Comparison of different time delay embedding strategies for urban water demand (UWD) forecasting using machine learning techniques

John Quilty\textsuperscript{1}, Jan Adamowski\textsuperscript{1}, Bahaa Khalil\textsuperscript{1}, & Maheswaran Rathinasamy\textsuperscript{2}

\textsuperscript{1}McGill University
Department of Bioresource Engineering

\textsuperscript{2}Indian Institute of Technology Delhi
Department of Civil Engineering
Introduction

- Urban water demand (UWD) forecasting
  - **Optimizing** system performance [Adamowski, 2008];
  - Implement water use restrictions;
  - **Least-cost** infrastructure expansion strategy [Tiwari and Adamowski, 2015]; and
  - Provide **risk assessment** for the water supply system [Yung et al., 2011].

- UWD is a **nonlinear process** [House-Peters and Chang, 2011] requiring **nonlinear modeling tools** such as:
  - Artificial Neural Networks (**ANN**);
  - Support Vector Regression (**SVR**); and
  - Fuzzy Logic (**FL**).
Introduction

- **ANN**: many hyper-parameters; iterative calibration; local minima.
- **SVR**: very sensitive to hyper-parameter settings; longer development times (than ANN); **global solution**.
- Extreme Learning Machine (ELM) [Huang et al., 2006]:
  - Same configuration as ANN but **much faster**;
  - **Global solution**;
  - **Better or comparable performance** to ANN and SVR [Huang et al., 2012];
  - Only **three simple steps**: design hidden layer, randomization, inversion of hidden layer to obtain solution.
Problem Definition

- Given an UWD time series: $x(t)$ for $t = 1, 2, \ldots N$.
- Predict the next value $x(t + 1)$ using nonlinear methods and historical records, then estimate the importance of the historical records considered in the forecast:

$$[1] \quad x(t + 1) = f([x(t), x(t - \tau), \ldots, x(t - (m - 1) \tau)]); \rightarrow \text{time lag space}$$

$$[2] \quad \hat{x}(t + 1) = \sum_{k=1}^{L} b_k \text{sig} \left( a_{k0} + \sum_{j=1}^{m} a_{kj} x_{i-(j-1)\tau} \right); \rightarrow \text{ELM}$$

$$[3] \quad \hat{S}((j-1)\tau) = \frac{1}{N-(j-1)\tau} \sum_{i=(j-1)\tau+1}^{N} \left| \frac{\partial \hat{x}_{i+1}}{\partial x_{i-(j-1)\tau}} \right|; \rightarrow \text{ELM output sensitivity to time lag } j$$

$$[4] \quad \frac{\partial \hat{x}_{i+1}}{\partial x_{i-(j-1)\tau}} = \sum_{k=1}^{L} a_{kj} b_k \text{sig} \left( a_{k0} + \sum_{c=1}^{m} a_{kc} x_{i-(c-1)\tau} \right) \left( 1 - \text{sig} \left( a_{k0} + \sum_{c=1}^{m} a_{kc} x_{i-(c-1)\tau} \right) \right)$$
Objectives

- The main goal of this study is to compare two different nonlinear methods (based on chaos theory [Takens, 1981]) for choosing which historical records to include in a forecast for a given UWD time series and to forecast the process at one-step ahead. The importance of each historical record considered in the forecast can then be evaluated via model-based and model-free approaches.
  - Entropy Ratio (ER) [Gautama et al., 2003];
  - Local Constant Modeling (LCM) [Small and Tse, 2004];
  - ELM;
  - Conditional Mutual Information (CMI) [Cover and Thomas, 1991].
Methodology

- Three daily UWD signals from **Canadian** water utilities (**Montreal (M)**, **Toronto (T)**, and **Victoria (V)**) were collected for this study;
- Each signal was determined to be **chaotic** through the largest Lyapunov exponent method [Wolf et al., 1985; Kodba et al., 2005];
- Each signal was then transformed to its time lag space via **ER** and **LCM** methods;
- For each signal the time lag space from each method (ER and LCM) were used as inputs to ELM (creating **ER-ELM** and **LCM-ELM**) to derive a one-step ahead prediction.
Methodology

- The final 365 records were used for validation;
- Each ELM considered up to 150-250 sigmoid hidden layer neurons (activation functions);
- The one-step ahead predictions (ER-ELM and LCM-ELM) were then assessed by measures of:
  - Precision (root mean square error (RMSE)); and
  - Efficiency (Nash-Sutcliffe Efficiency Index (EI)).
- Model-based (ELM output sensitivity) and model-free (CMI) approaches are used to quantify time lag importance.
Results (time delay embedding)

<table>
<thead>
<tr>
<th>Time Series</th>
<th>Optimal Time Delay Embedding Parameters</th>
<th># Hidden Neurons</th>
<th>Performance Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LCM*</td>
<td>ER</td>
<td>LCM</td>
</tr>
<tr>
<td>Montreal (M)</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Toronto (T)</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Victoria (V)</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td></td>
<td>106</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

* Time delay is fixed at 1 for the LCM method
Results (Victoria time series plot)

- Total UWD (ML/D)
  - Date
    - 1985-01-01
    - 1990-06-24
    - 1995-12-15
    - 2001-06-06
    - 2006-11-27
    - 2012-05-19

- Total UWD (ML/D)
  - Date
    - 2011-10-25
    - 2012-01-03
    - 2012-03-13
    - 2012-05-22
    - 2012-07-31
    - 2012-10-09

- Observed, LCM-ELM, ER-ELM
Results (Toronto time series plot)

![Graph showing time series data for Total UWD (ML/D) with dates from 2005-01-01 to 2014-12-05 and 2014-03-11 to 2015-02-24. The graph includes observed data and LCM-ELM and ER-ELM model predictions.]
Results (Montreal time series plots)

Average UWD (ML/D)

Date


Average UWD (ML/D)

Date

2009-08-07 2009-10-16 2009-12-25 2010-03-05 2010-05-14 2010-07-23

- Observed
- LCM-ELM
- ER-ELM
Results (Montreal: time lag importance)
Conclusions & Recommendations

- **LCM-ELM outperformed ER-ELM** in terms of precision (RMSE) and efficiency (EI) for each time series (e.g. V time series → EI scores: 0.957 (LCM) vs. 0.830 (ER));

- ELM based time lag importance measures **confirmed via model-free approach**;

- LCM provided **higher dimensional** lag spaces compared to ER; and

- **LCM-ELM models** contained a **larger number of parameters** than ER-ELM models (e.g. T time series → # Hidden Neurons: 91 (LCM) vs. 15 (ER)).
Conclusions & Recommendations (Maximum Monthly [Average] UWD Profile for Montreal)

- **LCM-ELM** more accurately captured the **extremes** of each time series (e.g. Montreal).

![Graph showing average UWD for different months]

- **Month**: January, February, March, April, May, June, July, August, September, October, November, December
- **Average UWD (ML/D)**: 1700 to 2300
- **Lines**:
  - **LCM-ELM**
  - **ER-ELM**
  - **Observed**
Conclusions & Recommendations

• Create **bootstrap** based versions of **ER-ELM** and **LCM-ELM** to provide **confidence intervals** for predictions and time lag sensitivities;

• Utilize **time-frequency localized algorithms** (e.g. wavelet transforms or empirical mode decomposition) to **improve** overall **forecast accuracy** and to improve **prediction of outliers**;

• Bootstrap based ER-ELM and LCM-ELM can be further **improved by coupling with time-frequency localized algorithms**; and

• Future studies should consider investigating a wider array of urban water supply system time series (e.g. **reservoirs**, transmission mains, **automated metering infrastructure**) to **decipher the best approach** (ER-ELM or LCM-ELM) to use in general chaotic water resources time series forecasting.
THANK YOU!
References


References


