Neuro-Fuzzy Logic Model for Breakup Forecasting at Fort McMurray, AB

Chandra Mahabir\textsuperscript{1,2}, Faye Hicks\textsuperscript{1} and Aminah Robinson Fayek\textsuperscript{1}

\textsuperscript{1}Dept. of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2W2, Faye.Hicks@ualberta.ca
\textsuperscript{2}River Engineering, Alberta Environment, Edmonton, AB, Canada, T5K 2J6, Chandra.Mahabir@gov.ab.ca

At many sites in Canada river breakup presents a source of flooding concern, as the transition from an ice covered river to open water can sometimes be dramatic. Water levels can rise sufficiently quickly to fracture and move the ice cover while the ice is still reasonably strong and an ice jam can form if this moving ice comes to a halt. The safety of residents and property along a river may be compromised by the rising water levels behind a jam or by the water associated with the release of a large ice jam.

Fort McMurray AB, is one such site and consequently an extensive database of relevant hydrological, hydraulic and meteorological variables has been created to support ice related research in the Athabasca River basin. Over one hundred variables have been investigated with multiple linear regression methods. Results indicate that several combinations of variables can be used to develop a model for maximum water level to be expected at breakup, but data from three seasons (fall, winter and spring) are needed to produce reasonably reliable forecasts. Therefore, only short term forecasts (only a few days warning) are possible with this model.

This paper reports on a more sophisticated non-linear analysis undertaken using fuzzy logic theory, a form of artificial intelligence modeling that can incorporate both expert knowledge and historical occurrences. Here, using a logic rule base derived using artificial neural networks, a breakup forecasting model is developed which provides comparable results using fall and winter data only, thus providing a long lead time forecast (several weeks warning) of potential maximum water levels at breakup.
Introduction

Each spring, numerous rivers across Canada are monitored closely, because of the potential risk to property and lives should an ice jam form during the river breakup process. River ice jams can produce rapid changes in water levels as water flow is impeded by an ice jam, or as water is suddenly released from storage behind an ice jam during release. Unlike open water flood events that are preceded by heavy rains or snowmelt, at present spring ice jams events have no single, generally identifiable, quantitatively predictable precursor. The lack of forecasting ability for such severe river conditions poses a serious threat to communities each year.

Fort McMurray, located in northeastern Alberta, is one example of a community where the river is actively monitored each spring to advise on the river ice breakup progress. The community has a history of river ice jams resulting in flooding. Severe floods occurred recently in 1977 and 1997, causing millions of dollars in damage. Minor flooding has also occurred several times in the last decade and although these have only resulted in minimal damage to residential and commercial properties, expenses are still incurred by governments monitoring and preparing for the potential of a much more severe event. The extensive monitoring over the past three decades has resulted in a significant volume of qualitative data, but not a rich wealth of quantitative data. Qualitative data exists in the form of photos, and written descriptions of the river breakup processes. The collection of quantitative data is limited by the inability to predict the formation of ice jams; their dynamic nature makes it extremely difficult to plan observations. Furthermore, critical reaches of this river are remote and relatively inaccessible. However, perhaps most important, directly measuring some of the key properties of an ice jam is logistically difficult, and in some cases impractical (e.g. porosity, internal strength, associated discharge, etc.). Consequently, many forecasting efforts have been focused towards probabilistic models (empirical and statistical techniques).

For ice jam flood forecasting in general, statistical and threshold models have been most frequently investigated. Threshold methods attempt to establish lower or upper limits that predict a positive or negative occurrence of a particular event, such as the formation of an ice jam. Multivariate threshold models have been applied with modest success (for example, Galbraith, 1981 and Wuebben et al., 1995). While threshold models provide a forecast of the potential for an ice jam event, they give no indication of the potential flood event that may accompany an ice jam. Single and multiple regression models have also been applied to forecast the maximum water level at breakup with moderate success by Shulickovski (1963), Beltaos (1984) and Robichaud (2003). White and Daly (2002) used stepwise selection of meteorological and hydrologic parameters to identify statistically significant input variables and then applied discriminant function analysis to predict ice jams. Massie et al. (2000) developed an artificial neural network to produce a daily forecast of jam/no jam that required 22 input variables.

As an alternative approach to statistical modeling, Mahabir et al. (2002) reported that fuzzy logic appeared promising for qualitative predictions of the potential severity of river ice breakup. Fuzzy logic is a type of modeling which is based on fuzzy set theory, rather than the crisp set theory of traditional mathematics. The approach allows for mathematical representation of linguistic terms and incorporation of subjective assessment.
The rule base is a critical component in the development of any fuzzy logic model. It contains the statements that dictate the mathematics that manipulate the input data sets from the membership functions into a resultant set. Two accepted approaches to rule base development involve using historical data or expert knowledge (or a combination of both) to define the cause and effect relationship. In this investigation, we explore the comparative value of each approach in the context of ice jam flood forecasting, and also compare the fuzzy logic modeling approach to more conventional multiple linear regression modeling. We use the Athabasca River at Fort McMurray, AB as the case study, due to its propensity for ice jam occurrence and the extensive database assembled for the site (Robichaud, 2003, Mahabir et al., 2004, 2005).

Description of Study Area

Figure 1 illustrates the Athabasca River Basin from its headwaters to the location of interest, immediately downstream from Fort McMurray. It is the largest, unregulated river in the province of Alberta, flowing eastward out of the Rocky Mountains. The Pembina river drains farmlands in the southern portion of the basin into the Athabasca. Further downstream, the river loops southwards towards to the town of Athabasca, before making a dramatic turn northwards at the town into a reach containing numerous rapids sections. About 140 km upstream of Fort McMurray, the Athabasca River the river turns east and becomes deeply entrenched and meandering, flowing through another series of rapids, before reaching Fort McMurray. At Fort McMurray, the river turns north, the slope reduces, and numerous sand bars and islands are distributed across the channel. These dramatic changes in the physical properties of the river in the vicinity of Fort McMurray, are responsible for the frequent formation of ice jams. The Clearwater River has its confluence with the Athabasca River immediately downstream of Fort McMurray.

Basin hydrology is a very relevant factor in river ice breakup processes. Because the Athabasca River flows northwards, river breakup occurs in the southern basin first and progresses northwards. While the headwaters of the basin remain snowbound until late May due to the mountainous topography, the mid basin can generate significant runoff resulting in a river ice breakup moving from upstream to downstream.

Because of the availability of historical data and the frequent occurrence of ice jams, the Athabasca River at Fort McMurray provides a good opportunity to investigate spring river ice jams. Social and economic impacts of river ice breakup have impacted Fort McMurray from its early years as a Hudson Bay Outpost and throughout its development and into the modern era as an oil boom municipality (Blench and Associates Ltd.,1964 and Alberta Environment, 1985). In response to the economic loss and the threat to human lives, the Alberta Government became more actively involved with this aspect of the Athabasca River. After it was realized that controlling ice jams on unregulated large rivers was not feasible by physical alteration of the river channel, the government established a monitoring and observation program with the ultimate goal of identifying important characteristics of the breakup process (Andres, 1980). Between 1977 and 1990, several groups were involved in observations and research regarding river ice on the Athabasca River including Alberta Environment, Alberta Transportation, Alberta Research Council and the Regional Municipality of Wood Buffalo. In 1997, an ice jam again
caused millions of dollars in damage and, again, there was little warning of the potential severity of river breakup.

Data and Preliminary Analyses

Although many groups have worked independently, Robichaud (2003) was the first to collect data from all sources, starting with the 1875 account and continuing through until 2001 to create a single database with a detailed description of documented qualitative descriptions, quantitative data and published research associated with the Athabasca River at Fort McMurray. While the database was extensive, there were still many processes which may be better represented by additional variable selections. For example, air temperature was only considered at Fort McMurray and not in the mid basin, where the majority of the spring runoff snowmelt originates. Therefore, for the purposes of this study, that database was extended further to include a total of 106 relevant variables. The types of variables that comprise the expanded database include:

- maximum water level that occurs at a specific location;
- fall and spring water levels water levels (prior to river freeze up, water levels after freeze up, water levels prior to spring runoff, water levels prior to ice movement);
- fall river flows;
- linear cumulative heat transfer variables (dependent on air temperature) and solar radiation;
- degree day indices (degree days of freeze and degree days of thaw);
- soil moisture indices (seasonal precipitation, groundwater levels);
- snow cover in the basin, (i.e. SWE) based on satellite and manual snow course data);
- climate indices (Pacific Decadal Oscillation); and
- ice thickness.

A more extensive description of the 106 variables contained in the database is provided by Mahabir et al. (2005).

Multiple linear regression (MLR) models for the database were explored by Mahabir et al. (2005), who found that the application of MLR analysis for river ice breakup models at this site is limited by the available data and multicolinearity between variables. Multicolinearity, or interdependence, must be carefully considered when regression analysis is applied because the influence of a particular variable can be hidden by another and can result in unstable predictive capabilities in the equation. Logical groupings of the variables have been established but because of the interrelationship of river ice breakup processes, it is not practical to simply consider variables by the process that they may influence. For example, solar radiation plays a large role in snowmelt processes, river ice decay processes and, to some extent, influences other meteorological variables.

To investigate the potential for developing long lead time forecasts with MLR models, Mahabir et al. (2005) grouped the data by season, and tested various model groups. It was found that no significant relationship for maximum spring breakup water levels could be obtained for the individual or combined fall and winter data sets. However, a model based on data from all three seasons (fall, winter, and spring) was successfully developed. This model, shown in Figure 2,
requires data for 8 variables (from all three seasons), has a $R^2_{adj} = 0.84$, and a standard error of 0.7 m. The relevant parameters were:

1. soil moisture
2. a measure of the intensity of winter cold
3. a measure of early spring runoff
4. SWE in the basin
5. intensity of cold weather immediately before breakup
6. intensity of the solar radiation in the mid-basin
7. rate of water level increase as measured below Fort McMurray prior to major ice movement
8. water level as measured below the town of Athabasca prior to spring runoff

Also shown in Figure 2 are two anomalous points which the regression software excluded in the development of the MLR model. The dashed lines in Figure 2 indicate ± 0.5 m from the line of perfect agreement, which represents the average accuracy to which breakup water levels can realistically be measured at this site. These lines indicate that the model predictions are relatively good, although the excluded points do represent a significant model limitation.

Mahabir et al. (2005) also demonstrated that there was no unique combination of variables for this site that would model the maximum water level during spring breakup. Specifically, they found that, in some cases, correlated variables could be substituted to create a new multiple linear regression equation that had comparable accuracy to the model presented on Figure 2.

**Overview of Fuzzy Logic Modelling**

Pioneered by Zadeh (1965), fuzzy logic has been effectively used in combination with other soft computing methods for predictive hydrological modeling, providing solutions to problems that are difficult to state empirically but which can be linguistically described. The basic components for fuzzy systems involve fuzzification of the input variables, application of a fuzzy operator, implication from an antecedent to the consequent, aggregation of the results from the evaluation of all the rules and potentially defuzzification (interpretation of resultant fuzzy set to a crisp or unique number).

The fuzzification of the input variables is the process by which a defined quantity (e.g. ‘40 mm of SWE in a snowpack) is redefined in terms of how representative it is of an idea (e.g. such as a “low” snowpack). The linguistic terms, such as “low”, are used to describe the input variables (like snowpack) as membership functions. Membership functions are logical linguistic descriptions of the input variables where each measured input quantity is evaluated by how well it is suited to the linguistic term. Figure 3 provides possible membership definitions for snowpack and illustrates that a single input value can belong to more than one membership function. The shape and number of the membership functions can be selected based on expert opinion, statistical distributions, cluster analysis or simply as logical groupings.

The selection of a fuzzy operator governs the interaction of the input variables, which are now expressed as data sets, with the familiar linguistic terms of “AND” or “OR”. For most
applications, the independence or interdependence of the input variables with respect to the outcome governs the selection of “AND” or “OR” respectively.

A rule base must be developed to define the implication from an antecedent combination of input variables to the consequent or outcome membership functions. These rules can be defined based on historical evaluations and/or based on expert knowledge. Rules follow the format of an if-then statement such that “If condition X AND/OR condition Y then condition Z”. Conditions X and Y describe the state of input variables, and condition Z describes the state of the output variables. The magnitude of the rule base can be a limiting factor in the development of fuzzy logic systems as the number of rules required increases substantially with the number of input variables. This is because a complete rule base has a rule defined for every combination of antecedent conditions. For example if a process is defined by two input variables each described by three linguistic terms, $3^2$ or 9 rules would constitute a complete rule base. However, if three variables are required to describe the process, then $3^3$ or 27 rules are required. Thus for complex problems involving a large number of causative factors (such as ice jam occurrence), the number of variable involved can necessitate an extremely large number of rules.

The sets of data from the evaluation of each rule are combined through the aggregation of the consequents across the rules. A common method used in fuzzy logic, and applied in this paper, is maximum. In terms of aggregation, maximum is a function that combines the maximum value attained by any rule evaluation into a single resultant set. Defuzzification is the process of evaluating the resultant set, often for the purpose of describing the result as a single crisp value. Features of the resultant set, such as the centroid or the maximum membership value, are used in the defuzzification process to produce the optimum description of the result.

**Fuzzy Logic Model Development for Athabasca River Breakup**

Mahabir et al. (2002) explored the applicability of fuzzy logic for obtaining long lead time qualitative predictions of the potential severity of river ice breakup. However, the question remains as to whether accurate quantitative forecasts could be achieved with a fuzzy logic model. Thus, the main focus of this investigation was to examine alternatives for the development of the fuzzy logic rule base, and to compare the resultant models to the MLR model developed by Mahabir et al. (2005). As discussed earlier, two rule based development approaches were compared: one based on expert knowledge and another based on historical occurrences.

**Rule Base Development based on Expert Knowledge**

The application of expert knowledge has long been accepted as one of the most advantageous features of fuzzy logic. Expert knowledge relies not on the historical occurrences, but on the ability of an expert to define the interaction of the variables to each other and to the outcome. In this approach, knowledge about the process to be modeled is extracted either by conscious development of the expert system or through less direct means. For example, an individual or group of experts may be asked to evaluate statements to be used directly in a fuzzy rule base, or this knowledge may be acquired by means of surveys particularly if a range of opinions is expected from a group with specialized knowledge on the subject.
One of the major criticisms of developing the rule base exclusively with expert knowledge is that the process is subjective in that interpretations may vary depending on the resource(s) from which the knowledge was acquired. Therefore, the rule base may be context specific and thus difficult to re-create. Another obvious disadvantage is that this approach practically limits the number of input variables that can reasonably be considered, given the exponential growth of the required rule base with increasing numbers of input variables. For example, the multiple linear regression analysis indicated that eight variables were relevant for breakup water level forecasting at Fort McMurray. If each of these variables employed three membership functions (e.g. defining ‘low’, ‘medium’, and ‘high’), then $3^8$ or 6561 rules would be required. Even if the number of input variables was reduced to four, the required number of rules would still be quite large ($3^4$ or 81).

In this case, an expert knowledge based fuzzy logic system was developed to predict the maximum water level at breakup on the Athabasca River at Fort McMurray, AB, using three input variables specifically: the amount of snow in the basin (SWE) on April 1, the ice thickness at Fort McMurray and the soil moisture in the basin prior to winter. Each of the three was defined with three linguistic terms (low, medium and high) thus requiring $3^3$ or 27 rules. These rules related all possible combinations of the input variables to three possible outcomes of the output variable, specifically a low, medium or high maximum water level at breakup. A complete rule base was then generated with all rules having equal strength. Expert knowledge combined with minimal site specific information (five years of record) was used to define the outcome of each input combination in the rule base. The resulting model is similar to the one developed earlier by Mahabir et al. (2002); however, it is different in that it is based on the extended database (Mahabir et al., 2005) and more consistent the ice thickness data (Mahabir et al., 2004).

The resulting fuzzy expert system was found to be able to correctly classify the maximum water levels during spring breakup as ‘low’, ‘average’ or ‘high’ which was a promising finding, given that all of the variables used as input to the model are known several weeks in advance of breakup. This indicates that the model is valuable for qualitative long term forecasting. However, when the resultant fuzzy set was defuzzified to crisp numbers, they were not found to be sufficiently accurate for water level forecasting purposes. Figure 4 illustrates these quantitative results, comparing this model’s results to those for the MLR model (Mahabir et al., 2005).

**Rule Base Development based on Historical Data**

The above results suggested that a fuzzy expert system built on a rule base developed using historical data might provide superior results, as a larger number of variables could potentially be considered if the rule base development could be automated. Furthermore, by relying completely on past occurrences to develop the rule base, the problem of subjectivity could be removed. Neuro-fuzzy training is one method that can be used to automate the generation of the rule base, incorporating past occurrences. The name neuro-fuzzy represents a combination of artificial neural network (ANN) and fuzzy logic modeling. Although this approach eliminates the subjectivity associated with developing a rule base based on expert knowledge alone, one key
potential disadvantage of ANN modeling is that it is essentially a ‘black box’ approach, and therefore the results may not necessarily be physically meaningful.

To test the neuro-fuzzy approach, an attempt was first made to develop a fuzzy logic model based on the same eight input variables employed in the MLR model. Each of these had three membership functions, thus requiring the development of $3^8$ or 6561 rule. As it turned out, even with the assistance of ANNs, resolving the database was not possible because of the limited available data (in terms of years of record). Therefore, in order to apply fuzzy analysis, the number of variables was reduced from 8 to 4, selecting the variables with the longest lead time (those available in the fall or late winter). This reduced the required number of rules to 81, and also facilitated the development of a long lead time forecasting model. A further advantage was that the reduction in the number of input variables effectively increased the period of record, since not all data were available in all years (a common scenario in ice jam databases).

The four variables evaluated in the neuro-fuzzy model were then: the soil moisture (as indicated by the cumulative summer precipitation at Fort McMurray); a measure of the intensity of winter cold (as indicated by the cumulative degree days of freeze during the winter), a measure of early spring runoff (specifically the change in groundwater levels from January to March), and snowpack in the basin at the beginning of April (SWE measured from passive microwave satellite data for the entire basin).

The strengths of the rules in the rule base were determined through neuro-fuzzy training, which consists of multiple iterations on the historical record, with incremental adjustments of the strength of each rule. The rule training was done with unsupervised learning using random presentation of the historical occurrences. Adjustments were required for the initial neuro-fuzzy model to achieve the desired level of performance since initially, the neuro-fuzzy model was unable to resolve high or low water level occurrences. Through trial and error, it was determined that by adding more definition to the output membership function, the ANN was better able to resolve extreme years. Specifically, the membership functions for the output results were increased from three to five linguistic terms by using “very low” and “very high”, in addition to the low, medium and high categories used in the expert knowledge based model.

Figure 4 illustrates the quantitative results for this model, as compared to the MLR model developed by Mahabir et al. (2005). Even with the reduced number of input variables, the neuro-fuzzy model is clearly able to model the water levels at breakup comparably well. In particular, at higher water levels (those of most interest in the context of flood risk) nearly all modeled values are within the ± 0.5 m band on the graph. Furthermore, the neuro-fuzzy model was able to consider all relevant years of record, including the two years excluded by the MLR modeling software. Perhaps most significant, this model is based entirely on data available several weeks in advance of breakup.

Conclusions and Recommendations

In this investigation, fuzzy logic based models for river ice breakup water level forecasting were developed using two approaches for rule base development: expert knowledge and ANN analysis of historical data. The results of this investigation clearly show the advantage of using
ANN’s to develop the rule base in a fuzzy expert system, as the resulting model was not only significantly superior to the model based on the expert knowledge rule base, it was also competitive with the MLR model incorporating twice as many variables. Furthermore, unlike the MLR model which required data from three seasons (including data in the few days preceding breakup), the neuro-fuzzy model was based on fall and winter data only, thus facilitating a long lead time forecast of approximately 3 weeks.

These results also suggest that combinations of expert knowledge and historical occurrences should probably be used to create the rule base. If experience is limited for a particular circumstance, there may be historical data available to guide the expert in developing or confirming a particular rule. Conversely, the rules in a neuro-fuzzy model may be evaluated by an expert to extend the models performance, as the model can only reproduce what has been presented to it for training. For example, experts can often define intermediate or extreme rules where no historical examples exist. At the least, expert knowledge should be employed to scrutinize the ANN rules base for deduced rules which are obviously physically unrealistic.

Acknowledgments
This research was supported in part through a grant to the second author from the National Sciences and Engineering Research Council of Canada through the MAGS Research Network. This support is gratefully acknowledged. The authors would also like to acknowledge the assistance of Alberta Environment, particularly Claudine Robichaud and Larry Garner, and the other “ice” graduate students at the University of Alberta.

References


Mahabir, C., F.E. Hicks and A. Robinson Fayek. 2005. “Multiple Linear Regression Applied to Ice Jam Flood Forecasting at Fort McMurray, AB.” prepared for submission to the Canadian Journal of Civil Engineering.


Mountain and foothill areas remain frozen until late spring.

Runoff from farmlands in the southern portion of the basin can contribute significantly.

Limited contributions in the immediate vicinity due to slower hydrological response.

Figure 1. The Athabasca River basin upstream of WSC Gauge Station 07DA001.
Figure 2. Results of multiple linear regression model of maximum breakup water level for the Athabasca River at Fort McMurray, AB (adapted from Mahabir et al., 2005). Dashed lines depict ± 0.5 m from the line of perfect agreement.

Figure 3. Example of membership functions describing quantitative snowpack (20 mm) in the linguistic terms of low (0.24 membership), average (0.76 membership) and high (0.00 membership)
Figure 4. Comparison of results from multiple linear regression and fuzzy logic models based on expert knowledge.

Figure 5. Comparison of results from multiple linear regression and neuro-fuzzy models.