Machine learning based urban water demand (UWD) forecasts incorporating stochastic weather inputs

John Quilty & Jan Adamowski

McGill University

Department of Bioresource Engineering

Background source: http://www.clevelandwater.com/sites/default/files/your-water_distribution.png

Introduction

- Urban water demand (UWD):
  - **Nonlinear** process [Adamowski et al., 2012];
  - Coupled **human-water-environment** system;
  - Function of **previous** demands, **climate**, socio-economic fluctuations, etc. [House-Peters and Chang, 2011]; and
  - **Outdoor water use** very **important** for system performance (**risk**, **resiliency**, **vulnerability**)…learning about **feedbacks** between humans, water, and the environment…generally **dominated by recent weather**.
- Models/forecasts of water resources time series may be generated using:
  - Deterministic; stochastic; and/or **quasi-stochastic** methods [Chow, 1978].

Introduction

- Opportunity to develop a **quasi-stochastic** model for UWD forecasting (**QS-UWDF**) using a new (**open-source**) stochastic weather generation tool (Chen et al., 2010)*:
  - **Stochastic** weather generation + **historical** UWD used for **input**;
  - **Multiple runs** generated (i.e. to create **ensemble members**);
  - Each run is used to develop a **deterministic machine learning** model; and
  - Each prediction is **combined** in an **ensemble forecast system**.
- **Earlier** stochastic based UWD forecasting approaches:
  - Coupled ANN with either stochastic weather generation (**similar to this presentation**) or GCM projected changes [Yung et al., 2011]; and
  - ARMA applied after removing dominant periodicities [Mamo et al., 2013].

* http://www.mathworks.com/matlabcentral/fileexchange/29136-stochastic-weather-generator--weagets-

Problem Definition

- Given a set of daily UWD records and historical daily weather measurements develop a ***quasi-stochastic*** UWD forecast model (**QS-UWDF**); and
- **Forecast** UWD using the **QS-UWDF** model for the **next 3 days ahead** and **compare performance** against a fully **deterministic** model (i.e. only historical UWD and weather inputs) to asses efficacy of use.

Objectives

- The main goal of this study is to determine if a machine learning based **quasi-stochastic** approach is suitable for short-term (daily) UWD forecasting during outdoor water-use periods by referencing its performance to a **deterministic** model.
- To accomplish our objective we **incrementally combine** the following methods/tools to **produce forecasts** using our proposed **QS-UWDF** model:
  - Input variable selection (**IVS**);
  - Weather Generator of Ecole de Technologie Superiere (**WeaGETS**) [Chen et al., 2010];
  - Support Vector Regression (**SVR**); and
  - Bayesian Model Averaging (**BMA**).

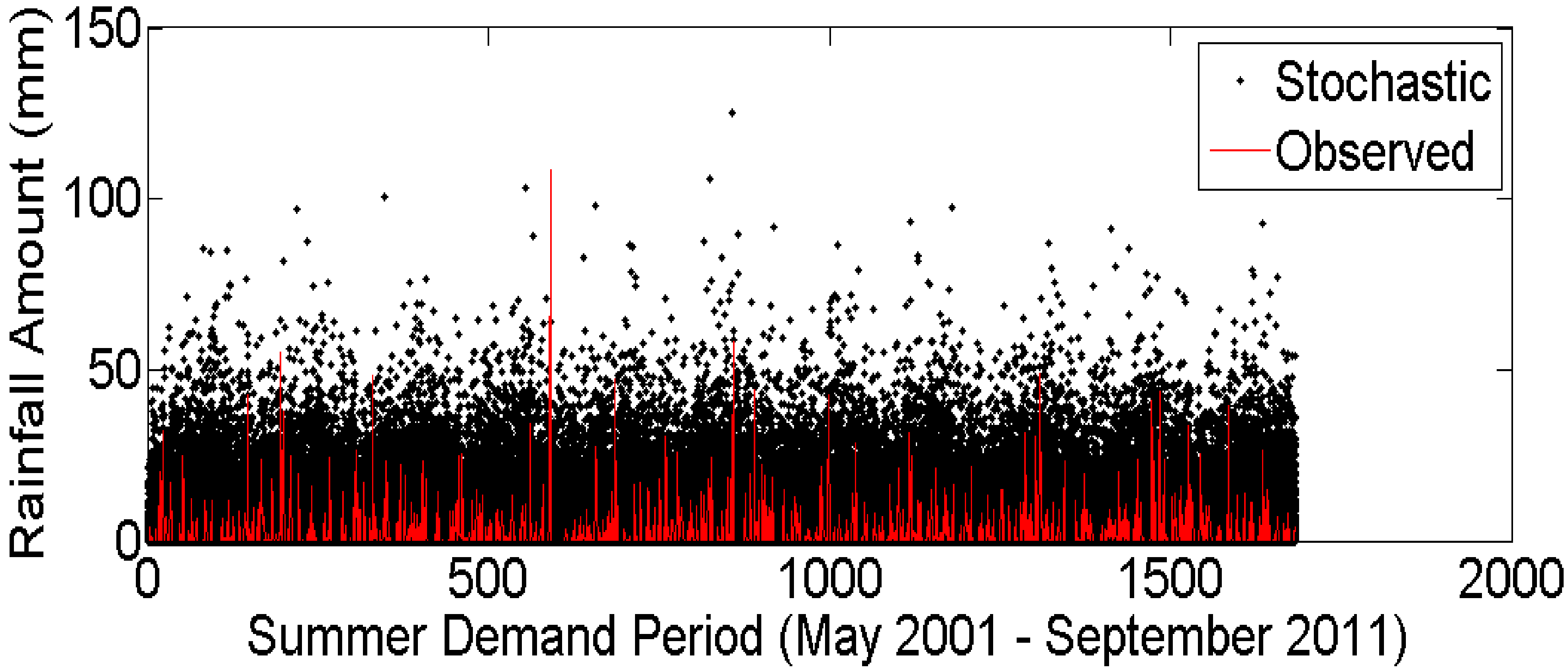
Methodology

- Study site → **Ottawa, Canada**
  - Daily **UWD records** (2001-2011); and
  - Historical daily **weather measurements** (1890-2011):
    - Max and min air temperature and;
    - **Rainfall depth**.
- Cross-correlation analysis **reveals significant dependencies** between UWD and weather up to a ~ **21 day time delay**:
  - **Each time series** (historical UWD and weather inputs) were time delayed **up to 21 days**;
  - Model **inputs**:
    - **Deterministic → Historical** weather measurements **only available up to 1 day time delay** (i.e. only previous information is used);
    - **QS-UWDF → Stochastic** weather inputs were **time lagged up to day of forecast** (i.e. weather on the day of the forecast is considered in the model **in addition to previous 21 days**).

Methodology

- WeaGETS:
  - Rainfall **occurrence**→ first-order Markov model;
  - Rainfall **amount** → mixed exponential distribution; and
  - Max and min **air temperatures** → first-order Markov model (conditioned on wet/dry status).
- SVR → Least- Squares SVR (LSSVR) optimized via PRESS (predicted residual sum of squares) [Cawley and Talbot, 2004]:
  - IVS → Input variables determined via Conditional Mutual Information (**CMI**).
- BMA → Individual LSSVR models combined to provide confidence intervals over ensemble prediction;
- The **2011 summer demand** period (153 records) used for **validating forecasts**; and
- **Forecast quality** determined via: Mean Absolute Error (**MAE**) and Correlation Coefficient (**CC**).

Results (weather generation)



Results (selected input variables)

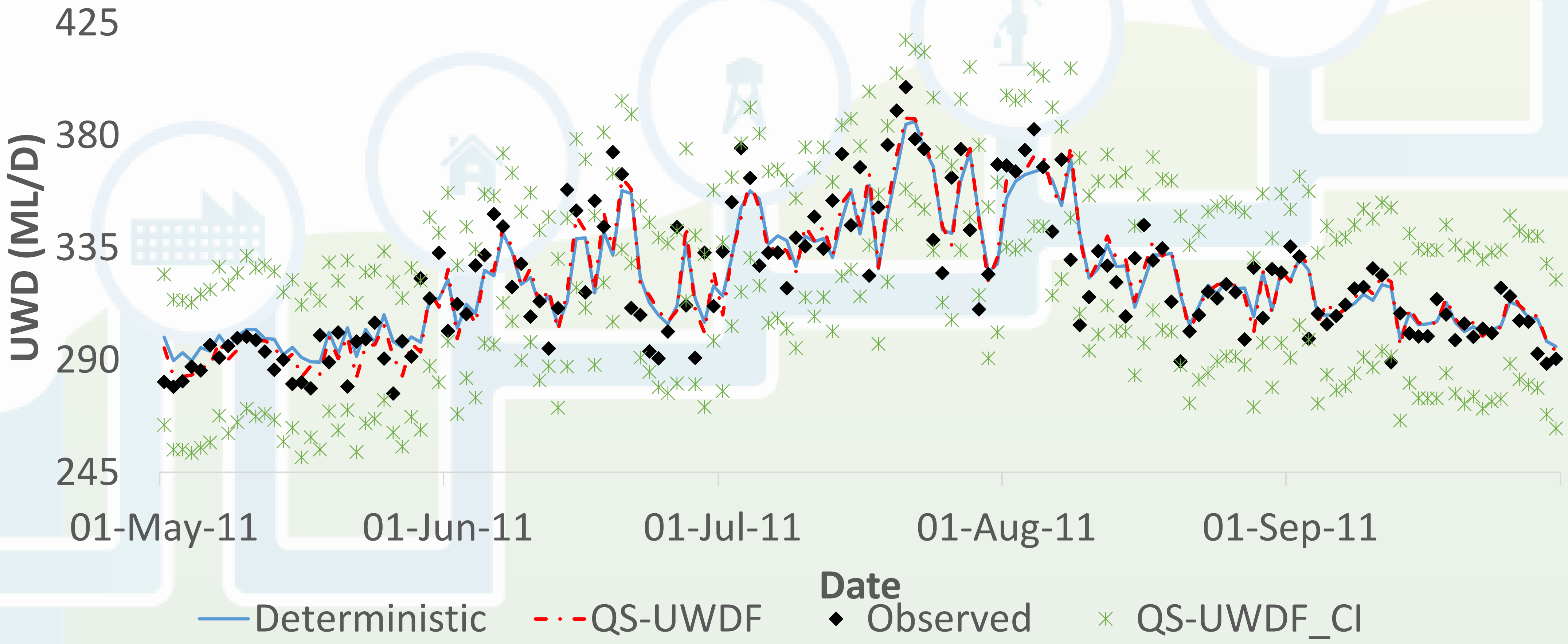
Input Variables (20 in total)

UWD (t)	UWD (t-1)	rain (t-1)	UWD (t-2)	rain (t-2)	UWD (t-4)	rain (t-4)	rain (t-5)	UWD (t-6)	rain (t-7)
rain (t-8)	rain (t-9)	rain (t-11)	rain (t-12)	rain (t-13)	rain (t-14)	rain (t-15)	UWD (t-17)	rain (t-17)	rain (t-19)

Results (short-term UWD forecasts)

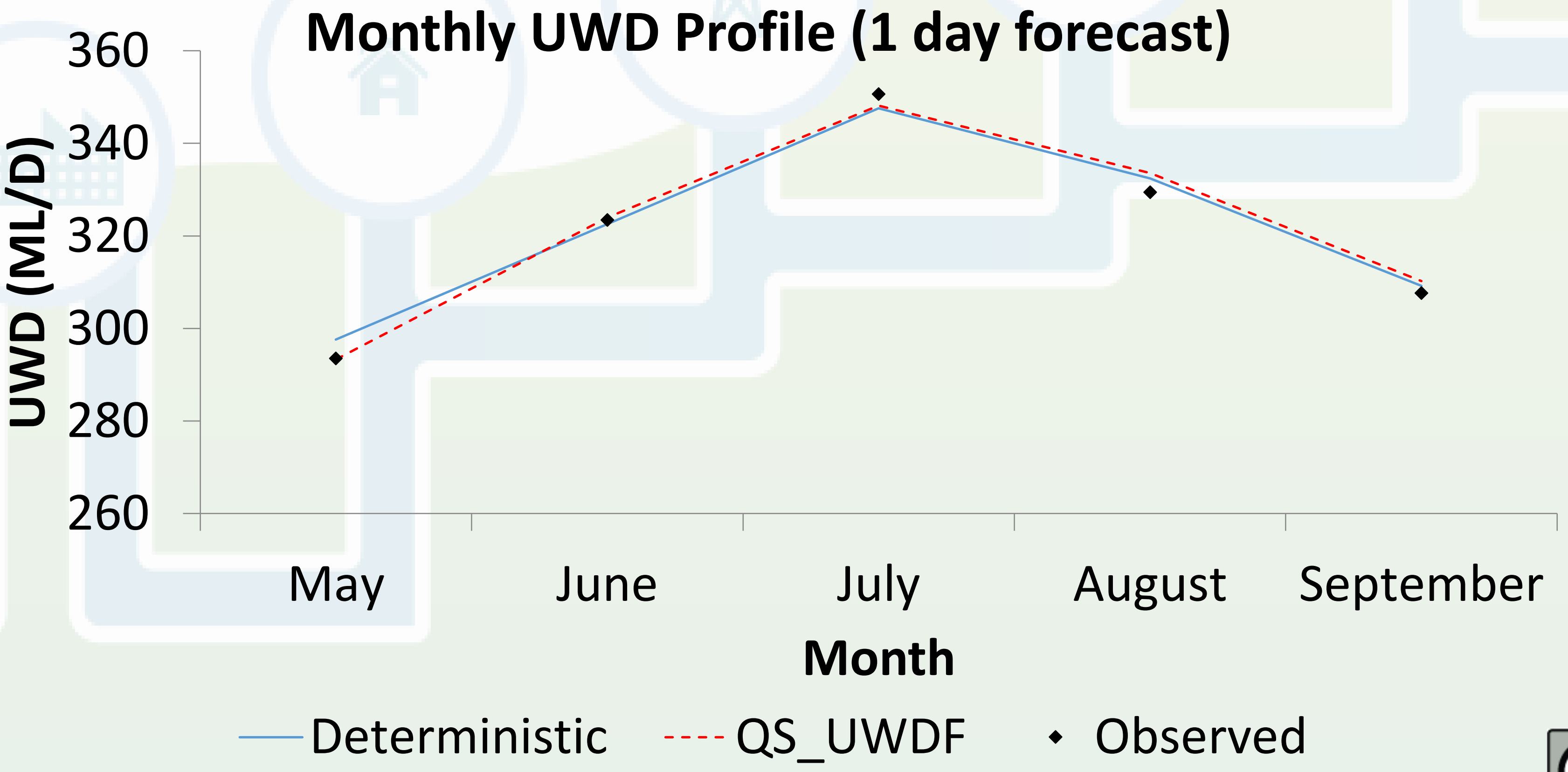
Method	Forecast Lead Time (Days) 1 MAE (ML/D)	CC	2 MAE (ML/D)	CC	3 MAE (ML/D)	CC
Deterministic	13.119	0.789	15.563	0.670	17.146	0.624
QS-UWDF	13.144	0.790	16.462	0.649	18.009	0.615

Results (1 day ahead forecasts)



Conclusions & Recommendations

- **Stochastic weather** inputs or **historical** records are both **viable inputs** for short-term **UWD forecasting** in **Ottawa, Canada;**
- **QS-UWDF** is **competitive** with **deterministic** modeling.



Conclusions & Recommendations

- Deterministic model has better performance **over longer lead times** (e.g. **CC** for QS-UWDF and Deterministic → **1 day**: 0.790 vs. 0.789; **2 day**: 0.649 vs. 0.670; and **3 day**: 0.615 vs. 0.624;
- **QS-UWDF** provide **uncertainty assessment** while deterministic only provides point forecasts; and
- QS-UWDF may be used for **risk**, **resiliency**, and **vulnerability** assessment of water supply systems.

Conclusions & Recommendations

- **Low-frequency variability correction** in weather generation should be explored;
- **Different stochastic simulation methods** should be compared (e.g. WeaGETS vs. k-NN stochastic simulation [Prairie et al., 2006]);
- **Different model ensemble approaches** can be implemented (e.g. via input variable selection);
- **Different machine learning techniques**, such as Extreme Learning Machines [Huang et al., 2006], can be utilized to improve computational efficiency; and
- The **QS-UWDF should be tested on numerous water supply systems** to further explore its practical applications.

THANK YOU!

References

Adamowski, J. et al. (2012), Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour. Res.,* 48, W01528.

Cawley, G.C. and N.L.C. Talbot (2004), A fast exact leave-one-out cross-validation of sparse least-squares support vector machines, *Neural Networks*, 17, 1467-1475.

Chen, J., F.P. Brissette, and R. Leconte (2010), A daily stochastic weather generator for preserving low-frequency of climate variability, *J. Hydrol.*, 388, 480-490.

Chow, V. T. (1978), Stochastic modeling of watershed systems, *Adv. Hydrosci.*, 11, 1-93.

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.

References

Mamo., T.G., I. Juran, and I. Shahrour (2013), Urban water demand forecasting using the stochastic nature of short term historical water demand and supply pattern, *J. Water Resour. Hydraul. Eng.,* 2, 92-103.

Prairie, J.R., B. Rajagopalan, T.J. Fulp, and E.A. Zagona (2006), Modified k-nn model for stochastic streamflow simulation, *J. Hydrologic Eng.*, 11, 371-378.

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.