for dissolved oxygen prediction. *Hydrological Sciences Journal*, DOI:10.1080/02626667.2014.900558
- Kim, K. J., Moskowitz, H., & Koksalan, M. (1996). Fuzzy versus statistical linear regression. *European Journal of Operational Research*, 92(2), 417-434. doi:10.1016/0377-2217(94)00352-1
- Ozbek, M. M., & Pinder, G. F. (2006). Non-probabilistic uncertainty in subsurface hydrology and its applications: an overview. *Water, Air, & Soil Pollution: Focus*, 6(1-2), 35-46. doi:10.1007/s11267-005-9011-4
- Robinson, K., Valeo, C., Ryan, M.C., Chu, A. & Iwanyshyn, M. (2009). “Modelling Aquatic Vegetation and Dissolved Oxygen After A Flood Event in The Bow River Alberta, Canada.” *Canadian Journal of Civil Engineering*, 36:492-503.
- Solomantine, D. P., & Ostfeld, A. (2008). Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*, 10(1), 3-22. doi:10.2166/hydro.2008.015
- Zhang, K., & Achari, G. (2010). Correlations between uncertainty theories and their applications in uncertainty propagation. In H. Furuta, D. M. Frangopol, & M. .. Shinozuka (Eds.), *Safety, reliability and risk of structures, infrastructures and engineering systems* (pp. 1337 - 1344). London, UK: Taylor & Francis Group