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

John Quilty[1], Jan Adamowski[1], Bahaa Khalil[1], & Maheswaran Rathinasamy[2]

[1]McGill University

Department of Bioresource Engineering

[2]Indian Institute of Technology Delhi

Department of Civil Engineering

Background source: http://aqualegion.com/wp-content/uploads/istock_000003810943medium1-1024x653.jpg

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] $x(t+1) = f([x(t), x(t-\tau), \ldots, x(t-(m-1))\tau])$; → **time lag space**

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

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

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

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

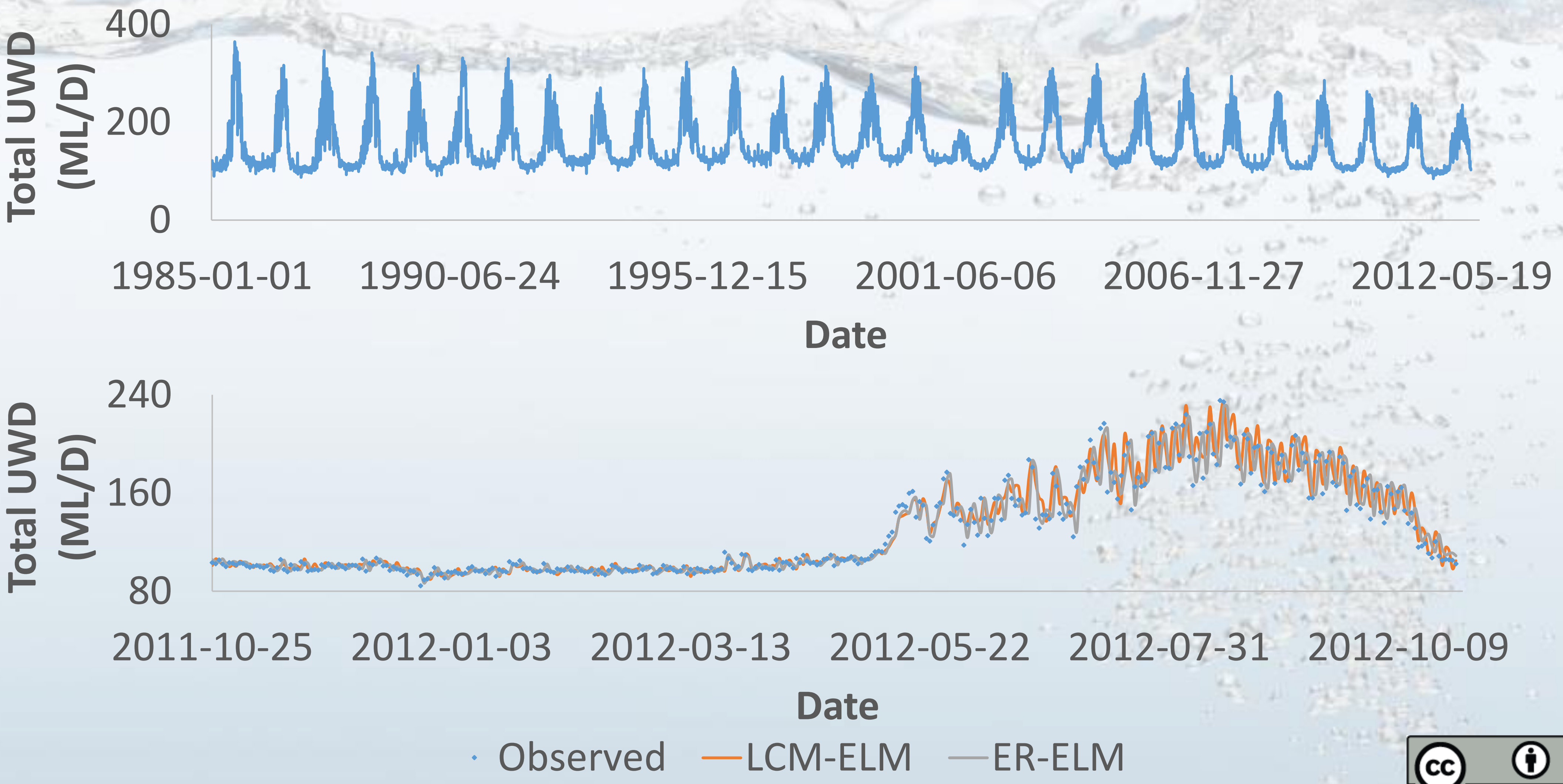
- The final **365 records** were used for **validation**;
- Each **ELM** considered up to **150-250 sigmoid** hidden layer neurons (**activation functions**);
- The one-step ahead predictions (**ER-ELM** and **LCM-ELM**) were then assessed by measures of:
  - **Precision** (root mean square error (**RMSE**)); 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)

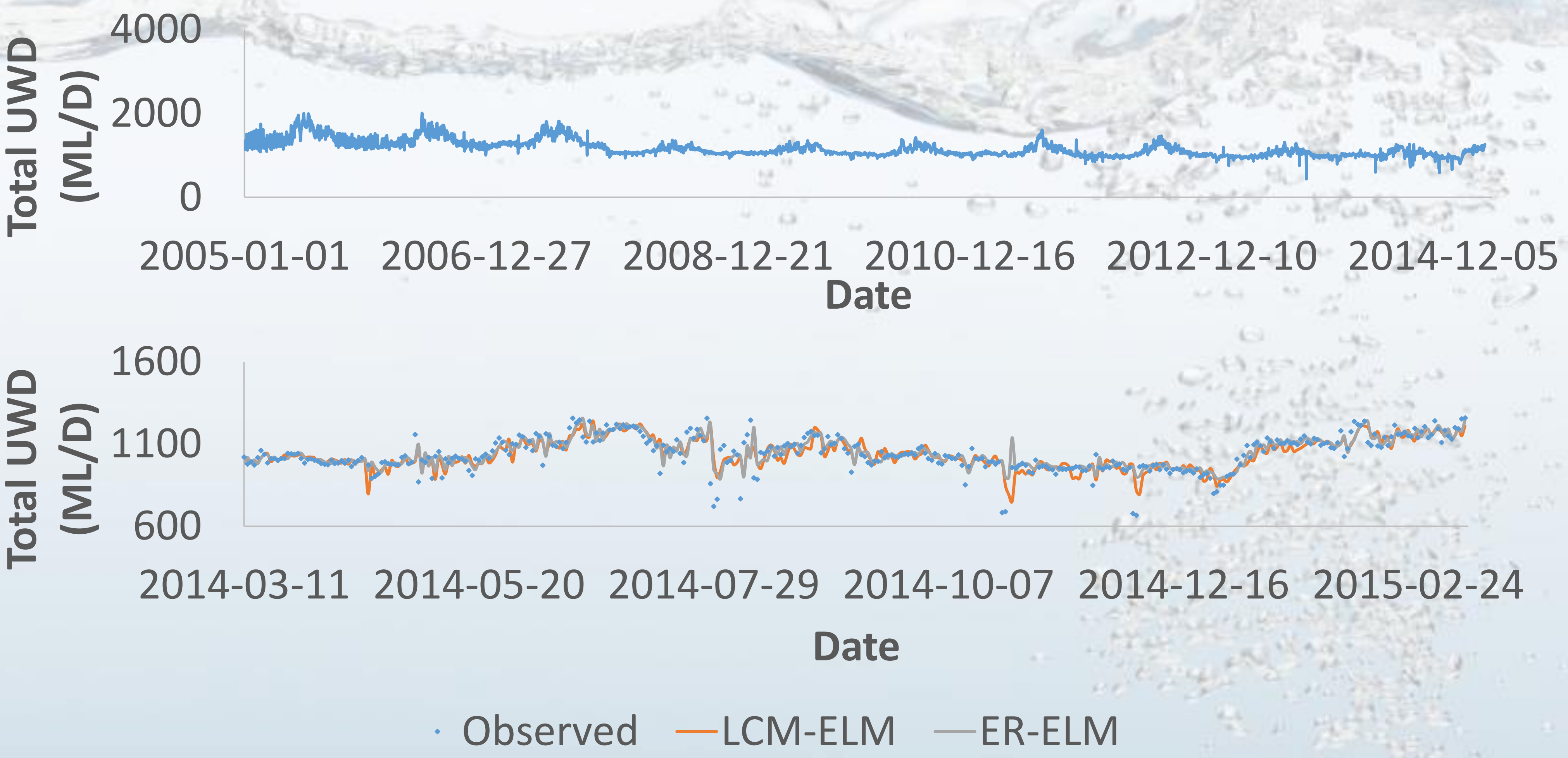
	Optimal Time Delay Embedding Parameters			# Hidden Neurons		Performance Evaluation			
	LCM*	ER		LCM	ER	LCM	ER	LCM	ER
Time Series	**m**	**m**	**τ**			**RMSE (ML/D)**		**EI**	
Montreal (M)	14	4	7	83	4	30.265	41.846	0.911	0.830
Toronto (T)	45	2	2	91	15	72.526	72.939	0.550	0.545
Victoria (V)	106	4	7	240	6	7.899	13.577	0.957	0.872

* Time delay is fixed at 1 for the LCM method

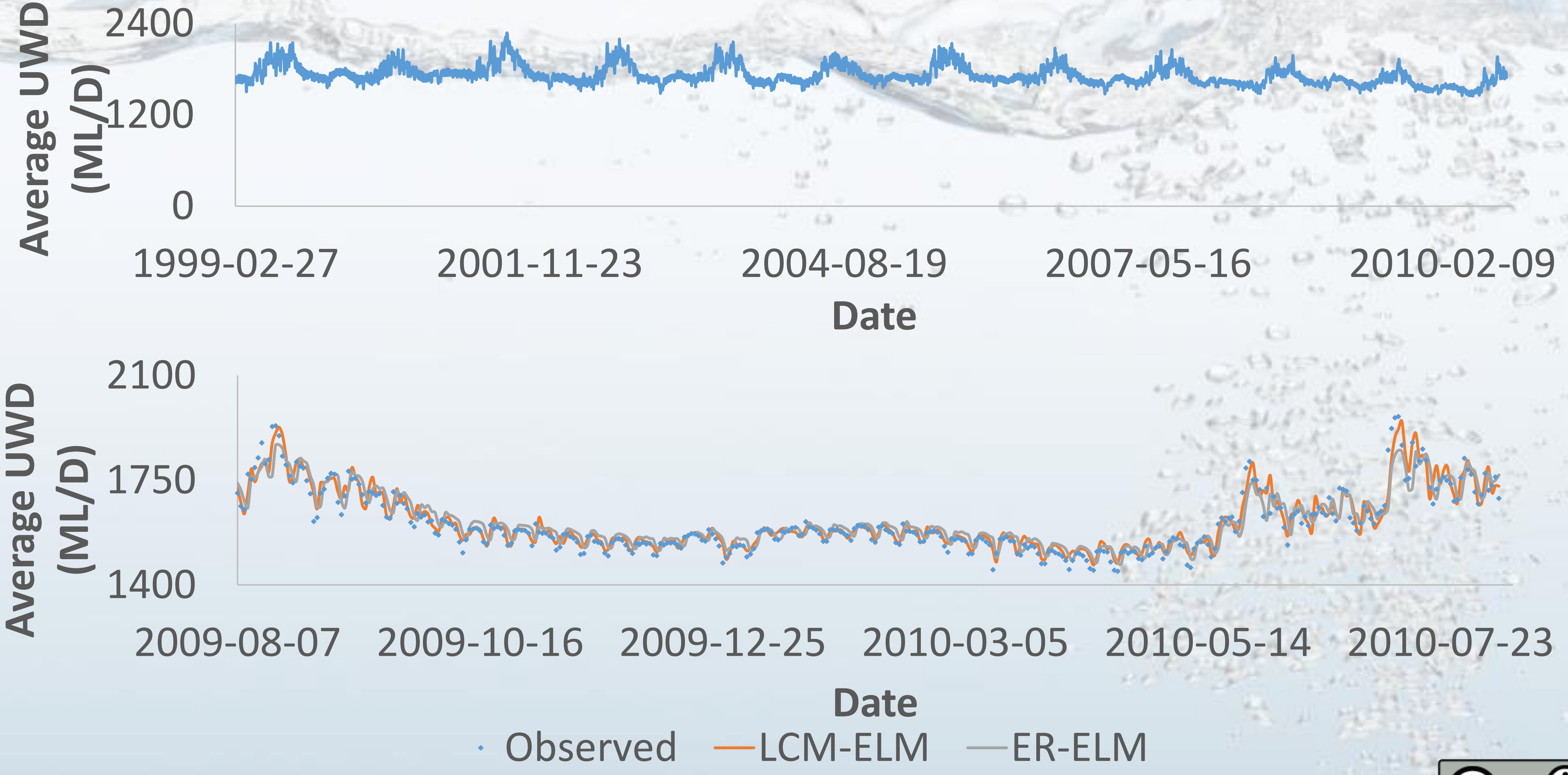
Results (Victoria time series plot)



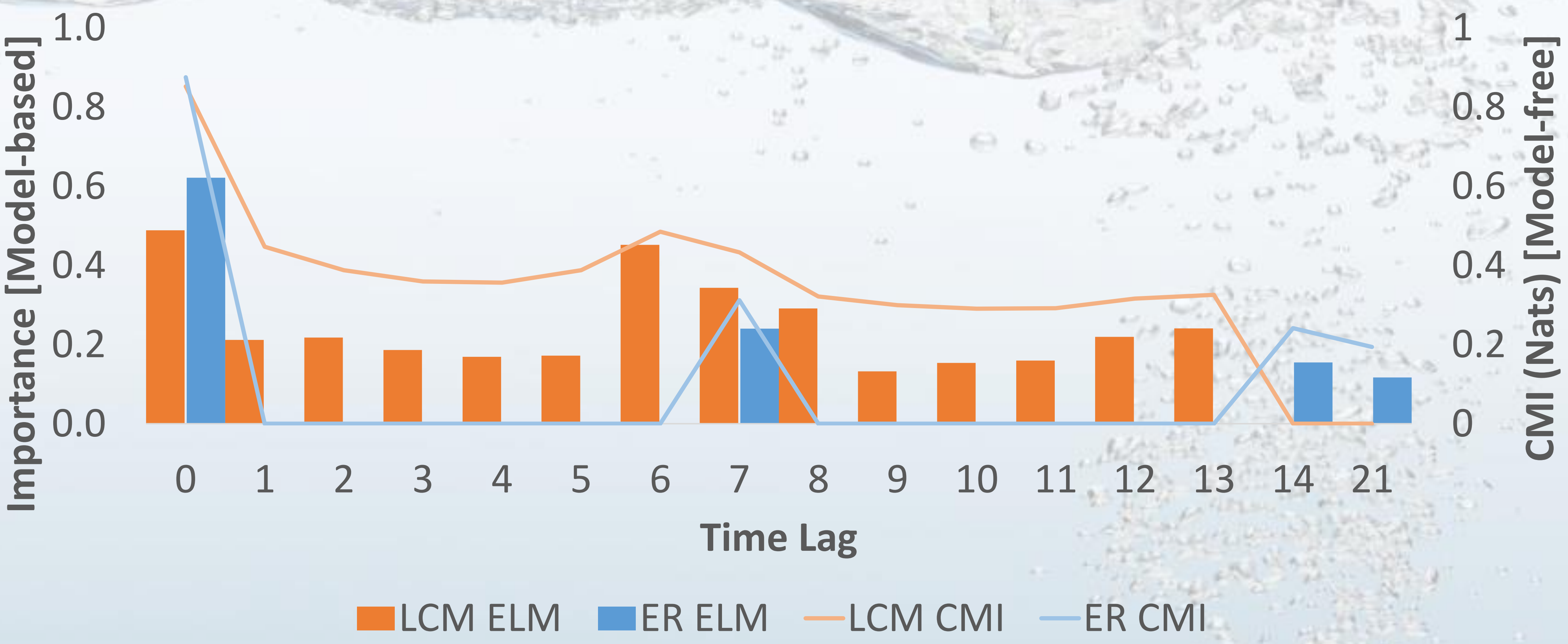
Results (Toronto time series plot)



Results (Montreal time series plots)



Results (Montreal: time lag importance)

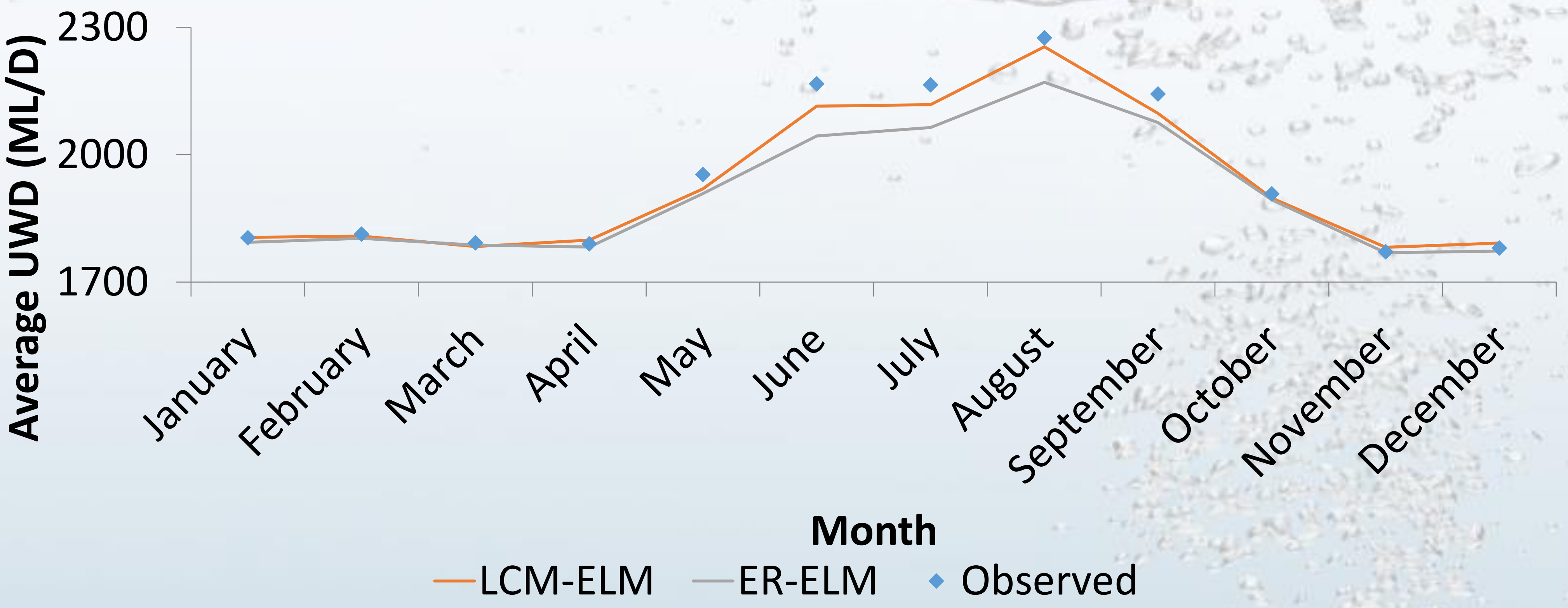


Conclusions & Recommendations

- **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).



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

Adamowski, J. F., 2008. Peak daily water demand forecast modeling using Artificial Neural Networks. *J. Water Resour. Plann. Manage.,* 134(2), 119-128.

Cover, T. M. and Thomas, J. A., 1991. *Elements of Information Theory.* Online ISBN 0-471-20061-1: John Wiley and Sons, Inc.

Gautama, T., D. Mandic, and M. M Van Hulle (2003a), A differential entropy based method for determining the optimal embedding parameters of a signal, in *Proceedings of the ICASSP-2003,* International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, China, 29–32.

House-Peters, L. A. and H. Chang (2011), Urban water demand modeling: Review of concepts, methods, and organizing principles, *Water Resour. Res.,* 47, W05401.

Huang, G.-B., Q.-Y. Zhu, and C,-K. Siew (2006), Extreme learning machine: Theory and applications, *Neurocomputing,* 70(1-3), 489-501.

Huang, G.-B., H. Zhou, X. Ding, and R. Zhang (2012), Extreme learning machine for regression and multiclass classification, *IEEE Trans. Syst. Man Cybern. Part B Cybern.,* 42(2), 513-529.

Kodba, S., M. Perc, and M. Marhl (2005), Detecting chaos from time series, *Eur. J. Phys.,* 26(1), 205-215.

References

Small, M. and C.K. Tse (2004), Optimal embedding parameters: a modelling paradigm, *Physica D,* 194, 283-296.

Takens, F. (1981), Detecting strange attractors in turbulence, In: D. Rand and L. Young, eds. *Dynamical Systems and Turbulence,* New York: Springer, 366-381.

Tiwari, M. and J. Adamowski (2015), Medium-term urban water demand forecasting with limited data using an ensemble wavelet–bootstrap machine-learning approach, *J. Water Resour. Plann. Manage.*, 141(2), 04014053.

Wolf, A., J.B. Swift, H.L. Swinney, and J.A. Vastano (1985), Determining Lyapunov Exponents from a time series, *Physica D*, 16, 285-317.

Yung, B. B., B.A. Tolson, and D.H. Burn (2011), Risk assessment of a water supply system under climate variability: a stochastic approach. *Can. J. Civ. Eng.,* 38. 252-262.