2019

Smart Water: Short-Term Forecasting Application in Water Utilities

Mo'tamad Bata
University of Windsor

Smart Water: Short-Term Forecasting Application in Water Utilities

By

Mo'tamad H. Bata

A Thesis

Submitted to the Faculty of Graduate Studies

through the Department of Civil and Environmental Engineering

in Partial Fulfillment of the Requirements for

the Degree of Master of Applied Science

at the University of Windsor

Windsor, Ontario, Canada

2019

Smart Water: Short-Term Forecasting Application in Water Utilities

By

Mo’tamad H. Bata

APPROVED BY:

____

J. Johrendt

Department of Mechanical, Automotive & Materials Engineering

____

T. Bolisetti

Department of Civil and Environmental Engineering

____

D. Ting, Co-Advisor

Department of Mechanical, Automotive & Materials Engineering

____

R. Carriveau, Advisor

Department Civil and Environmental Engineering

April 18, 2019

DECLARATION OF CO-AUTORSHIP / PREVIOUS PUBLICATION

I. Co-Authorship

I hereby declare that this thesis incorporates material that is result of joint research. Chapters 3 & 4 of this thesis were completed under the supervision of Dr. Rupp Carriveau and Dr. David Ting. In all cases, the key ideas, primary contributions, experimental design, data analysis, interpretation, and writing were performed by the author. The contribution of co-authors was primarily through the provision of checking and comments on the literature review, methodology and modeling, results interpretation, providing feedback on refinement of ideas, editing of the manuscript, and advice on selecting peer reviewed journals for publication.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include the above materials in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

II. Previous Publication

This Thesis consists of three original papers that have been published / submitted or to be submitted for publication in peer-reviewed journals, as follows:

Thesis Chapter	Publication	Publication Status
Chapter 3	Bata, Mo'tamad; Carriveau, Rupp; Ting, David S.-K. 2018. "Short-term Water Demand Forecasting Using Nonlinear Autoregressive Artificial Neural Networks (ANN)". Engineering Application of Artificial Intelligence.	Under review

Chapter 4	Bata, Mo'tamad; Carriveau, Rupp; Ting, David S.-K. 2018. "Hybrid Self Organizing Map and Regression Tree Short-term Water Demand Forecasting Model".	To be submitted

I certify that I have obtained written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

III. I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works of any other author, in any form, is properly acknowledged at their point of use.

ABSTRACT

The unyielding interconnection between water and energy has made demand forecasting a necessity for water utilities. Electricity prices driven by the time of use has impelled water utilities towards short-term water demand forecasting. The progressive new Smart Water Grid platform has helped water utilities in utilizing their Water Distribution Networks. This two-way platform has provided developers and decision makers with robust models that rely on consumer feedback. Among these models is the water demand forecasting models. Multitudinous demand forecasting methods have been developed but none have utilized model implementation practicality. Utilities differ in size, capacity, and interest. While small size utilities focus on model simplicity, larger utilities prioritize model accuracy. This work focuses on a water utility located in Essex County, Ontario, Canada. This study presents three papers that focus on investigation and evaluation of short-term water demand forecasting techniques. The first paper compares water usage between two crops (tomatoes and bell peppers) in an effort to evaluate a crop to crop forecast technique that relies on one crops watering data in order to produce forecasts for another crop, The second paper examines the effect of model type, input type, and data size on model performance and computational load. The third paper proposes a new methodology where model performance is not sacrificed for model simplification.

DEDICATION

To Majeda and Hussein (Mom and Dad) for their endless support and continuously encouraging me to strive. To Emily (my wife) for her patience and abiding this education with me.

ACKNOWLEDGMENTS

I would like to thank my advisors Dr. Rupp Carriveau and Dr. David S.-K. Ting as well as the entire Turbulence and Energy Laboratory for their support, guidance, and patience they have shown throughout my studies. I would also like to acknowledge my committee members Dr. Jennifer Johrendt and Dr. Tirupati Bolisetti for their input along with Mr. Rodney Bouchard for providing valuable insight and data, without which this would not be possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background

Existing conventional water networks are aging. Conventional distribution systems operate in a one-direction supply process. Minimal feedback is provided from the demand side, i.e. the consumers. With their centralized one-directional flow characteristics (Newbold J., 2009), these systems are more prone to undetected failures and leaks. This can cause network disruptions and waste water, which is discouraging in solving the two major water issues, scarcity and security. Therefore, conventional water networks are unpredictable, economically infeasible, energy inefficient, and hard to maintain and manage.

Over the past decade, the water security and scarcity issues have propelled governments, water supply systems, and researchers to upgrade current networks and develop preventive measures. These measures aim to avert the global population increases and the escalating global warming effects; including but not limited to, droughts, fires, floods, and climate change.

New water management infrastructures have been successfully established to tackle the aforementioned challenges. The new platform of tools, technologies, and models is known as the Smart Water Grid (SWG). SWG is a two-way real time network with sensors and devices that continuously and remotely monitor the Water Distribution System (WDS). Smart water meters can monitor many different parameters such as pressure, quality, flow rate, temperature and others (Martyusheva O., 2014). The gathered data is then used in various tools, models, and decision make systems. For instance, data collected through pressure monitoring sensors can be utilized to detect and locate damaged pipelines or system leaks. Compared to the conventional water grid, SWG is more resilient, reliable, sustainable, and energy efficient.

The SWG approach can help solve one of the most complicated challenges in the Water Distribution System (WDS), the imbalance between water supply and demand. Water supply, measured by system production capacity, is usually a constant rate. On the other hand, the water demand represented by water consumption varies throughout the day and from day to another. For example, water demand is higher during daylight business hours than during night-time (Leirens et al. 2010). Factors other than time of the day affect water demand, such as: number of consumers, type of consumer, consumption seasonality, new technologies, etc. With this dynamic fluctuation in the water demand side, it would be near impossible for the supply side to keep the network in balance. The development of water demand forecasting models is essential in developing pumping schedules and operating the network in balance (Fodya & Harley, 2014).

Researchers have extensively studied and proposed water demand forecasting models. A rich review of over 30 papers published between the years of 2000 and 2010 was presented by (Donkor et al. 2014). Models seemed to perform well on the studied data. However, one common issue was found for the vast majority of the models, model practicality. Models were deemed impractical due to: difficulty in acquiring input data, complexity in models architectures, and ineffective development cost. Most models have focused more on model accuracy without any consideration on how practical it would be for a specific utility to adapt such model. Here, practicality is defined as model suitability for a specific utility. Model suitability is governed by utility size, capacity, and interest to deploy a model. Small size utilities, for instance, may assign more weight to model simplicity over accuracy. While, bigger utilities prioritize model accuracy over model simplicity.

Problem statement

The Union Water Supply Systems (UWSS) located (see Figure 1) in Essex County, Ontario, Canada, operates their network on reactive mode. When water levels in main reservoir drop, operators turn pumps switches on. By doing that, operators respond to the demand in the past and select pumps according to their knowledge and experience. UWSS faces new challenges where the reactive mode is no longer efficient. UWSS receives continuous requests of additional water demand by consumers. UWSS can supply the extra requested water demand, however, their capability is constrained by the time of use. Moreover, UWSS is impacted by the energy time of use prices fluctuation and the evolving renewable energy market penetration and its pressure on big consumers.

UWSS can tackle the aforementioned challenges by switching into proactive mode. The anticipated future demand is responded to within this mode instead of the past. Pumping schedules are prepared ahead, pumps are selected accordingly, time of the use prices can be benefited from. Also, demand peaks can be reduced which will result in less water and energy losses in the system. Proactive mode is based on pre-knowledge of the demand patterns and can be achieved through short-term water demand forecasting models.

The objective of this study was to propose and investigate short-term water demand forecasting models while considering model practicality. The proposed models are designed to assist water utilities in developing their pumping schedules. To do this, multiple forecasting methodologies were adapted for this purpose. Once the models were developed, the next step was to evaluate their performance and deployment practicality.

Figure 1. Studied area location

This study is made up of three research papers. The first paper (reviewed in chapter 2) has been published in the Agriculture journal; Evaluation of Crop to Crop Water Demand Forecasting: Tomatoes and Bell Peppers Grown in a Commercial Greenhouse (Rice et al. 2017) and proposed a simple model where an end-user of only 80% of the utility consumers (i.e. commercial greenhouses) are studied. This paper addresses the model complexity and practicality issue. The issue of model type, data size, and input selection are studied in Chapter 3. The second paper, which is under review at the Journal of Water Resources Planning and Management, investigates three model architectures with different levels of complexity in order to determine if complex models with more data instances and more data inputs perform significantly better than simple models with less data instances and less inputs. The third and final paper (chapter 4), which will be submitted to the Journal of Water Resources Planning and Management, addresses the complexity issue in another way. A hybrid model is presented where model practicality is considered without sacrificing the forecast accuracy.

References

Alvisi, S., Franchini, M., & Marinelli, A., (2007). “A short-term, patternbased model for water-demand forecasting.” J. Hydroinf., 9(1), 39–50.

Billings, B., & Jones, C., (2008). Forecasting urban water demand, 2nd Ed., American Waterworks Association, Denver, CO.

Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Roberson, J. A., (2014). Urban Water Demand Forecasting: Review of Methods and Models. Journal of Water Resources Planning and Management, 140(2), 146–159.

Fodya, C., & Harley, C., (2014). Suppliers View of Water Demand for a City- Modelling the influencing Factors. In S. Mambretti & C. A. Brebbia (Eds.), WIT Transactions on the Built Environment (pp. 27–37). Algarve, Portugal: WIT Press.

Gardiner, V., & Herrington, P., (1990). Water demand forecasting, 1st Ed., Spon Press, Norwich, U.K.

Ghiassi, M., Zimbra, D., & Saidane, H., (2008). “Urban water demand forecasting with a dynamic artificial neural network model.” J. Water Resour. Plann. Manage., 10.1061/(ASCE)0733-9496(2008)134:2 (138), 138–146.

Jain, A., Varshney, A., & Joshi, U., (2001). “Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks.” Water Resour. Manage., 15(5), 299–321.

Leirens, S., Zamora, C., Negenborn, R.R., & Schutter, B. D., (2010). Coordination in urban water supply networks using distributed model predictive control, in: Proceedings of the 2010 American Control Conference, ACC 2010, Baltimore, MD, art. no. 5530635, 2010, pp. 3957–3962.

Martyusheva O., 2014. Smart Water Grid – Plan B Technical Report. Department of Civil and Environmental Engineering. Colorado State University. Thesis and Dissertation.

Newbold, J.R., (2009). Comparison and simulation of a water distribution network in EPANET and a new generic graph trace analysis based model, MSc. Thesis, Environmental Engineering Department, Virginia Polytechnic Institute and State University, urn: etd-02082009-075711.

Rice, D.C.J., Carriveau, R., Ting, D.S.-K. & Bata, M. H., 2017. "Evaluation of Crop to Crop Water Demand Forecasting: Tomatoes and Bell Peppers Grown in a Commercial Greenhouse." Agriculture 7, no. 12: 104.

CHAPTER 2

Forecasting Methodology

Introduction

This chapter provides an overview of an end-user previous study on short-term water demand forecasting models for the same water utility. The content here presents in depth the methodology used in the previous study. Also, a general introduction of the methodologies used in this study is presented.

C2C Methodology

Crop to crop is a water demand forecasting methodology presented by Rice et. Al. (2017). The study focused on the UWSS's main consumer, the commercial greenhouses. UWSS services over 720 hectares of greenhouse operations with more than 57% of the crops is tomato. The study proposed simplified methodology where one crop's water demand is predicted based on another's. The base model was suggested to be an Artificial Neural Network (ANN) model where the water demand for the targeted crop is forecasted with a simple linear model. Two simple linear models were proposed, Linear Regression (LR) and Quotient Method (QM). LR and QM were deployed to forecast 24 hours ahead pepper water demand based on tomato's and vice versa.

The results showed that both LR and QM had on average a NRMSE of 25% more than the base ANN model when predicted demand compared to the actual demand. The increase in the ANN model accuracy was due its capability of capturing nonlinearity in the actual data. The results also showed that when the ANN base model predicted data was used in LR and QM, an average increase of 12% and 22% in the NRMSE was noticed, respectively. This disparity in error can be attributed to the magnitude of error present in the base model. Although, C2C is dependent on the base model, C2C provided an improvement to the current fixed demand methods used in the region. Also, the study suggested that ANN models outperforms the simple proposed linear models.

ANN Theory and Models

ANN is a biologically inspired network of connected nodes called artificial neurons configured to perform specific tasks (Hagan et. al., 1995). ANN consists of a Multi-Layer Perceptron (MLP): input layer, one or more hidden layers, and output layer (see Figure 1). The input layer is where data is fed at the beginning of the training process and anytime the network is used to predict outputs (training, validation, and test phases). The hidden layer is where most of the computations occur and it has neurons, connections of weights and biases, propagation function, and a learning rule. Neurons in the hidden layer are trained where only one neuron is activated for each input. This is achieved through forward or backward propagation where neurons are assigned weights after comparing the input to the output value and calculating the gradient of the loss function. The output layer contains the final form of the input data after it is processed. ANNs are widely used in different applications such as: pattern recognition, classification, regression, and clustering.

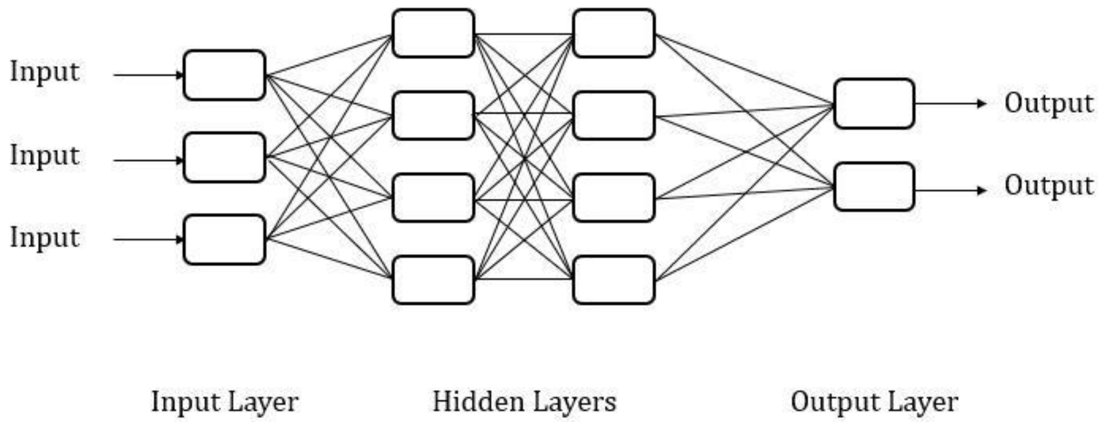

Figure 1. Typical Artificial Neural Network structure.

ANNs have advantages over the conventional models in demand forecasting applications. ANNs can generalize and model nonlinear complex relationships. That is important because most real-life problems have a nonlinear and complex relationship between inputs and outputs. In addition, ANNs have a high computational capacity, the ability to generalize, and infer unseen relationships on unseen data. Researchers (Adamowski & Karapataki 2010, Ghiassi et al. 2008; Jain & Ormsbee 2002) have used ANNs in short-term water demand forecasting application and showed that ANNs outperforms traditional forecasting technique. In this study, four ANN models are deployed to forecast the water demand t-time steps ahead. Then, these models are compared to a traditional adequate forecasting model based on two criteria: model forecasting accuracy, and computational load.

Forecasting methodology

The focus of this study is on short-term water demand forecasting. Here, short-term refers to how many t-time steps are forecasted ahead. Proposed models produced 8-hour, 24-hour, and 7-day forecasts. These spans were selected to utilize the operations and working shifts at the water utility.

Two forecasting methodologies are presented in this study. The first methodology is the nonlinear autoregressive with and without exogenous parameters. This group of models is investigated in chapter 3. The second methodology, studied in chapter 4, is a hybrid model resulted from the fusion of regression trees and self-organizing maps.

Nonlinear Autoregressive Models

Because of the seemingly nonlinear relationship in the water demand data, this nonlinear ANN model was chosen for forecasting. Nonlinear Autoregressive (NAR) and Nonlinear Autoregressive with Exogenous inputs (NARX) are a supervised machine learning Multi-Layer Perceptron (MLP) feedforward ANN models. Theses two models use historical time series data to predict in the future. The difference between NAR and NARX is that NARX utilizes more than a single input time series to predict a target series. In this thesis, for example, NARX used the historical water demand and weather parameters to predict the future water demand. Meanwhile, NAR only used the historical water demand to predict the future water demand. Both models have a similar workflow of training, testing, and validating. Both models were also built using the same Guide User Interface (GUI) neural net time series package in MATLAB R2017b platform.

Processed data is loaded to the workplace. Then, target data and input data are selected to be fed to the network input layer. At this step, the time series format is selected where data is converted into a standard neural network cell array form. The formatting function chosen in this study was "tonndata" because of its compatibility with neural networks. After conversion, data is divided up into three categories: training data, validation data, and testing data. Training data is the group presented to the network during training step and the network will adjust based on its errors and weights. Validation data is the portion that is used to measure the network generalization. Network will stop training when no generalization improvement is measured. Testing data is the last portion that is used to measure the network performance. In this study, default recommended data division of 70% training, 15% validation, and 15% testing was used. Ideally, datasets are divided into training dataset to build the model, and testing dataset to evaluate the model performance. However, a validation dataset is used in ANN models to avoid overfitting issue. Overfitting occurs when the model is excessively trained and its accuracy reaches 100%. Overfitted models are not preferred due to their poor performance on unseen data. The validation held back dataset estimates the model parameters independently from the training dataset.

At this step, the input layer is fully developed. Next, "narnet" or "narxnet" is used to define the NAR or NARX network architecture, respectively. These two functions define the feedback delays number, the hidden neurons number, and the training algorithm. Both NAR and NARX models were first trained with the default feedback delays of 2, hidden neurons of 10, and Levenberg-Marquardt “trainlm” as training algorithm. Levenberg-Marquardt is an iterative technique that locates the minimum of a multivariate function that is expressed as the sum of squares of nonlinear real-valued function (Levenberg, 1944). This regularization algorithm is developed to solve the overfitting problem in ANN models. The selection of this algorithm is based on obtaining the lowest mean squared error possible (Hagan et. al., 1995). The training algorithm, the number of delays, and hidden neurons were adjusted based on the network performance. After the network architecture is defined, performance measures are selected to evaluate the network. The Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) were selected for performance evaluation (i.e. equations 1 & 2, respectively). In case the network performance was not satisfactory, parameters such as number of neurons, number of delays, training algorithm, and network retrain option, could be calibrated. After reaching a steady state, where performance does not change with changing parameters, the closed-loop network could be exported and used to forecast water demand.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\mathrm{Yi} - \hat{\mathrm{Y}}\mathrm{i}}{\mathrm{Yi}}\right| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y_i}^2 - Y_i^2)} \tag{2}$$

Where,

n = the number of data points; Y_i and $\hat{Y}_i$ represent and and water outflow

i = the data point number (i.e. 1, 2, …, n)

Yi = the actual water outflow

Ŷi = the forecasted water outflow

Hybrid Model

A supervised machine learning technique is fused with an unsupervised machine learning model to perform short-term water demand forecasting. Supervised machine learning is a technique where a corresponding target output is known for each input instance, in contrast to unsupervised machine learning where no target values are known. Regression Trees (RT) technique was selected for the supervised machine learning part. RT implicitly perform feature selection which reduces data dimension (Breiman et al., 1984). This feature is important in a predictive model because it minimizes irrelevant data noise, which improves the model performance and reduces time of training. Other advantages of the RT technique are the minimal effort in data preparation, and the ease of model interpretability. Self-Organizing Map (SOM) is a clustering unsupervised model fused with RT technique to further reduce the data space where forecasting is performed. SOM groups input data into a defined number of clusters based on specific similarities (Kohonen T., 1982). In this study, the similarity criterion is set as the water demand intensity. This clustering technique is performed before data is fed to the forecasting model, RT model. Again, both models were built using the Guide User Interface (GUI) neural net time series package in MATLAB R2017b platform.

First, the SOM model is built, trained, validated, and tested. Input data is selected in matrix form within the neural clustering tool (nctool). After input data is loaded, the SOM network architecture is defined where clustering function and number of neurons are selected. In this study, "selforgmap" function is used to define the: row vector dimension size, number of training steps, initial neighborhood size, layer topology function, and neuron distance function. Values of these parameters were initiated as the tool recommended default (i.e. 2 neurons, 100 training steps, 3 neighborhood size, hexa-topology, "linkdist"). These parameters could be replaced by different values; however, the number of neurons was the dominant parameter that affected the SOM performance. Next, SOM is trained using the batch unsupervised weight/bias algorithm. Weights and biases are calculated and updated at the end of the entire pass of input data. The output of SOM is binary determining if an instance belongs to a specific cluster or it does not.

Second, SOM cluster number is added to the input data and uploaded to the RT model workplace. RT model requires defining the parameters of: number of cross-validation folds, maximum number of splits and minimum leaf size, Principal Component Analysis (PCA) feature, and the fitting algorithm.

- Cross-validation folds: It is important to define and include this feature to prevent the common training issue, overfitting. Number of folds was initially chosen to be 5 (default).
- Maximum number of splits and Minimum leaf size: RT algorithms grow deep trees by default, determining the number of maximum splits and leaf size helps in reducing model complexity and training time. The number of maximum splits and minimum leaf size were initiated as default (10 and 4, respectively.)

- Principal Component Analysis (PCA): this feature was enabled to transform features and remove redundant dimensions. PCA prevents the model from learning a previous learnt information. PCA extracts a small number of variables that best explain variance in the dataset.
- “fitrtree”: this was the algorithm used to fit the RT model. “fitrtree” returns a regression tree based on the input value contained in the predictor matrix and the output value in the response matrix.

After defining and selecting the previous parameters, RT model is trained, and performance is calculated. Further modifications are applied if the performance is unsatisfactory. This includes: adding extra neurons in the SOM model, add folds, increase the maximum splits and minimum leaf size values in the RT model.

References

Adamowski, J. & Karapataki, C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering*, 15(10), 729–743.

Breiman L., 1984. Classification and Regression Trees. Belmont, CA: Wadsworth International group.

Ghiassi, G., Zimbra, D.K., Saidane, H., 2008. Urban water demand forecasting with a dynamic artificial neural network model. Journal of Water Resources Planning and Management 134 (2), 138-146.

Hegan M. T., Demuth H. B., Beale M., 1995. Neural Network Design. Boston: PWS-.

Jain, A. & Ormsbee, L., 2002. Short-term Water demand forecast modeling techniques — Conventional Methods versus AI. *Journal American Water Works Association*, 94(7), 64–72.

Kohonen T., 1982. Clustering, Taxonomy, and Topological Maps of Patterns. Proceedings of the 6th International Conference on Pattern Recognition.

Levenberg K., 1944. A Method for the Solution of Certain Non-Linear problems in Least Squares. Quarterly of Applied Mathematics, 2, 164-168.

Rice, D.C.J., Carriveau, R., Ting, D.S.-K. & Bata, M. H., 2017. "Evaluation of Crop to Crop Water Demand Forecasting: Tomatoes and Bell Peppers Grown in a Commercial Greenhouse." Agriculture 7, no. 12: 104.

CHAPTER 3

Short-term Water Demand Forecasting Using Nonlinear Autoregressive Artificial Neural Networks (ANN)

Mo'tamad H. Bata, Rupp Carriveau, and David S.-K. Ting

Turbulence and Energy Laboratory, Ed Lumley Centre for Engineering Innovation, University of Windsor, Windsor, ON N9B 3P4, Canada.

Nomenclature

The following symbols are used in this chapter:

ANN	= Artificial neural network
NARX	= Non-linear autoregressive with exogenous inputs model
NAR	= Non-linear autoregressive model
SARIMA	= Seasonal Autoregressive Integrated Moving Average
MAPE	= Mean absolute percentage error
NRMSE	= Normalized root mean squared error
m^3	= Cubic meter
hr	= Hour
HD	= Hot Deck
^oC	= Degree Centigrade
g	= Gram
W/m^2	= Watt per squared meter
kPa	= Kilo Pascal
PCC	= Pearson Correlation Coefficient
MLP	= Multi-layer perceptron
ARX	= Autoregressive with exogenous inputs model
Y_i	= Current observed value of time series
X_i	= Current value of exogenous time series
t	= Time
n	= Number of data point (i.e. 1, 2, 3, etc.)
RMSE	= Root mean squared error
$\overline{Y}_i$	= Data set mean
$\widehat{Y}_i$	= Forecasted water outflow

Introduction

Providing reliable, secure potable water to consumers is a vital core aspect of any water utility's daily operations. Short-term water demand forecasting can help a water utility to more efficiently manage many of their principal operations. These include the distribution system, pumping schedule, and storage assets. Water demand is temporally and spatially dynamical, often in short time frames. Water demand profiles are very dependent on regional influences. Some of the main driving factors are consumer type, water prices, population growth, economic growth, technological practices, and management strategies. A forecasting model's capacity to take into consideration the aforementioned variables is important; three main factors are deemed the main engine of a relatively accurate model. The first and most important factor is the input. Researchers (Guo et al., 2018; Arandia et al., 2016; Alvisi et al., 2007; Homwongs et al., 1994; Shvartser et al., 1993; Jowitt and Chengchao, 1992) have used historical water demand as a single input in their forecasting models. Their models showed reasonably accurate forecasts. Others (Rice et al., 2017; Herrera et al., 2010; Ghiassi et al., 2008; Bougadis et al., 2005; Aly and Wankule, 2004; Jain and Ormsbee, 2001; Jain et al., 2001) have used extra (Exogenous) inputs in addition to the historical water demand, weather and seasonality data, such as, temperature, rainfall, and evaporation, time of the day, and day of the week. The inputs used in this study are explained in the Methods section.

The second driving factor is the time horizon, which contains two-time dimensions, one to determine how far back the historical data is needed, and one to set the length of the forecast. Previous studies have chosen the historical time span without clarifying the reasons for selection. In this paper, historical hourly data of the last 5 years (2013-2017), the most recent year of the 5 years data (2017), and the 4 most recent months of the most recent year (September 2017-December 2017) were used as inputs. These were chosen to investigate the effects of historical data length on the forecast accuracy. The other time dimension is how far ahead the model will forecast, this horizon is governed by the model application (Bakker et al., 2003). Short-, medium-, and long-term forecast horizons have been extensively discussed by (Donkor et al., 2012; Billings and Jones 2008; Gardiner and Herrington 1990). For our purposes here, short-term forecast horizons of 24 hours and 1 week ahead were utilized for the purpose of daily operations and pumping schedule control.

The third significant factor is the model method itself. Many studies have assumed the presence of a linear relationship between the predictors and the response in demand forecasting; however, nonlinearity oftentimes manifests itself in these variables (Ghiassi et al., 2008). ANN models have proven themselves in linear and non-linear applications. This is evident in their ability to progressively learn, self-train, and improve performance of demand forecasting systems. They have been broadly used in demand forecasting in various applications. Kandananond, 2011; Chang et al., 2011; Feilat & Bouzguenda, 2011; Amjady & Keynia, 2011; Azadeh et al., 2007 used ANN in electricity load forecasting. Guo et al., 2018; Bennet et al., 2013; Adamowski & Karapataki, 2010; Mitrea et al., 2009; Ghiassi et al., 2008; Jain & Ormsbee, 2002 used ANN in water demand forecasting and showed that it outperformed traditional water demand forecasting techniques. Although many ANN approaches have proven useful in modeling complex demand forecasting, a few undetermined parameters of the model architecture have made ANN models harder to elucidate and optimize. Theses architectural parameters include the training algorithm, and the optimal number of neurons and hidden layers.

Guo et al., 2018 broadly classified water demand forecasting methods into traditional methods and learning algorithms. Traditional statistical methods have used linear time series or linear regression to tackle this challenge and have proposed simple statistical models (Wong et al., 2010; Zhou et al., 2000) to tackle it. A popular model used extensively among this group is the Autoregressive Integrated Moving Average (ARIMA). These models have been widely used (Kofinas et al., 2014) because they have a simple structure, are easy to implement and interpret, and do not require a lot of input data. Learning algorithms identify water demand forecasting as a nonlinear problem. With the progressive advancements in machine learning and data analytics, these algorithms have achieved higher accuracy in prediction models. Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Deep Learning models are the most studied within this group. More detailed description and the use of traditional methods and learning algorithms can be found in Arandia et al., 2016 and Guo et al., 2018.

In this manuscript, two nonlinear ANN models and one linear ARIMA model were employed with a specific architecture to forecast water demand. Model nature and methodology (i.e. linear vs. nonlinear) was first investigated. Secondly, input type (single input vs. exogenous inputs) influence on the predicting model was studied. Lastly, data size was scrutinized in an attempt to determine how far back the data is required to be.

Study Area and Data

Union Water Supply System (UWSS) is a municipal water supply system owned by the Ontario municipalities of Leamington, Kingsville, Essex, and Lakeshore. UWSS supplies water to approximately 65,000 residents and in addition to commercial, industrial and agricultural customers. A consumer breakdown inspection reveals that on an annual average; 78% of the utility's outflow is consumed by commercial greenhouses. The rest, 22%, is consumed by the residential and industrial sectors.

Five years (Jan 2013 – Dec 2017) of continuous hourly data was used to train the models. It was grouped into three main categories, the plant outflow measured in m^3/hr (see Figure 1), the weather, and the seasonality. Outflow data had 3.75% of missing and erroneous (i.e. zero data) data, missing data was imputed using the Hot Deck (HD) method, and the imputation was based on the highest correlated factor, previous day same hour, shown in Table 1. Further description of the imputation is discussed in the *Data Pre-processing* section.

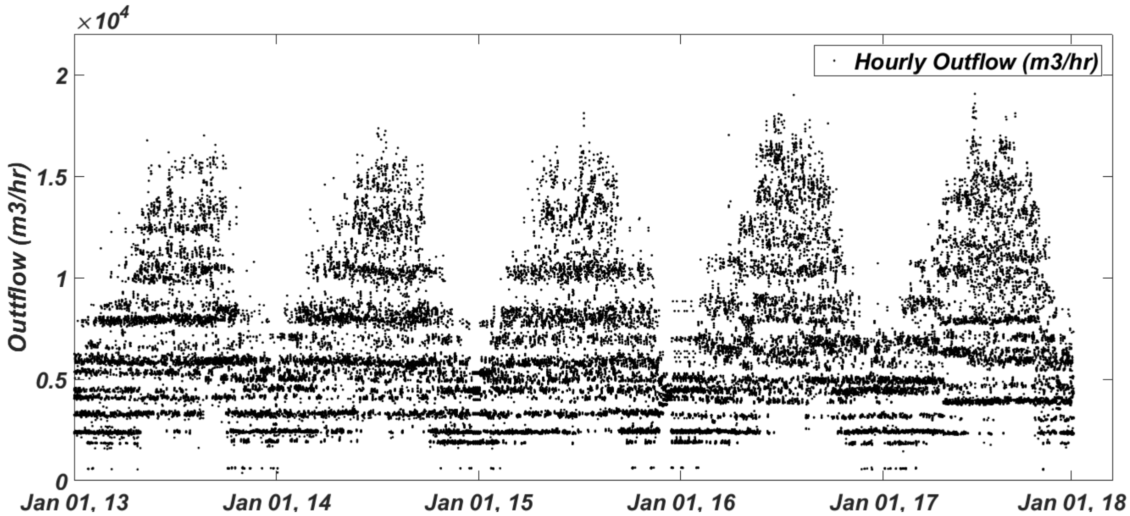

Figure 1. Utility water outflow m^3/hr

The observed weather data at t=0 is used as an exogenous input in the proposed NARX models to predict at t=24 and t=168; hourly data of ambient temperature (^{0}C) (see Figure 2), dew point (^{0}C), absolute humidity (g H_2O / m^3 air), solar radiation (W/m^2), and station pressure (kPa). A total of 2.30% missing data, mainly solar radiation, was imputed similarly using the HD method. The last category is the seasonality and it contains the year (2013-2017), the month (1-12; 1 represents January and 12 represents December), the day of the month (1-28,29,30 or 31), the day of the week (1-7; 1 represents Saturday and 7 represents Friday), and hour of the day (1-24; 1 represents 12:00 -1:00 am and 24 represents 11:00 pm – 12:00 am).

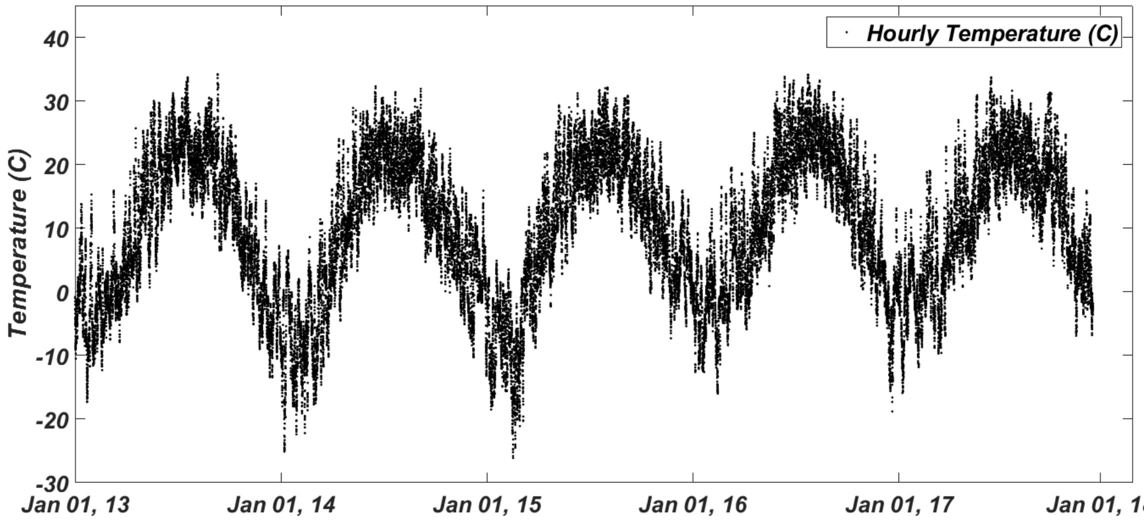

Figure 2. Ambient temperature (^{0}C)

In terms of input data, the Pearson Correlation Coefficient (PCC, also referred to as Pearson's r) is evaluated for 14 different demand, weather, and seasonality predictors. The 14 input predictors were selected according to data availability and previous consideration in demand forecasting literature. In Table 1, the predictors PCC is shown in one column and the correlation strength is shown in the other column. The correlation strength is a description using the guide proposed by (Evans, 1996) for the absolute value of r. PCC ranges between 1 and -1, where 1 is the total positive linear correlation, 0 is no linear correlation, and −1 is the total negative linear correlation. While PCC is not robust in terms of measuring dependency (Wilcox, 2005), other robust estimators of correlation are not as interpretable as PCC.

Table 1. PCC measured for the predictors

Predictor	Rank	PCC	Strength
Previous day same hour	1	0.835	Very strong
Previous week same hour	2	0.758	Strong
Previous 24hrs average	3	0.569	Moderate
Temperature	4	0.478	Moderate
Absolute humidity	5	0.397	Weak
Dew point	6	0.393	Weak
Solar radiation	7	0.363	Weak
Year	8	0.064	Very weak

Month	9	0.036	Very weak
Hour of the year	10	0.035	Very weak
Hour of the day	11	0.026	Very weak
Day of the week	12	0.020	Very weak
Station pressure	13	- 0.010	Very weak
Day of the month	14	0.007	Very weak

Data Pre-processing

Missing data is an untold story of a record. Missing data is a common occurring issue that significantly affects the information that can be drawn from the data. The missingness could be a result of many reasons. Some of these common reasons are: incomplete data collection, faulty equipment, non-response in data, and impractical feature case.

Handling missing data is essential prior to deploying it. Missing data could be discarded; this is the simplest technique of handling the missingness. However, it is not effective if the missingness rate is high. Also, missing data could be replaced, estimated, or filled based on domain knowledge (Salvador et al. 2015). In this paper, missing data is handled by imputation, where the missing values are filled to form a complete dataset. One used popular strategy of imputation is the *Hot Deck*.

The term "Hot Deck" originally used as computer punch cards for data storage, and refers to the deck of cards for donors available for a non-respondent. The deck was "hot" since it was currently being processed, as opposed to the "cold deck" which refers to using pre-processed data as the donors, i.e. data from a previous data collection or a different data set (Andridge et al. 2010).

Hot Deck is a common imputation technique and is considered in this paper. With HD imputation, the missing value is replaced by a similar responding unit. This method is extensively used in practice where a non-respondent missing value (called the recipient) is replaced by a respondent value (called the donor) with respect to characteristics observed by both cases (Andridge et al. 2010). Selection of the donor leads to two versions of the Hot Deck imputation. The first version is the *Random Hot Deck* method where the donor is selected randomly from a donor pool. Whereas for the second version, the *Deterministic Hot Deck*, one single donor is selected; usually the nearest neighbor.

In this paper, the *Deterministic Hot Deck* is applied to impute the missing data on provided dataset. The nearest neighbor donor is selected which is similar to the *k-Nearest Neighbor (K-NN)* where k is equal one. The nearest neighbor is determined though a distance function, herein, the *Euclidean Distance* presented in equation 1.

$$d = \sqrt{\sum_{i=1}^{n} (Qi - Pi)^2} \tag{1}$$

Where,

d is the Euclidean Distance

i is the record number

n is the dimension number

Q is the reference point

P is the target point

The methodology of the HD imputation is simple. If Y and X are two variables/features where Y is the target (i.e. the Donee) column that has the missing values and X is the covariant (i.e. the Donor). For the missing value Y_4, the distance function (i.e. Euclidean Distance) determines the nearest neighbor to X_4. Then, the corresponding Y_n of that nearest X_n is replaced for the original missed value, Y. Figure 3 illustrates the process.

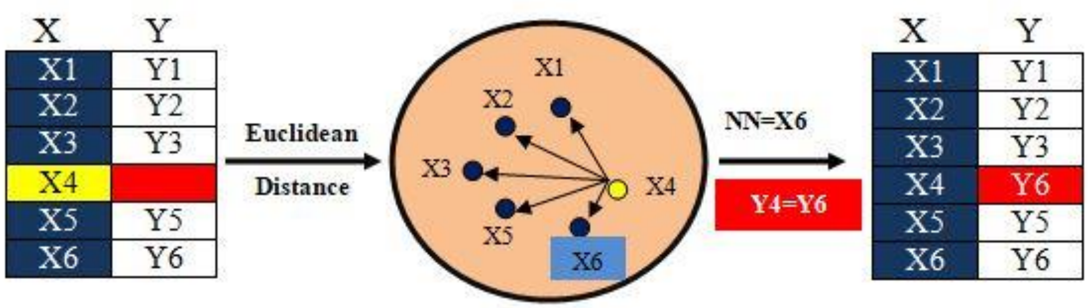

Figure 3. Nearest Neighbour Hot Deck Imputation

Methods and Model development

SARIMA Model

SARIMA is a simple statistical model that is used to analyze and predict a time series. SARIMA is a seasonal generalization of the Autoregressive Moving Average (ARMA) with the addition of seasonality and integration. SARIMA models consist of both seasonal and non-seasonal parameters and are denoted by ARIMA *(p, d, q)* x *(P, D, Q)$_S$* (Shumway and Stoffer, 2000). The *(p, d, q)* non-seasonal order of the model is the number of Autoregressive (AR) parameters, differences, and Moving Average (MA) parameters. The *(P, D, Q)$_S$* order of the seasonal component of the model is the AR parameters, differences, MA parameters, and periodicity. These non-seasonal and seasonal parameters are iteratively identified through plotting the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). Regarding the data needed to properly fit a SARIMA model, *Arandia et al., 2016,* showed that a 7-day window is enough to forecast 24 hours ahead. In this paper, two SARIMA models with two different training windows were fitted to forecast water demand. The first model, ARIMA (0, 1, 1) X (0, 1, 1)$_{24}$, used 7 day hourly water demand data to forecast 24 hours ahead. The second model, ARIMA (0, 1, 1) X (0, 1, 1)$_{168,}$ used four recent months of hourly water demand data to forecast 7 days ahead.

Nonlinear Autoregressive with Exogenous inputs (NARX)

NARX models are recurrent dynamic networks with feedback connections that enclose several layers of the network. NARX is a Multi-Layer Perceptron (MLP) feedforward ANN. These models are based on the linear ARX model, which is commonly used in time-series modeling. (Pisoni et al., 2009) claimed that the nonlinear neural network models (NARX) have better performance than the polynomial NARX models. Later (Mustafaraj et al., 2011) validated that claim and showed that neural network NARX models outperform the linear ARX models. A NARX model delineates the current value of a time series (Y_i) to both past values of the time series (Y_{i-1}, Y_{i-2},…, Y_{i-n}) and current (X_{ai}, X_{bi},…, X_{ni}). It also demarcates past values of the correlated (Exogenous) series (X_{ai-1}, X_{bi-1}, …, X_{ni-1}, X_{ai-2}, X_{bi-2}, …, X_{ni-2}, X_{ai-n}, X_{bi-n}, …, X_{ni-n}). Equation 2 illustrates the defining equation by (Demuth et al., 1998) for the NARX model and the NAR model explained in the next section.

$$Y(t) = f\,(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-n}, X_{t-1}, X_{t-2}, \ldots, X_{t-n}) \qquad (2)$$

Figure 4 (a) illustrates the algorithm used to develop NARX models in the MATLAB R2017b platform. First, data is divided into input data (Exogenous data) and its corresponding target data (water outflow data). Then, a training function is selected based on memory performance and problem type. Input delays, feedback delays, and hidden layer size are then estimated and the NARX net is created. After creating the net, input and feedback pre-/post-processing functions are chosen. Data is then prepared for training and simulation where inputs and delays are selected to feed a new network design at each time step. Also, at each time step, data samples are divided randomly into training, testing, and validation. At this point, the network is ready to be trained, however, its performance is not yet evaluated. Therefore, one or more performance functions are selected to check whether the network performs as desired or the network parameters and structure need to be estimated again. If the desired performance is met, network parameters are returned, and the network is enclosed to forecast steps ahead. All NARX and NAR networks in this paper are MLP with the same training, processing, preparing functions (i.e. trainbr, proccessfncs, and preparets, respectively).

Three NARX models were built to forecast the water demand 24 hours and one week ahead. The criterion that was considered to differentiate between the three models is the historical data span length (5 years, 1 recent year, and 4 recent months) for both water outflow and exogenous inputs. The selection of 5 years was based on the use of all available data. Meanwhile, 1 recent year and 4 recent months were selected to investigate how using less data would impact model accuracy versus model complexity. Further selections could be made. However, the consumer activity periodicity must be acknowledged, here that's commercial greenhouses. A researcher could have different data span length according to the utility consumer breakdown, for this specific utility, the division into three main seasons was assumed to accommodate the main consumer (Commercial greenhouses) seasons of planting.

```
1:  Read input data, target data
2:  Choose training function
3:  Create NARX net
4:  Choose input, feedback pre-/post- processing function
5:  Prepare data for training and simulation
6:  Setup data division for training, testing, and validation
7:  Choose a performance function
8:  Train, test, and validate network
9:  Run performance checking
10: if passes then
11:     Select network
12: else
13:     Go to 3
14: end if
15: Return network parameters and close for prediction steps
    ahead
```

```
1:  Read input data
2:  Choose training function
3:  Create NAR net
4:  Choose input, feedback pre-/post- processing function
5:  Prepare data for training and simulation
6:  Setup data division for training, testing, and validation
7:  Choose a performance function
8:  Train, test, and validate network
9:  Run performance checking
10: if passes then
11:     Select network
12: else
13:     Go to 3
14: end if
15: Return network parameters and close for prediction steps
    ahead
```

Figure 4. a) NARX network algorithm **b)** NAR network algorithm

Nonlinear Autoregressive (NAR)

NAR models are recurrent dynamic networks with feedback connections enclosing several layers of network. NAR also is a Multi-Layer Perceptron (MLP) feedforward ANN. NAR models are used extensively in time series demand forecasting, a comprehensive literature review by (Moreno- Chaparro et al., 2011) lists different types of time series and methods that used NAR models for prediction in various applications. A NAR model predicts the current value of a time series (Y_i) using past values (Y_{i-1}, Y_{i-2},…, Y_{i-n}), see Equation 3 for the defining equation. NAR models were developed with an algorithm (see Figure 4 (b)) similar to the NARX's. However, NAR models are fed with only input data. Three NAR models were built in this study and used to forecast the water demand for 24 hours and one week ahead; the same criterion used to produce the three NARX models was used similarly in the NAR models.

$$Y(t) = f\,(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-n}) \tag{3}$$

Baseline comparison and model performance

The actual measured water outflow is used as a basis for evaluation. The forecasted target value ($\hat{Y}_i$) is compared to the target actual measured value (Y_i) and overall model performance is measured by: (1) Mean Absolute Percentage Error (MAPE) and (2) Root Mean Squared Error (RMSE). However, the dataset that was used as an input for the models has different ranges and means, the Normalized Root Mean Squared Error (NRMSE) is included to syncretize those differences. Equations 4, 5 and 6 represent the MAPE, RMSE and NRMSE, respectively, where n represents the number of data points; Y_i and $\hat{Y}_i$ represent the actual and the forecasted and water outflow, respectively, and $\overline{Y_\iota}$ is the data set mean.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\mathrm{Yi} - \hat{\mathrm{Y}}\mathrm{i}}{\mathrm{Yi}}\right| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\,\widehat{Y_\iota}^{\,2} - Y_i^2)} \tag{5}$$

$$NRMSE = \frac{RMSE}{\bar{Y}_i} \qquad (6)$$

Results and Discussion

Linear vs. Nonlinear

The results for the seven proposed models forecasting 24 hours ahead and 1 week ahead are shown in Table 2. It can be observed that the forecast in both time horizons have a relatively better overall performance when nonlinear models (i.e. NAR and NARX) were employed compared to the linear models (i.e. SARIMA). On average, MAPE error for NAR models compared to the SARIMA models is 3% and 15% less, forecasting 24 hours ahead and 1 week ahead, respectively. Likewise, NARX models performed with 30% and 36% less error compared to SARIMA models, forecasting 24 hours ahead and 1 week ahead, respectively. That is because of: (1) nonlinear models are able to capture both linearity and nonlinearity in the time series; and (2) nonlinear models employed in this paper, ANN models, have a complex multi-layer structure that finds correlation between input, feedback, and output parameters. Regarding model stability, nonlinear models seem to perform with greater stability as the relative error values are lower compared to linear models.

Table 2. Models overall performance

Forecast ahead	Historical data	Model	MAPE (%)	NRMSE (%)	Relative Error		
						Percentile	
					Standard deviation	5%	95%
24 Hours	7 days	SARIMA-24	17.5	17.7	0.0821	±0.86%	±9.65%
	5 years	NAR-5y	18.6	18.9	0.0873	±0.78%	±9.16%
	1 year	NAR-1y	17.5	17.6	0.0816	±0.75%	±8.48%
	4 recent months	NAR-4m	14.8	15.2	0.0701	±0.72%	±8.62%
	5 years	NARX-5y	18.1	18.5	0.0864	±0.64%	±7.95%
	1 year	NARX-1y	13.4	13.6	0.0622	±0.46%	±6.80%
	4 recent months	NARX-4m	04.3	06.6	0.0438	±0.23%	±5.78%
7 Days	4 recent months	SARIMA-168	19.7	22.6	0.0934	±0.79%	±11.55%
	5 years	NAR-5y	19.1	20.2	0.0898	±0.74%	±10.84%
	1 year	NAR-1y	18.7	19.2	0.0904	±0.83%	±9.98%
	4 recent months	NAR-4m	12.6	15.0	0.0735	±0.52%	±8.32%

5 years	NARX-5y	18.5	18.7	0.0868	±0.69%	±10.45%
1 year	NARX-1y	13.6	13.9	0.0703	±0.57%	±7.12%
4 recent months	NARX-4m	05.6	07.0	0.0486	±0.16%	±6.14%

Single Input vs. Exogenous inputs

Another perspective that affected model performance was the input type. Here, the proposed models are grouped into single input models, and exogenous models (i.e. models that have X in their identifier name). SARIMA models and NAR models belong to the first group by feeding only the water outflow as a model input. Whereas, NARX models have been fed by extra correlated weather inputs. NARX average overall performance in terms of error, MAPE, decreased by 30% and 25% compared to NAR models forecasting 24 hours ahead and 1 week ahead, respectively. NARX average error also decreased, as mentioned before, by 30% and 36% compared to SARIMA models. Even though NAR and NARX models have a similar structure, the inclusion of exogenous parameters has advantaged the ANN model. Again, this is due to extra correlations drawn from the extra inputs. In addition, NARX models have lower relative error, which may suggest greater stability.

Data span

Data span here refers to how far back data is needed to adequately train a model. The data span focus in this research was on the ANN models. Results show that four months data was sufficient to train both NAR and NARX models. Comparing 4 months to 5 years and 1 year data models, error dropped by 20% and 15% in NAR model and by 76% and 68% in NARX model, respectively, forecasting 24 hours ahead. Further, there were reductions by 35% and 33% in NAR model and by 70% and 59% in NARX model, respectively, forecasting 1 week ahead. While it may be intuitive to expect that a time series model would improve with more historical data; this was not the case in this research. Both NAR and NARX models had a drastic error decrease when fed with less historical data. The reason is due to the data itself rather than the model structure. A major change in water consumption profiles had occurred in the region where the utility is located. Greenhouses, the main consumer, started to switch from growing vegetables to Marijuana after the new Canadian legalization act occurred in 2018. The change in agricultural activities in the studied area had affected water demand profiles and left the historical data with little information to add. Another important observation is the dramatic error decrease in NARX model compared to NAR model. Again, this emphasizes the importance of including correlated exogenous parameters in the model. These exogenous parameters reinforced the model by adding more information when the studied parameter (i.e. water outflow) no longer had the same trend, seasonality, and consumption profiles.

Computational Load

As mentioned earlier, SARIMA models have a simple structure and do not require a large quantity of data for training. On the other hand, NAR models have a complex structure and use more parameters. Even more complicated, NARX models are fed with exogenous inputs. This increases the model complexity and number of parameters. Model performance indicators are a good measure for model accuracy. However, they do not necessarily embody the model complexity issue. Therefore, the computational load was evaluated to better reflect the model's complexity and deployment practicality. Computational load, Table 3 was evaluated from two perspectives: (1) Akaike Information Criteria (AIC), which penalizes models that use more parameters; and (2) Computation time spent during data training steps. The computational tool used in this research was HP Pavilion TS 14 Notebook PC with a 1.6 GHz Intel Core i5 processor and 8 GB memory.

Table 3. Computation load of forecasting models on training data

Model	Historical data	Mean (m^3/hr)	Standard deviation (m^3/hr)	Identifier	Indicators	
					AIC	Time (s)
SARIMA	7 days	3858	1394	SARIMA-24	2,379	21
	4 recent months	6418	3251	SARIMA-168	9,481	39
NAR	5 years	6114	3407	NAR-5y	107,284	82
	1 year	6472	3669	NAR-1y	29,641	42
	4 recent months	6418	3251	NAR-4m	18,486	28
NARX	5 years	6114	3407	NARX-5y	228,652	346
	1 year	6472	3669	NARX-1y	103,829	187
	4 recent months	6418	3251	NARX-4m	41,451	138

AIC and training time results reveal that the SARIMA model had the lowest and most preferred performance. SARIMA-24 has had a relatively low AIC due to the shortest data span, 7 days, used to feed the model. SARIMA-168 with a similar structure had a 4-fold higher AIC. This was because the model was trained with a longer data span, 4 recent months. Ranked second was the NAR model. The ANN single input model with complex structure had approximately twice the AIC over the SARIMA model. That is NAR-4m compared to SARIMA-168 where both models were trained with the same four recent months data. Lastly, for the ANN model with exogenous inputs, the NARX-4m showed an AIC 4-times higher than the SARIMA-168, and 2-times higher than the NAR-4m. This ranking was anticipated because AIC consists of two terms, likelihood and number of parameters. Table 3 and Figure 5 show that the number of parameters was the dominant determinant in AIC values. Exogenous inputs along with longer data span added more parameters and errors to their associated models.

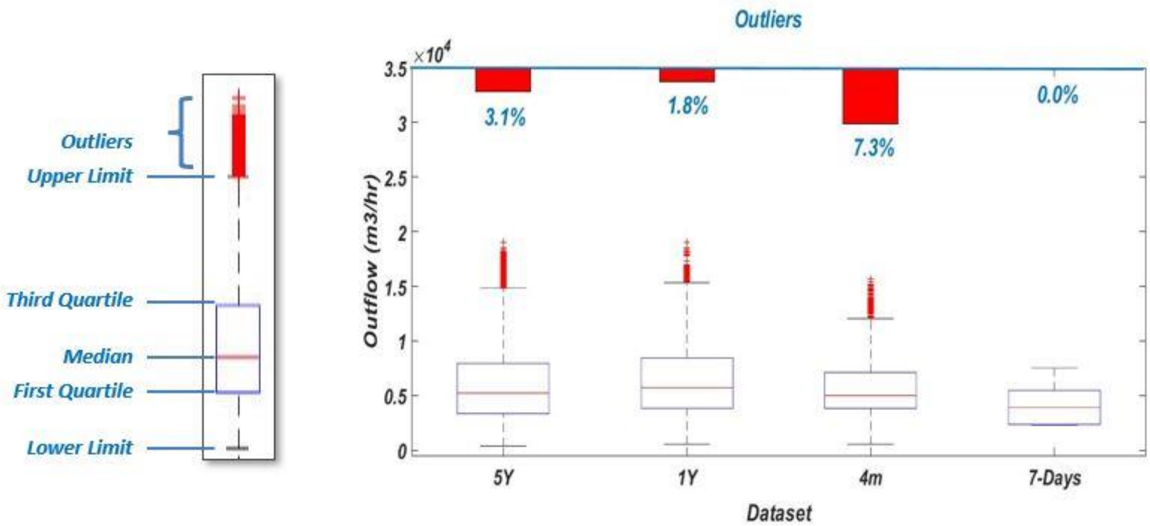

Figure 5. Datasets distribution and outliers.

Model Error Histogram

The error histogram for one of the proposed models, specifically, NARX-4m, is shown in Figure 6. For a forecast of 24 hours ahead with a mean of 5852m^3/hr, the model predicted roughly 35.50% of the data with an overestimation average of 38.9 m^3/hr. Also, 20.15% and 24.34% of the data were overestimated and underestimated by an average of 461m^3/hr and 447m^3/hr, respectively. In total, approximately 80% of the 2238 predicted hours were estimated with an error of less than 10% of the mean. That said, approximately 5%, 110 hours of the 2238 hours tested to forecast 24 hours ahead, had an overestimation or underestimation by 1000-2000 m^3/hr. This relatively high error in forecasting is deemed to be due the dramatic peak during some random mid-day hours. This drastic change in the outflow was not tracked by the one-hour data models. A shorter time span dataset, e.g. 15 minutes, would have decreased the percentage and amount of overestimation or underestimation.

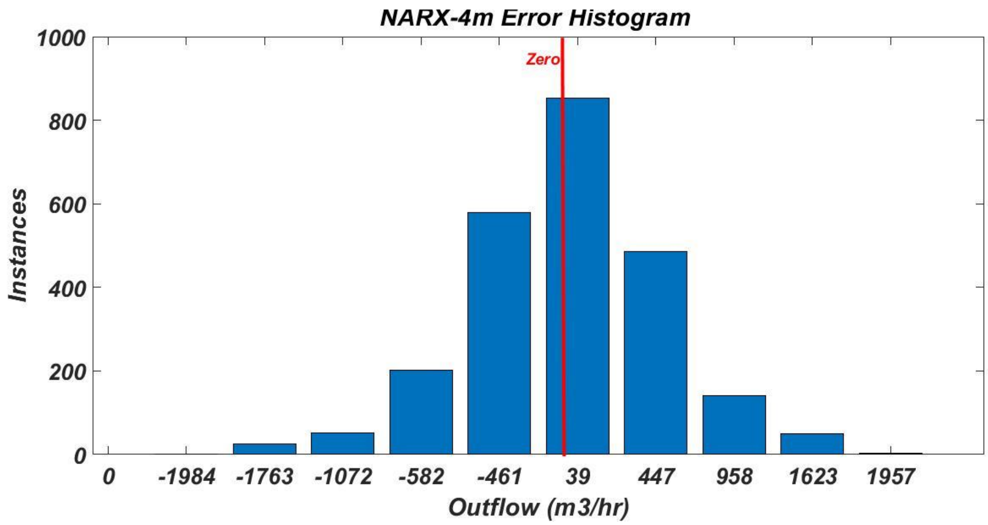

Figure 6. NARX-4m model error histogram

Models application

In this section, one week (12/02/2107 3:00pm – 12/09/2017 2:00pm) water outflow data was extracted from the utility daily operational data log and held back. SARIMA, NAR, and NARX models were deployed to forecast the water outflow for 24 hours ahead and 1 week ahead. The results of the performance in terms of MAPE are shown in Table 4. In addition, the forecasted patterns were compared to the actual held back water outflow for the targeted time horizon, Figures 7 and 8 display the results for models forecasting 24 hours ahead and 1 week ahead, respectively. Both results forecasting 24 hours ahead and 1 week ahead show a better performance for nonlinear models (i.e. NAR and NARX) over linear (i.e. SARIMA) models. Also, the models with exogenous parameters have outperformed the models with single input. Furthermore, shorter data span models, 4m, performed better than models with longer data span, 5y and 1y.

Table 4. Deployed models performance

Forecast ahead	Model	Error (%)	
		MAPE	NRMSE
24 Hours	SARIMA-24	17.9	18.3
	NAR-5y	15.6	15.9
	NAR-1y	16.5	16.5
	NAR-4m	13.8	13.5
	NARX-5y	14.1	14.0
	NARX-1y	13.4	13.6

	NARX-4m	12.3	12.7
7 Days	SARIMA-168	19.7	19.4
	NAR-5y	16.1	16.3
	NAR-1y	15.7	15.4
	NAR-4m	14.4	14.6
	NARX-5y	16.5	16.1
	NARX-1y	13.1	13.3
	NARX-4m	10.6	11.0

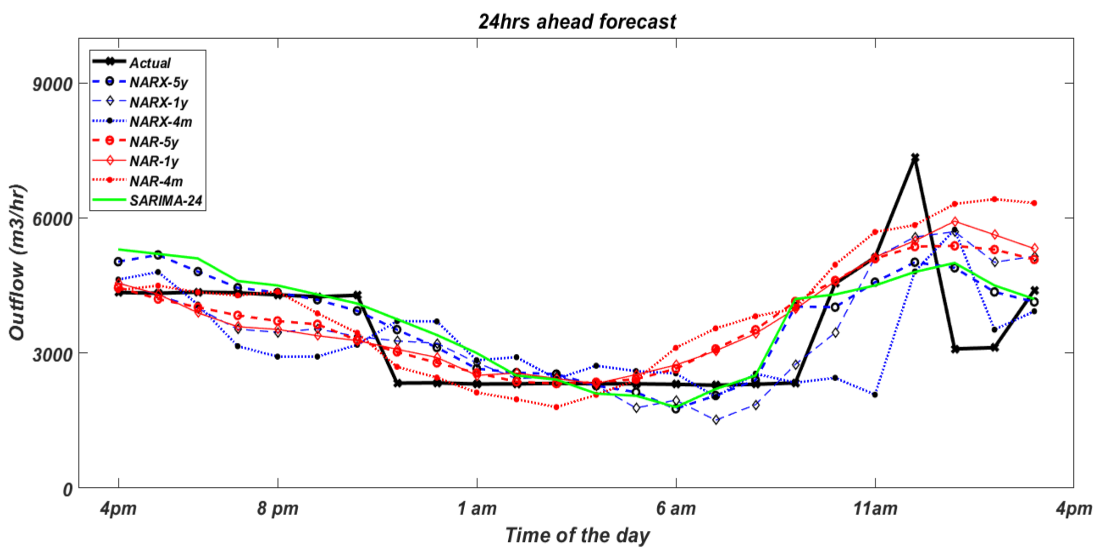

Figure 7. Models forecasting water demand 24 hours ahead (12/02/2107 3:00pm – 12/03/2017 2:00pm)

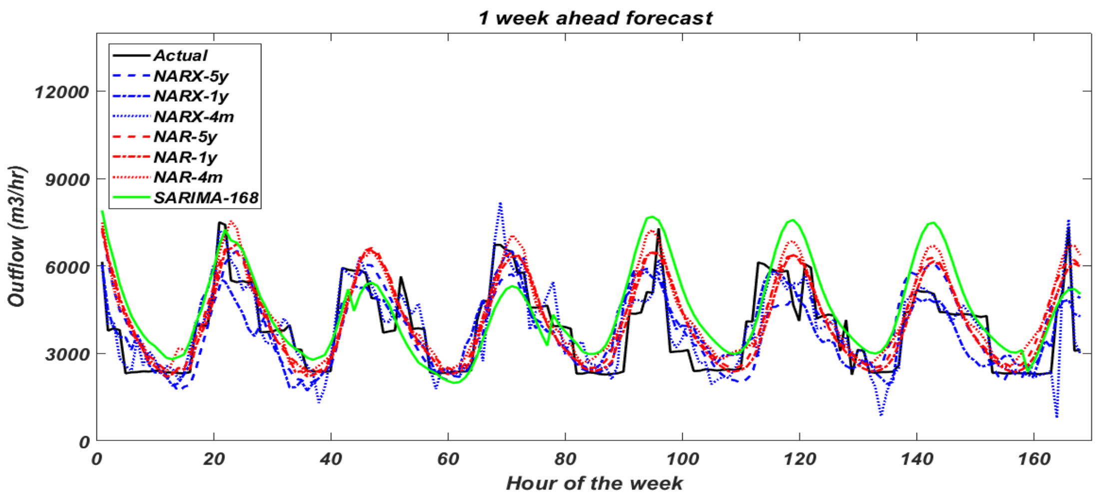

Figure 8. Models forecasting water demand 1 week ahead (12/02/2107 3:00pm – 12/09/2017 2:00pm)

Conclusion

Simulations with linear SARIMA and nonlinear autoregressive ANN models were employed to forecast water demand 24 hours ahead and 1 week ahead. The first objective of this study was to scrutinize the ability of nonlinear ANN models compared to an existing, adequate linear model. The results showed that both ANN models, NARX and NAR, have outperformed the linear SARIMA model. The second objective was to evaluate the efficacy of using exogenous data combined with historical demand. The results show that all models with exogenous inputs (NARX) outperformed those that used only the historical demand (SARIMA and NAR). The third objective was to investigate the influence of historical record length on prediction accuracy. The results showed that predictions based on the four most recent months of data outperformed those trained with five and one continuous years of data. NARX-4m model was shown to have the best performance with the lowest prediction error. However, this model may not be the best choice for all water utilities as the utility studied herein had a rather unique consumer breakdown. Abrupt changes in consumer breakdown and/or demand patterns throughout the year could challenge the NARX-4m model without recalibration.

NARX-4m model has had the best performance amongst the proposed models forecasting 24 hours ahead and 1 week ahead. However, it had 5% of the data with a high over- and underestimated water outflow. This 5% can be further reduced if the change in the outflow could be tracked on a finer level. A 15 minutes or 5 minutes outflow dataset are highly recommended to be used in the NARX-4m model instead of the one hour used in this analysis. Further, the NARX model had a moderate AIC value and training time, compared to SARIMA and NAR models. This was because of its complex structure its utilization of correlated exogenous parameters. It should be noted that AIC and training time values can be reduced through the utilization of a shorter data span.

Considering the outcomes evidenced here, we believe this important multi-dimensional balance between model accuracy and model complexity can be optimized based on a water utility's interest and resource capacity.

Data Availability

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions.

Acknowledgments

The Authors would like to show their gratitude to Mr. Rodney Bouchard, Dale Dillen of Union Water Supply System, and Mr. Klaus Dohring of Green Sun Rising Inc. for their sustained support and superintendence.

Conflicts of Interests

The authors declare no conflict of interest.

References

Adamowski, J. & Karapataki, C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering*, 15(10), 729–743.

Alvisi, S., Franchini, M., Marinelli, A., 2007. A short-term, pattern-based model for water-demand forecasting. Journal of Hydroinformatics 9 (1), 39-50.

Aly, A., and Wanakule, N. (2004). "Short-term forecasting for urban water consumption." J. Water Resour. Plann. Manage., 10.1061/(ASCE) 0733-9496(2004)130:5(405), 405–410.

Amjady, N., Keynia, F., 2011. A New Neural Network Approach to Short Term Load Forecasting of Electrical Power Systems. Energies, 4, 488–503.

Arandia, E., Eck, B., and McKenna, S. 2016. "Tailoring seasonal time series models to forecast short-term water demand." J. Water Res. Plann. Manage. 142 (3): 04015067. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000591.

Azadeh, A., Ghaderi, S., Sohrabkhani, S., 2008. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. Energy Convers. Manag., 49, 2272–2278.

Bakker, M., Van Schagen, K., Timmer, J., 2003. Flow control by prediction of water demand. Journal of Water Supply: Research and Technology - AQUA 52 (6), 417-424

Bennett, C., Stewart, R., Beal, C., 2013. ANN-based residential water end-use demand forecasting model, Expert Systems and Applications 40 (4), 1014–1023.

Billings, B., and Jones, C. 2008. Forecasting urban water demand, 2nd Ed., American Water Works Association, Denver, CO.

Bougadis, J., Adamowski, K., and Diduch, R. (2005). "Short-term municipal water demand forecasting." Hydrol. Processes, 19(1), 137–148.

Chang, P., Fan, C., Lin, J., 2011. Monthly electricity demand forecasting based on a weighted evolving fuzzy neural network approach. Int. J. Electr. Power Energy Syst., 33, 17–27

Demuth H., and Beale, M., 1998. *Neural Network TOOLBOX for use with Matlab*, version 3 Mathworks, Natick, MA, USA.

Donkor, E., Mazzuchi, T., Soyer, R., Roberson, J., 2012. Urbanwater demand forecasting: a review of methods and models. Journal of Water Resources Planning and Management.

Evans, J., 1996. Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Brooks/Cole Publishing.

Feilat, E., Bouzguenda, M., 2011. Medium-term load forecasting using neural network approach. In Proceedings of the IEEE PES Conference on Innovative Smart Grid Technologies—Middle East (ISGT Middle East), Jeddah, Saudi Arabia, 17–20, 1–5.

Gardiner, V., and Herrington, P., 1990. Water demand forecasting, 1st Ed., Spon Press, Norwich, U.K

Ghiassi, G., Zimbra, D.K., Saidane, H., 2008. Urban water demand forecasting with a dynamic artificial neural network model. Journal of Water Resources Planning and Management 134 (2), 138-146.

Guo, G., Liu, S., Wu, Y., Li, J., Zhou, R., and Zhu, X. (2018). Short-Term Water Demand Forecast Based on Deep Learning Method. J. Water Resour. Plann. Manage., 2018, 144(12): 04018076.

Herrera, M., Torgo, L., Izquierdo, J., Pérez-García, R., 2010. Predictive models for forecasting hourly urban water demand. Journal of Hydrology 387 (1-2), 141-150.

Homwongs, C., Sastri, T., Foster, J.W., 1994. Adaptive forecasting of hourly municipal water consumption. Journal of Water Resources Planning & Management - ASCE 120 (6), 888-905.

Jain, A. & Ormsbee, L., 2002. Short-term Water demand forecast modeling techniques — Conventional Methods versus AI. *Journal American Water Works Association*, 94(7), 64–72.

Jain, A., Varshney, A., and Joshi, U. (2001). “Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks.” Water Resour. Manage., 15(5), 299–321.

Jain, A., Varshney, A., Joshi, U., 2001. Short-term water demand forecast modeling at IIT Kanpur using artificial neural networks, Water Resources Management 15 (5), 299– 321.

Jowitt, P.W., Chengchao, X., 1992. Demand forecasting for water distribution systems. Civil Engineering Systems 9 (2), 105-121.

Kandananond, K., 2011. Forecasting Electricity Demand in Thailand with an Artificial Neural Network Approach. Energies, 4, 1246–1257.

Kofinas, D., Mellios, N., Papageorgiou, E., and Laspidou, C. 2014. “Urban water demand forecasting for the island of Skiathos.” Procedia Eng. 89: 1023–1030. https://doi.org/10.1016/j.proeng.2014.11.220.

Mitrea, C., Lee, C. & Wu, Z., 2009. A Comparison between Neural Networks and Traditional Forecasting Methods: A Case Study. *International Journal of Engineering Business Management*, 1(2), 19–24.

Moreno-Chaparro, C., Salcedo-Lagos, J., Rivas, E. y Orjuela Canon, A., 2011. A method for the monthly electricity demand forecasting in Colombia based on wavelet analysis and a nonlinear autoregressive model. En: Ingeniería, Vol. 16, No. 2, 94 – 106.

Mustafaraj, G., Lowry, G., and Chen, J., 2011. Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. Energy and Buildings, vol. 43, no. 6, 1452-1460.

Pisoni, E., Farina, M., Carnevale, C., and Piroddi, L., 2009. Forecasting peak air pollution levels using NARX models. Engineering Applications of Artificial Intelligence, vol. 22, no. 4-5, 593-602.

Rice, D., Carriveau, R., Ting, D., Bata, Mo’tamad, B., 2017. "Evaluation of Crop to Crop Water Demand Forecasting: Tomatoes and Bell Peppers Grown in a Commercial Greenhouse." Agriculture 7, no. 12, 104.

Shumway, R., and Stoffer, D. (2000). Time series analysis and its applications, Springer, Berlin.

Shvartser, L., Shamir, U., Feldman, M., 1993. Forecasting hourly water demands by pattern recognition approach. Journal of Water Resources Planning & Management - ASCE 119 (6), 611-627.

Wilcox, Rand R., 2005. Introduction to robust estimation and hypothesis testing. Academic Press.

Wong, J., Zhang Q., and Chen, Y. 2010. “Statistical modeling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast.” Water Res. Res. 16 (3): 335–343. https://doi.org/10.1016/0043-1354(82)90194-4.

Zhou, S., McMahon, T., Walton, A., and Lewis, J. (2000). “Forecasting daily urban water demand: A case study of Melbourne.” J. Hydrol., 236(3–4), 153–164.

Zhou, S., McMahon, T., Walton, A., and Lewis, J. (2002). “Forecasting operational demand for an urban water supply zone.” J. Hydrol., 259(1–4), 189–202.

CHAPTER 4

Hybrid Self Organizing Map and Regression Tree Short-term Water Demand Forecasting Model

Mo'tamad H. Bata, Rupp Carriveau, and David S.-K. Ting
Turbulence and Energy Laboratory, Ed Lumley Centre for Engineering Innovation, University of Windsor, Windsor, ON N9B 3P4, Canada.

Nomenclature

The following symbols are used in this chapter:

RT	= Regression tree
SOM	= Self-organizing map
ANN	= Artificial neural network
NRMSE	= Normalized root mean squared error
WDS	= Water distribution system
ARIMA	= Autoregressive integrated moving average
SARIMA	= Seasonal autoregressive integrated moving average
MLP	= Multi-layer perceptron
RVM	= Relevance vector machine
N	= Number of neurons
I_n	= Input number n
n	= Data point number (i.e. 1, 2, 3 …etc.)
RMSE	= Root mean squared error
AIC	= Akaike information criterion
m^3/hr	= Cubic meter per hour
am	= Ante Meridiem (i.e. Before midday)
ρ	= Pearson correlation coefficient
d	= The difference between the ranks of the two columns in ρ calculations
K	= The cluster number
2D	= Two dimensional
$\overline{Y}_i$	= Data set mean
$\widehat{Y_i}$	= Forecasted water outflow
Y_i	= Actual water outflow

Introduction

Forecasting water demand is essential for optimal operation of water supply systems. The accurate forecasts of the future demands can enable utilities to supply water at lower costs, with less energy, and lighter loads on the network infrastructure. Improved pressure management and leakage control can also be achieved. These positive outcomes also help water utilities move steps closer to true sustainability. Smart decision making in water distribution systems (WDS) is crucial for water utilities adapting Smart Water Grid platform (Joong K., 2018).

Short-term water demand forecasting can be used to schedule pumping operations, system maintenance, and infrastructure development (Zhou et al., 2002). The accuracy of the forecast output controls the efficiency of the system response. This has been a topic of significant interest to researchers and developers alike. The literature features a broad spectrum of forecasting models. An extensive review of the forecasting models has been presented by (Donkor et al., 2012; House-Peters et al., 2011; Kozlowski et al., 2018). These models can be largely classified into two groups, linear and nonlinear (Zhang G., 2001). Table 1 highlights short-term load forecasting studies. The two groups can be further distinguished into two other groups: solo and hybrid. The solo models are standalone forecasters, where the hybrid ones are a combination of two or more forecasting techniques.

Linear models are used extensively owing to their simplicity and the practicality of the required data acquisition. The ease of implementation and ability to update make these models very attractive. A number of researchers (Hughes T., 1980; Maidment et al., 1986; Maidment et al., 1985; Zhou et al., 2000) have proposed innovative Autoregressive Integrated Moving Average (ARIMA) and univariate time series analysis models to forecast daily urban water demand.

These linear models can be used to forecast water consumption; however, the accuracy of the forecast can be unsatisfactory (Kozlowski et al., 2018).

Nonlinear models are complex to develop, implement, and update. However, their ability to analyze multiple parameters and concurrently find the nonlinearity relations between variables, have make them powerful prediction tools. Artificial Neural Networks (ANN), nonlinear regression models, fuzzy logic, and other nonlinear models are among the most popular for forecasting water demand (Adamowski et al., 2012; Adamowski J. & Karapataki C., 2010; Bennett et al., 2013; Boguadis et al., 2005; Cutore et al., 2008; Ghiassi et al., 2008; Ghiassi et al., 2005; Hippert et al., 2001; Jain et al., 2001; Mitrea et al., 2009; Nasseri et al., 2011; Tiwari M. & Adamowski J., 2015)

Beyond this, researchers have also combined two or more model types to accomplish the water forecasting. (Hiroyuki et al., 2001), presented a hybrid model that consists of fuzzy regression tree and a multi-layer perceptron (MLP) of ANN. The proposed fuzzy regression tree is here employed to reveal rules in the actual data and help to organize the input data into specific classes. The MLP then used to predict the load one step ahead. Fusing the simplified fuzzy method into the regression tree helped to determine the split values. The hybrid model was proven to effectively forecast one step ahead for a power system.

Another hybrid model of Relevance Vector Machine (RVM) and Regression Tree (RT) was proposed by Mori et al., 2011. Based on some similarity of data characteristics, the RT model classified the data into clusters. Then, the RVM model was constructed to predict the load one step ahead in each cluster. The proposed model was used successfully to forecast the electric load in some Japanese utilities.

Researchers have noticed a strong similarity between water and electricity demand patterns, it has also followed that forecasting approaches are similar (Perry P., 1981).

In this paper, a nonlinear hybrid Self Organizing Maps (SOM) clustering model is co-developed with a Regression Tree (RT) forecasting model. This hybrid model was then deployed to forecast utility water outflow 8 hours into the future. A brief description on the two models can be found in *Data and Models* section.

Table 1. Highlights of linear and nonlinear short-term forecasting models

Model Category	**Performance**	**Purpose**	**Reference Number**
Linear	Solo	Forecasting daily urban water demand	Hughes T., 1980; Maidment et al., 1986; Maidment et al., 1985; Zhou et al., 2000
Nonlinear	Solo	Short-term water demand forecast	Adamowski et al., 2012; Adamowski J. & Karapataki C., 2010; Bennett et al., 2013; Boguadis et al., 2005; Cutore et al., 2008; Ghiassi et al., 2008; Ghiassi et al., 2005; Hippert et al., 2001; Jain et al., 2001; Mitrea et al., 2009; Nasseri et

			al., 2011; Tiwari M. & Adamowski J., 2015
	Hybrid	Forecast step-ahead for a power system	Hiroyuki et al., 2001
		Forecast the electric load	Mori et al., 2011

Methodology

Self Organizing Maps (SOM), also known as Kohonen Neural Networks (Kohonen T., 1982), are able to map input data into an N-dimensional grid of neurons. This mapping technique preserves the patterns (i.e. topology) of the input data space. Simply stated, the patterns that are close in the input space will be mapped to units that are close in the output space (i.e. grid) (Bação et al., 2005). Each data point will pass through the set of neurons and only one neuron can win the point. This competition between the neurons is based on how close the data point is to the center of the clusters. After all input data points are processed, a topological map will appear. Figure 1 illustrates the mechanism of a 2-dimentional grid of 2 neurons resulting in 4 clusters. The process begins by assigning the same initial weight to all connections between the inputs (I_1, I_2, …, I_n) and the neurons. Then, the Euclidean distance is computed between all neurons and the node (the input multiplied by the weight). Only one neuron with the shortest distance wins that data input. The rest of the neurons are arranged topologically based on a neighbourhood function. Before moving to the next input, the weights are adjusted according to the previous neighbourhood topology. By the end of the training, the input data is grouped in 4 clusters that have the same topology as the input space. In this paper, the SOM clustering model does not operate as a forecaster. Rather, it serves as an auxiliary component in the hybrid model by decreasing the dimensionality of the target data. Having the target data grouped into fewer clusters assists the forecasting (Regression Trees) model by locating the cluster before predicting the point. As a result, accuracy, speed, and performance of the forecasting model are improved.

Regression Trees (RT) are a supervised machine learning technique that uses neural networks. The models are obtained by repeatedly dividing the data space and fitting a simple prediction model within each division. As a result, the data division can be represented graphically as a decision tree (Loh W., 2011). Model development begins with feeding the input data to the tree root, then the data will be filtered and sent to a branch and then to another branch until it reaches the leaf. The leaf is where the final decision is made, it is called the Response.

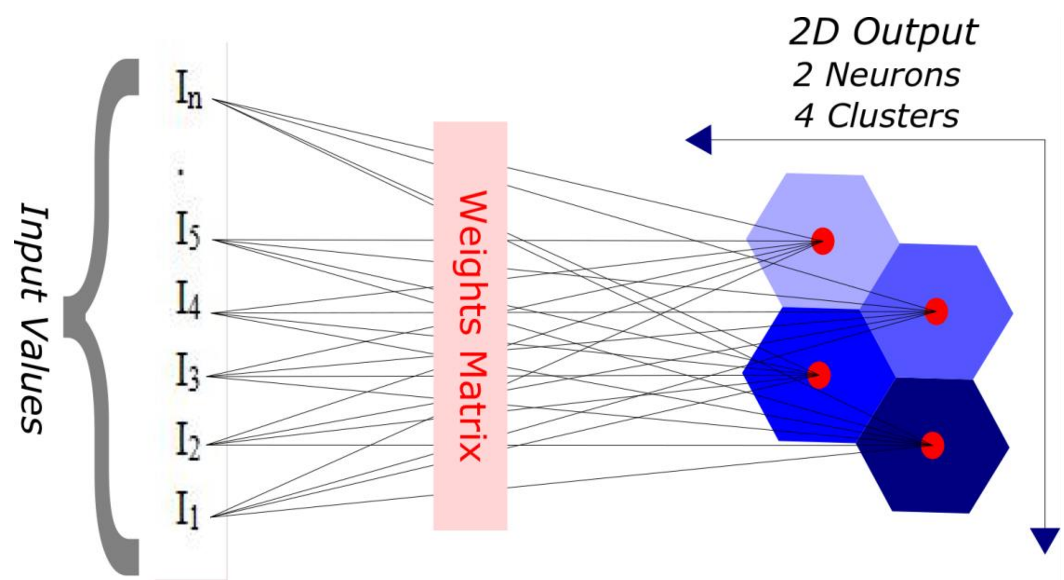

Figure 1. 2-dimentional (2D) SOM structure with number of neurons equal 2.

The practice for the proposed hybrid model is to simply feed the output of the SOM clustering model, accompanied by other desired correlated inputs, to the RT forecasting model. Figure 2 illustrates the hybrid model flowchart forecasting the response time t ahead. The process begins with gathering the required (as available) raw data. Raw data is then pre-processed. Here, all missing, erroneous, and noisy data are imputed and smoothed. After the raw data is processed, the target is isolated and fed to the SOM clustering model. Within the SOM model, the response is grouped into an initial number of clusters. The output of the SOM model is then fed to the RT forecasting model along with the input data. Once all required data is fed to the RT model, the model is trained, tested, and validated. The performance of the model then is assessed. Here the Root Mean Squared Error (RMSE) is calculated between the actual and predicted target values. If the RMSE is satisfactory, a future dataset excluding the target can be fed to the model to forecast t-time steps ahead. More often than not, the initial forecast iterations are not satisfactory. Where this is the case, extra neurons can be added to the SOM clustering model. It is important to note that at some point model performance increases only marginally regardless of how many neurons are added to the model. A final modification, to the RT model, is to increase the cross-validation folds and/or the number of leaves in the RT architecture. Increasing the cross-validation folds can protect the model from overfitting. While increasing the number of leaves will lead to a finer and more flexible tree.

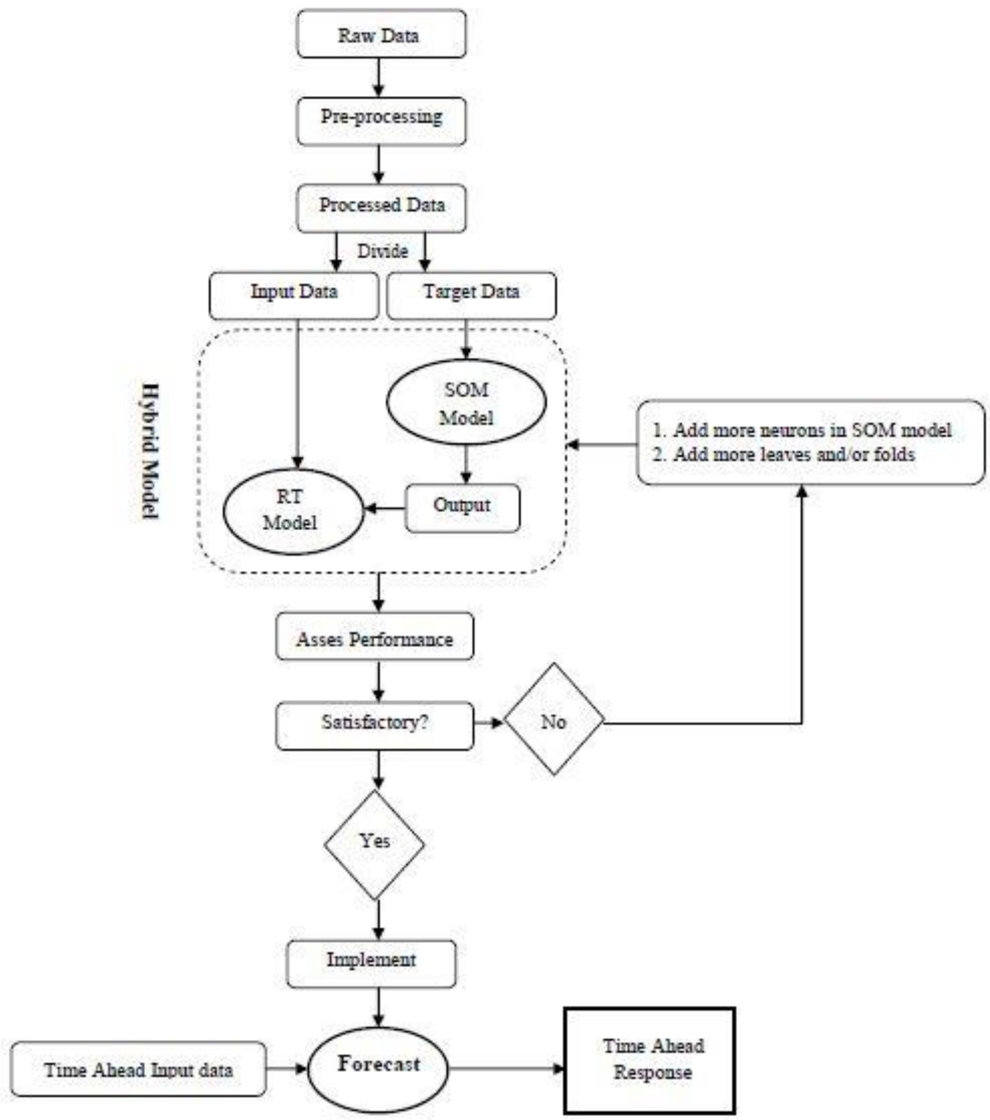

Figure 2. Hybrid model flowchart.

Data and Models

Raw Data

Preliminary study developed two ANN models to forecast the water outflow 24 hours and one week ahead using the same outflow data. Both models revealed that using one recent year of data, divided tri-annually, is easier to handle and more efficient for short-term forecasting; compared to the available five continuous years of data. Therefore, the data used to feed, train, test, and validate the models spans the last tri-annual portion (August – December) of 2017. The data consists of the water outflow in (m^3/hr) (see Figure 3) and other seasonality data, such as, the month of the year (1-12; 1 represents January and 12 represents December), the day of the month, and hour of the day (1:24; 1 represents 12:00 am – 01:00 am). In Figure 3, water demand tends to drop gradually as you move from August (summer) to December (winter). This seasonality varies year to year and also from one season to the other. In addition, there are no specific demand characteristics that can be outlined during the day/night or weekday/weekend. That is because most of the demand, about 80%, is consumed by commercial greenhouses. These commercial greenhouses grow different crops in different seasons and use different techniques in watering their crops.

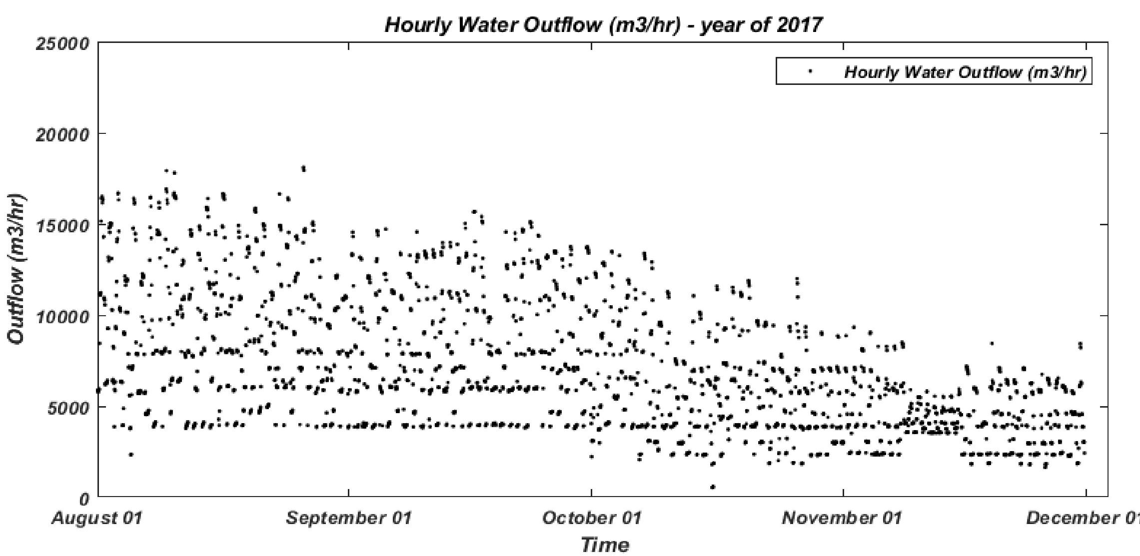

Figure 3. The utility hourly water outflow in m^3/hr.

The strength of the correlation between the water outflow and the predictors is calculated using a Spearman Correlation Coefficient (ρ). Equation 1 describes Spearman Correlation Coefficient, where d is the difference between the ranks of the two columns and n is the number of data points. Here ρ, shown in Table 2, is a statistical parameter that determines the strength of the monotonic relationship between two variables. Inspection of the table reveals that same day previous hour K values have the strongest correlation with the water outflow. However, this predictor is unknown and thus useless if the intended forecasting is for more than one hour ahead. For this reason, other predictors, ranked 2-4, are used in training and could be used in forecasting 24 hours and one week ahead.

$$\rho = 1 - \left[\frac{6\sum_{n=1}^{n=n}(d^2)}{n(n^2-1)}\right] \tag{1}$$

Table 2. Spearman Correlation Coefficient ρ between the utility water outflow and the predictors used in forecasting

Predictor	**ρ**	**Rank**
K[a] same day previous hour	0.868	1
Outflow previous day same hour	0.855	2
Outflow previous week same hour	0.820	3
K previous week same hour	0.781	4
Outflow average previous 24 hours	0.585	5
Month of the year	-0.563[b]	6
Hour of the day	0.181	7
Day of the month	-0.076	8

[a] K is the cluster number extracted from the SOM output

[b] Negative ρ means inverse correlation

SOM Clustering Model

Four of the abovementioned two-dimensional SOM (2D-SOM) were implemented to cluster the data. The four models change with the number (N) of neurons used in the 2D grid layer. For our purposes, N ranged between 2 and 5. For N equal to 3 in a 2D layer, the number of clusters is equal to 9 in total, which is 3 in each dimension. This resulted from calculating the distance between each neuron and its own center. The positions of these neurons are randomly assigned at the outset. Then, based on the input data distribution, they are self-arranged in positions (see Figure 4) that reflect the variance of the input data space. Lastly, each dataset point is entered through the layer of neurons and only one neuron is assigned as a winner for that point. Figure 5 illustrates the output of the SOM for N=2. Four clusters representing four different outflow ranges. For instance, Clusters 1, 2, 3, and 4 here represent, respectively, low, moderate, high, and very high water demand. This representation may change when N changes and it should still reflect a physical characteristic in the WDS. The output of the SOM clustering model is a binary form, 0 or 1. For each data point, one winning neuron receives 1 and the rest of the neurons receive a 0. All data points that were won by the same neuron are clustered together and labeled with the cluster number (i.e. 1, 2, 3, or 4). This pattern is then used as one of the inputs for the Regression Tree (RT) model presented in the *RT Forecasting Model section*.

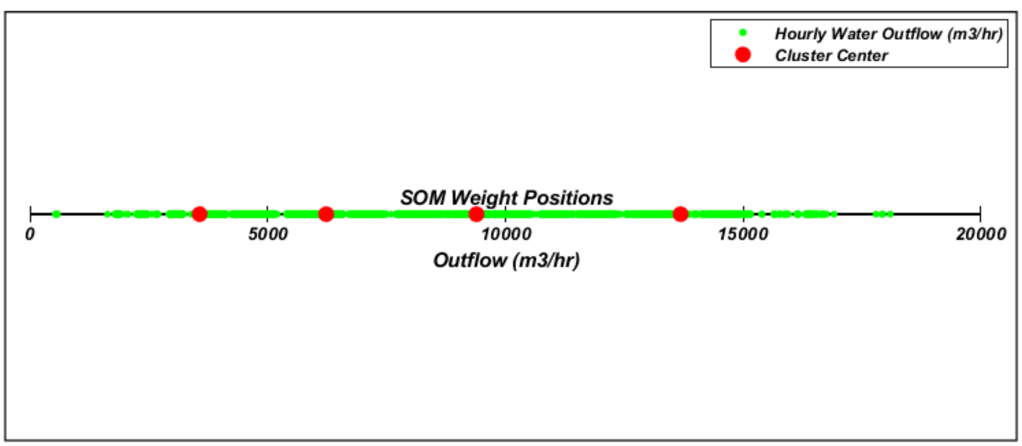

Figure 4. SOM Weight positions – Clusters positions.

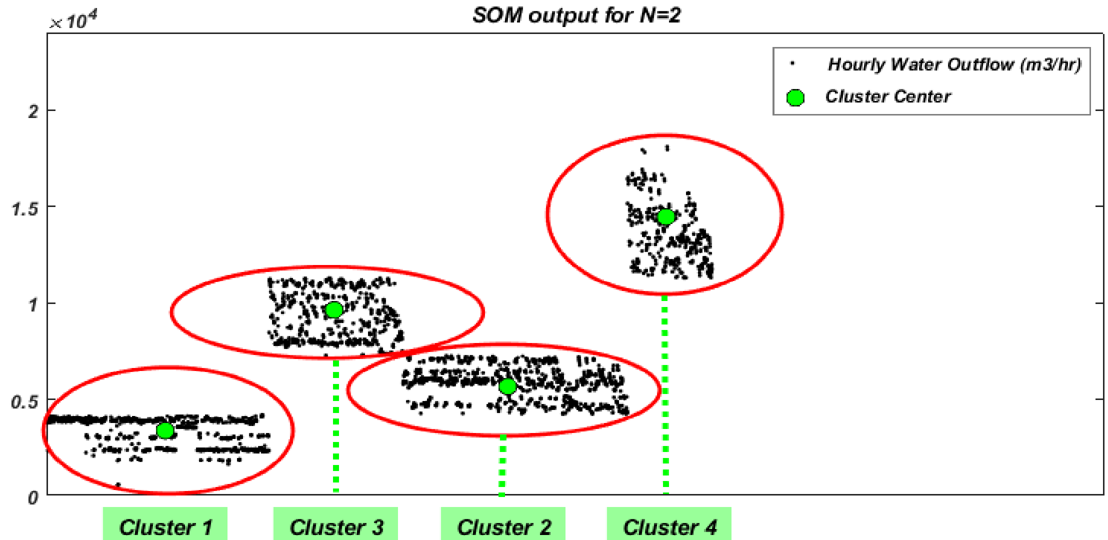

Figure 5. Self-Organizing Maps output where 2 neurons (N) are used

RT Forecasting Model

Four Regression Tree (RT) models were developed to forecast 8 hours ahead. The first model (HYB-1hr) executes the forecast hourly; all predictors are fed to the model every time the model predicts the future outflow demand. The second model (HYB-8hr) forecasts the 8 hours at once using all predictors excluding K same day previous hour. Meanwhile, the third and fourth models (RT-1hr and RT-8hr) do not use any of the SOM output (i.e. no K inputs). All models are identical in terms of the input data time span, the number of tree leaves, and the cross-validation folds.

Model Performance

The predicted outflow was compared to the actual outflow and the performance of the four models was measured with the Root Mean Squared Error (RMSE). However, the results are shown using the Normalized Root Mean Squared Error (NRMSE) because the datasets used in forecasting have different means. Equation 2 represents the NRMSE, where n represents the number of data points; $\overline{Y_i}$ is the data set mean, $\hat{Y}_i$ and Y_i represent the forecasted and the actual water outflow, respectively.

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y_\iota}^2 - Y_i^2)}}{\bar{Y_\iota}} \qquad (2)$$

Results and Discussion

Models Overall Performance

Figure 6 displays all models performance in terms of NRMSE. The figure reveals that RT-1hr has an NRMSE of 0.13 which is approximately double the NRMSE of HYB-1hr. Likewise, the RT-8hr has an NRMSE of 0.21 which is about double the NRMSE of HYB-8hr. Forecasting 8 hours on 1 hour increments using the hybrid model (i.e. HYB-1hr) has outperformed HYB-8hr which forecasts 8 hours ahead on one increment. For all proposed numbers of neurons (N=2, N=3, N=4, and N=5) HYB-1hr showed less error than HYB-8hr. For N=2, the NRMSE for HYB-1hr is 0.087 which is 45% less than 0.159 for HYB-8hr. Also, the NRMSE for HYB-1hr compared to HYB-8hr dropped from 0.141 to 0.067, 0.131 to 0.045, and 0.129 to 0.042 by 52%, 65%, and 67% for N equal to 3, 4, and 5, respectively. Moreover, increasing the number of neurons in SOM models did not significantly improve the performance of the hybrid model when 8 hours were forecasted in one increment. The NRMSE for HYB-8hr dropped only by (2% -11%) from 0.159 to 0.141 to 0.131 to 0.129 for N equal to 2, 3, 4, and 5 respectively. Meanwhile, the NRMSE for HYB-1hr dropped from 0.087 to 0.067 to 0.045 to 0.042 by 23%, 33%, and 7% when N increased from 2 to 3 to 4 to 5.

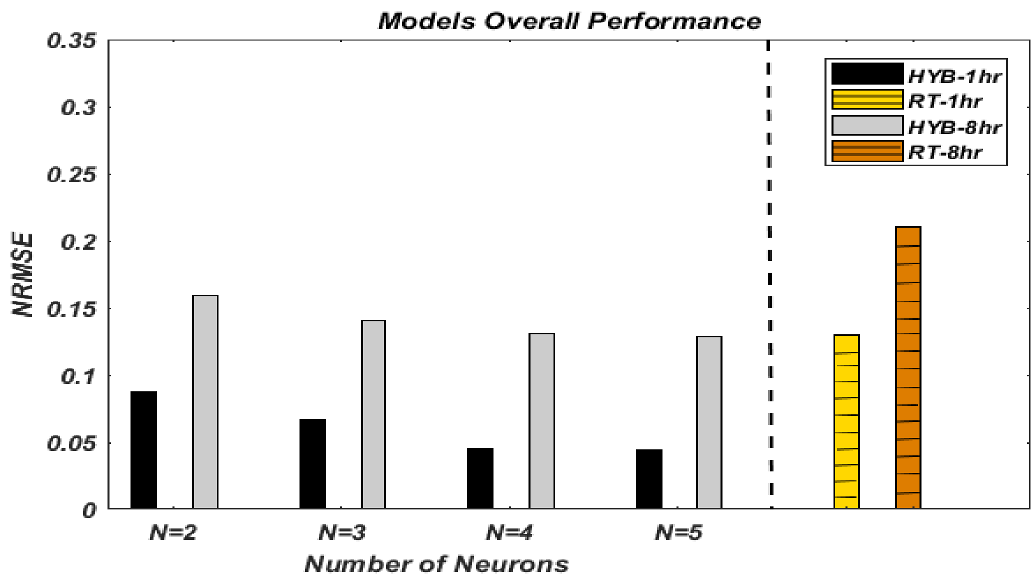

Figure 6. Models overall performance measured in NRMSE.

Model Selection

Relying only on model performance, the HYB model would be the accurate selection for forecasting water demand t-time steps ahead. However, the accuracy in the HYB was gained because of the fusion of two models (i.e. SOM and RT models). Essentially, this fusion has added more parameters and more computations to perform in the model. Therefore, the HYB model is expected to have a higher computational load and more complex architecture. Selecting the higher performance model is not accurate is this case. So, another measure is deemed to add more information in model selection decision. That is the Akaike Information Criterion (AIC). AIC penalizes models for using extra parameters. Hence, models that use more parameters, to gain additional information, have undesired high AIC. Table 3 summarizes the AIC values and time of model training using HP Pavilion TS 14 Notebook PC with a 1.6 GHz Intel Core i5 processor and 8 GB memory.

Table 3. Computation load of forecasting models on training data

Model	Identifier	Indicators	
		AIC	Time (s)
Regression Trees	RT	26,738	41
Hybrid	HYB	29,643	57

Models Application

After the models were trained, tested, and validated on the dataset that contains the water outflow parameter; the models were tested on a held back dataset for a random day on November 2017. Figures 7, 8, 9, and 10 present the results for N equal to 2, 3, 4, and 5, respectively. Analogous to the overall performance, HYB-1hr has shown better fits over the tested time span. It can be proposed here again that when the number of clusters in SOM was increased, the NRMSE performance of HYB-1hr decreases significantly. The results for RT-1hr and RT-8hr were not shown because no SOM output was used for these models.

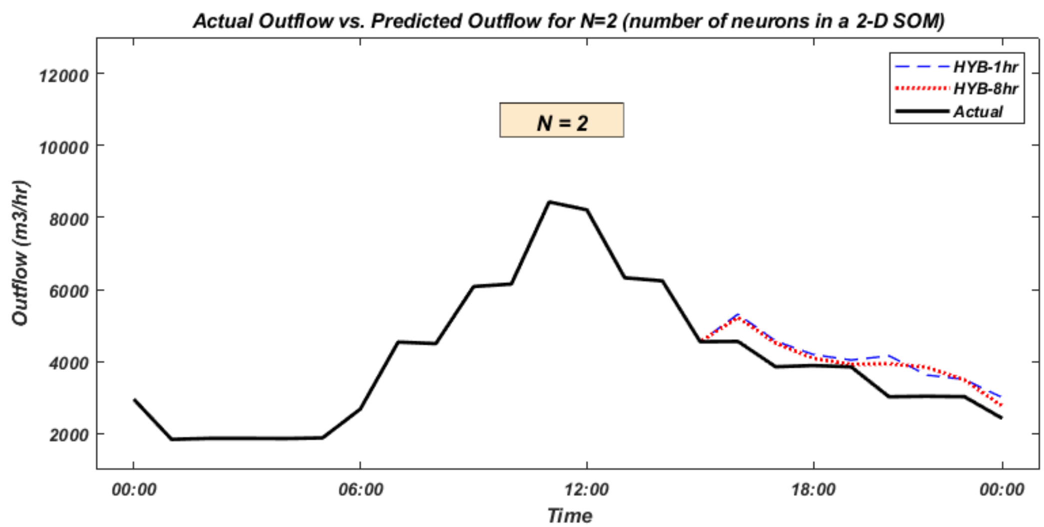

Figure 7. Actual vs. predicted outflow for HYB-1hr and HYB-8hr for N=2.

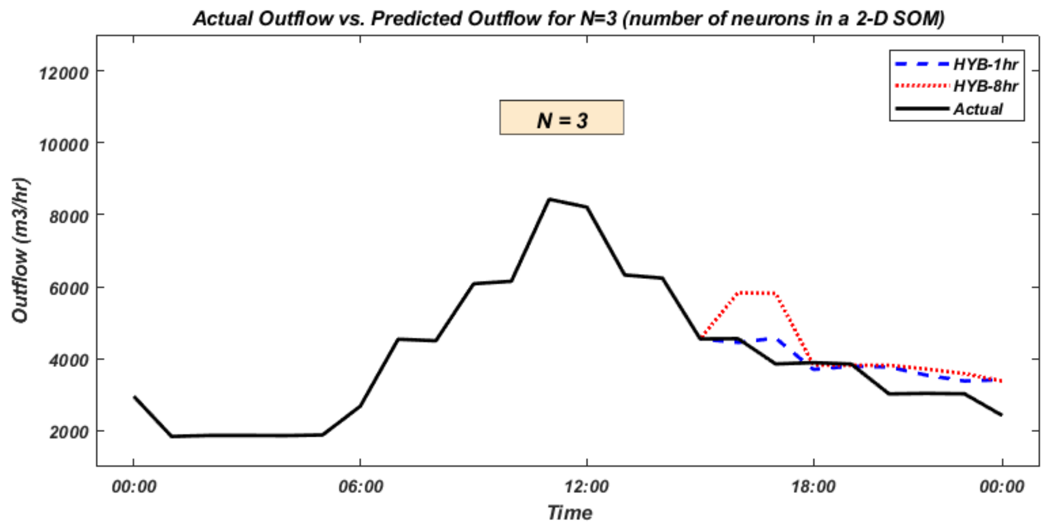

Figure 8. Actual vs. predicted outflow for HYB-1hr and HYB-8hr for N=3.

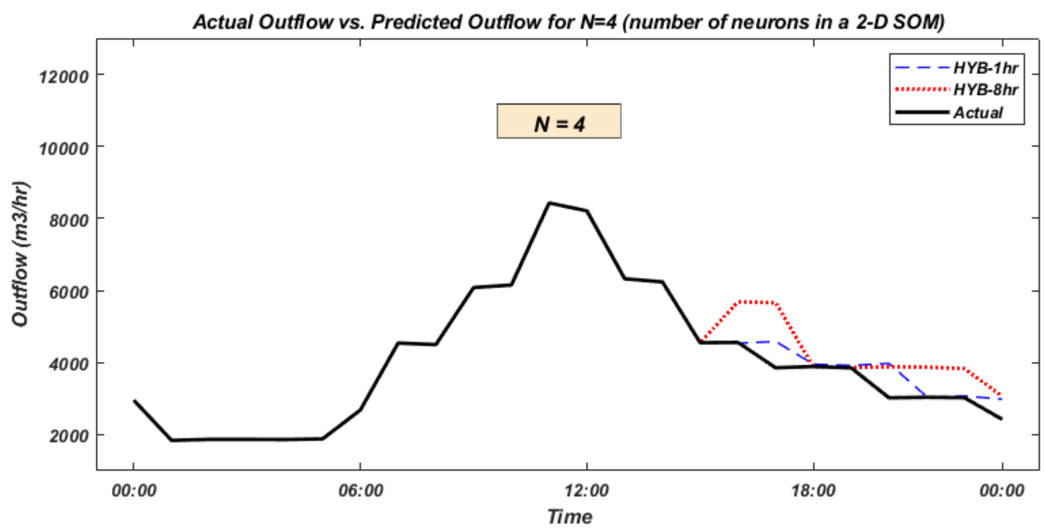

Figure 9. Actual vs. predicted outflow for HYB-1hr and HYB-8hr for N=4.

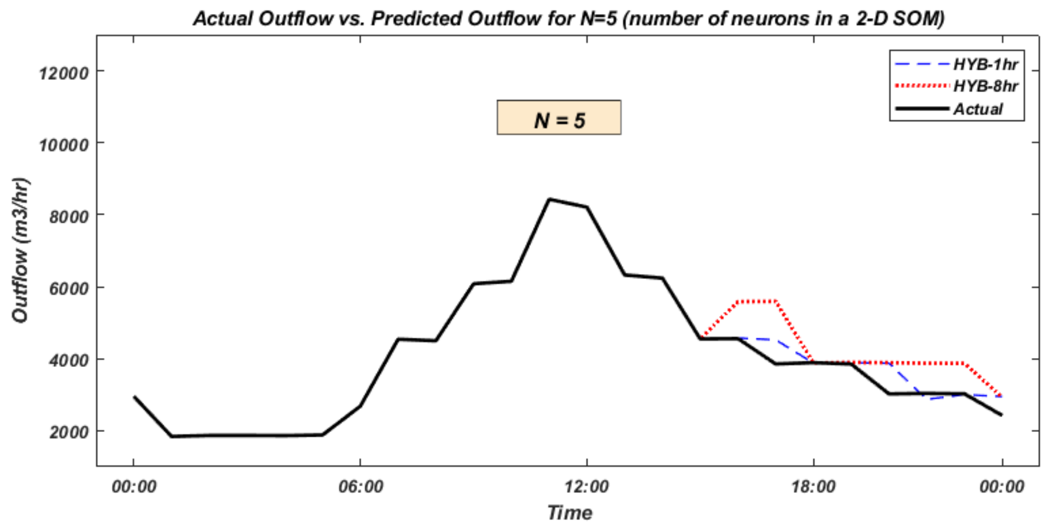

Figure 10. Actual vs. predicted outflow for HYB-1hr and HYB-8hr for N=5.

Concluding Remarks

Four models were presented to forecast a water utility outflow 8 hours into the future. The primary purpose of this paper was to study the influence of combining both supervised and unsupervised machine learning techniques on the performance of a short-term forecasting model. SOM as a cluster model and RT as a forecasting model were integrated to accomplish the forecast for HYB-1hr and HYB-8hr model. Where the RT-1hr and the RT-8hr models were constructed as a standalone only RT supervised model.

The first major takes away from this study was that both hybrid models (i.e. HYB-1hr and HYB-8hr) have shown a better performance than the standalone (i.e. RT-1hr and RT-8hr). The Normalized Root Mean Squared Error (NRMSE) for HYB-1hr and HYB-8hr was shown to be 50% and 35%, respectively, less than the NRMSE for RT-1hr and RT-8hr. Secondly, a significant drop in the NRMSE was noted when more clusters were used in the SOM model. The NRMSE dropped by 50% when the number of clusters increased from 4 (for N=2) to 25 (for N=5). Thirdly, performing the 8 hours forecast on one-hour increments, HYB-1hr and RT-1hr, surpassed HYB-8hr and RT-8hr which execute it in one increment. Finally, the increase in HYB models accuracy compared to RT models was because of the inclusion of SOM clustering model. That has led to a more complex model with an approximate increase of 12% and 35% in AIC value and time of training, respectively.

To conclude, implementing HYB-1hr would be the best of the proposed models. With its unique water outflow, a water utility should prioritize the option of combining a SOM clustering model with a RT forecasting model in order to obtain an accurate forecast. In terms of complexity of implementation, the hybrid model is as easy to develop and implement as any other standalone Artificial Neural Network (ANN) model. This significant increase in modeling accuracy could help the water utility improve operational efficiencies and infrastructure reliability. Specifically, sudden frequent peaks could be avoided, leading to energy, water, and maintenance conservancy. Energy conservation achieved through the reductions in friction and minor losses. The conservation of water would be attained where unnecessary higher pressures cause or exacerbate leaks leading to waste. This same flattening of pressure peaks will also contribute to lighter system mechanical loads, which should support a reduced maintenance requirement.

Acknowledgements

Authors would like to express their sincere appreciation to Rodney Bouchard of Union Water Supply System, and Klaus Dohring of Green Sun Rising Inc. for their perpetual succor.

Conflicts of Interests

The authors declare no conflict of interest.

References

Adamowski J. & Karapataki C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. Journal of Hydrologic Engineering, 15(10) 729–743.

Adamowski J., Chan H.F., Prasher S.O., Ozga-Zielinski B., Sliusarieva A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban demand forecasting in Montreal, Canada, Water Resources Research 48 (1) W01528.

Bação F., Lobo V., Painho M., 2005. Self-organizing maps as substitutes for k-means clustering, Lecture Notes in Computer Science, V. S. Sunderam, G. v. Albada, P. Sloot, and J. J. Dongarra (Eds.), Vol. 3516, PP. 476-483, ISSN 0302-9743, Berlin Heidelberg, Springer-Verlag.

Bennett C., Stewart R., Beal C., 2013. ANN-based residential water end-use demand forecasting model, Expert Systems and Applications 40 (4), 1014–1023.

Boguadis J., Adamowski K., Diduch R., 2005. Short-term municipal water demand forecasting, Hydrological Processes 19 (1) 137–148.

Cutore P., Campisano A., Kapelan Z., Modica C., Savic D., 2008. Probabilistic prediction of urban water consumption using the SCEM-UA algorithm, Urban Water Journal 5 (2)125– 132.

Donkor E., Mazzuchi T., Soyer R., Roberson J., 2012. Urban water demand forecasting: a review of methods and models. Journal of Water Resources Planning and Management.

Ghiassi M., Saidane H., Zimbra D.K., 2005. A dynamic artificial neural network model for forecasting time series events, International Journal of Forecasting 21 (2) (2005), 341-362.

Ghiassi M., Zimbra D.K., Saidane H., 2008. Urban water demand forecasting with a dynamic artificial neural network model. Journal of Water Resources Planning and Management 134 (2), 138-146.

Hippert H. S., Pedriera V, Souza R. C., 2001. Neural networks for short-term load forecasting: a review and evaluation, IEEE Transactions on Power Systems 16 (1), 44-55.

Hiroyuki M., Noriyuki M., Kenta, I., 2001. Short-term load forecasting with fuzzy regression tree in power systems. IEEE International Conference on Systems, Man and Cybernetics, October 7, - October 10. Institute of Electrical and Electronics Engineers Inc., pp.1948-1953.

House-Peters L.A, Chang H., 2011. Urban water demand modeling: review of concepts, methods, and organizing principals. Water Resources Research 47 (5).

Hughes T., 1980. Peak period design standards for small western U. S., Water Supply 16 (4) 661–667.

Jain A., Varshney A., Joshi U., 2001. Short-term water demand forecast modeling at IIT Kanpur using artificial neural networks, Water Resources Management 15 (5) 299– 321.

Joong K., 2018. Accelerating our current momentum toward smart water, Smart Water 3:2.

Kohonen T., 1982. Clustering, Taxonomy, and Topological Maps of Patterns. Proceedings of the 6th International Conference on Pattern Recognition.

Kozlowski E., Kowalska B., Kowalski D., 2018. Water demand forecasting by trend and harmonic analysis. Archive of Civil and Mechanical Engineering 18 140-148.

Loh W., 2011. Classification and regression trees. Wiley interdisciplinary reviews: data mining and knowledge discovery, 1(1):14–23.

Maidment D., Miaou S., 1986. Daily water use in nine cities, Journal of Water Resources Planning and Management. 110 (1) 90-106.

Maidment D., Miaou S., Crawford M., 1985. Transfer function models of daily urban water use, Water Resources Research 21 (4) 425–432.

Mitrea C., Lee C., Wu Z., 2009. A Comparison between Neural Networks and Traditional Forecasting Methods: A Case Study. International Journal of Engineering Business Management, 1(2), 19–24.

Mori A., and Takahashi A., 2011. Hybrid intelligent method of relevant vector machine and regression tree for probabilistic load forecasting. 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies, ISGT Europe.

Nasseri M., Moeini A., Tabesh M., 2011. Forecasting monthly urban water demand using extended Kalman filter and genetic programming, Expert Systems and Applications 38 (6) 7387–7395.

Perry P., 1981. Demand forecasting in water supply networks, Journal of Hydraulic Division 107 (9), 1077-1987.

Tiwari M., Adamowski J., 2015. Medium term urban water demand forecasting with limited data using an ensemble wavelet bootstrap machine-learning approach, Journal of Water Resources Planning and Management 141 (2) 04014053.

Zhang G., 2001. An investigation of neural networks for linear time-series forecasting. Computers and Operations Research 28 (12) 1183-1202.

Zhou S., McMahon T.A., Walton A., Lewis J., 2002. Forecasting operational demand for an urban water supply zone, Journal of Hydrology 259 (1–4)189–202.

Zhou S., McMahon T.A., Walton A., Lewis L., 2000. Forecasting daily urban water demand: a case study of Melbourne, Journal of Hydrology 236 (3–4) 153–164.

Chapter 5

Conclusions and Recommendations

Summary

This study aimed to develop short-term water demand forecasting models for a water distribution network. These models were intuitively constructed to assist the water utility upgrading towards a 2-way SWG. Models were selected to further investigate performance versus complexity and practicality. Previous work, C2C, have focused on the utility's major end-user, the commercial greenhouses. C2C suggested forecasting one crop water demand based on another's. The significance of C2C is that instead of developing a very complex and accurate model for all consumers, only one crop's consumption is forecasted accurately, and other crops consumption could be forecasted based on that one crop. Two simple models, LR and QM, were deployed to forecast the tomatoes water consumption based on the pepper's and vice versa. Again, the base model is required to be accurate regardless its complexity. So, ANN models were developed as optimal base models. With the similarities between the two crops, the results show an average increase in tomatoes forecast error of 29% and pepper forecast error of 12% when forecasting with base model data compared to real data. The disparity in the error increase is caused by a larger error in the tomatoes base model. The average increase of 12% might not seem significant. However, when compared to the base model, the increase in error ranged between 75% and 126%. This indicates that the ANN complex base model outperformed C2C simple models by a large margin.

This study proposed two groups of models. The first group consists of SARIMA linear model, and NAR and NARX nonlinear models. This group was employed to study the effect of model architecture, historical data size, and exogenous parameters on forecasting performance. The forecast of 24 hours ahead and 1 week ahead revealed: (1) nonlinear models outperformed the linear SARIMA model. An average drop in overall performance of 3%, 15% and 33%, 36% when NAR and NARX were deployed to forecast both horizons, respectively. That is because nonlinear models can capture better the time series nonlinearity and have a complex multi-layer structure that finds correlation between input, feedback, and output parameters. (2) 4 months of historical data was adequate to train all models. Results show an error decease of 20% and 15% in NAR model and by 76% and 68% in NARX model when compared to 5 years and 1-year data, respectively, forecasting 24 hours ahead. And also, by 35% and 33% in NAR model and by 70% and 59% in NARX model, respectively, forecasting 1 week ahead. Although time series models seem to perform better with more historical data, however, this was not the case in this research. Both NAR and NARX models had a drastic error decrease when fed with less historical data. The reason is due to the data itself instead of the model structure. A major change in water consumption profiles had occurred in the region where the utility is located. Greenhouses, the main consumer, started to switch from growing vegetables to Marijuana after the new Canadian legalization act occurred in 2018. The change in agricultural activities in the studied area had affected water demand profiles and left the historical data with little information to add. Another important observation is the dramatic error decrease in NARX model compared to NAR model. (3) Adding exogenous parameters to the nonlinear model has improved overall forecasting performance. NARX average overall performance in terms of error, MAPE, decreased by 30% and 25% compared to NAR models forecasting 24 hours ahead and 1 week ahead, respectively. NARX average error also decreased, as mentioned before, by 30% and 36% compared to SARIMA models. Even though NAR and NARX models have a similar structure, the inclusion of exogenous parameters has advantaged the ANN model. Again, this is due to extra correlations drawn from the extra inputs.

The second group comprises of RT model and a hybrid model. This group of models was employed to investigate the fusion of supervised and unsupervised machine learning models. The HYB model consisted of SOM classification model and RT forecasting model. The results show an error decrease of 50% when HYB model is compared to RT standalone model forecasting one hour and eight hours ahead. In addition, HYB model had an error drop in the range of 45% and 67% when forecasting 1 hour ahead compared to 8 hours ahead for different number of neurons in the SOM architecture. Moreover, increasing the number of neurons (from N=2 to N=5) in SOM fused model had decreased error by 25% to 50% forecasting 1 hour and 8 hours ahead.

In regard to practicality, C2C models are simple to develop, train, and deploy. However, this group of models relies on a complex model as a base model, which indirectly increases the complexity of C2C. Also, the base model has forecasting error to begin with, which makes C2C more vulnerable and unstable.

SARIMA models are also simple to develop, train, and deploy. SARIMA has a relatively low AIC values (ranges between 2,000 and 10,000) and short training time (less than 40 seconds). This set of models, however, has a moderate forecasting accuracy (15%-25% forecasting error).

ANN models, specifically NAR and NARX, are complicated, and hard to develop and train. NAR and NARX models have high forecasting capacity (5%-15% forecasting error). However, these models have a high AIC values (ranges between 20,000 and 200,000) and longer training time (1-3 minutes).

HYB model is moderately hard to develop, train, and deploy. HYB model has a moderate AIC value (10,000 – 30,000) and a moderate training time (40 - 60 seconds). And also, have a high forecasting accuracy (5% - 10% forecasting error).

Conclusion

The main conclusions that can be taken from this study are:

- C2C forecasting methods are simple but have a high forecasting error
  - C2C methods are highly susceptible to base model error
  - No significant difference in performance between LR and QM methods
- SARIMA models are simple and have a moderate forecasting error
  - 7 days of hourly data is enough to train SARIMA and forecast 24 hours ahead
  - SARIMA is a stable model for short-term water demand forecasting application
  - SARIMA has a low AIC values and short training time
- NAR and NARX show a high forecasting capacity
  - Observed exogenous parameters can improve their forecasting accuracy
  - The model architecture is complicated
  - Four months of hourly data is enough to train a model
  - High AIC values and relatively long training time is expected
- HYB model has a very high forecasting accuracy
  - The increase in model performance is due to adding the SOM unsupervised classification model.
  - HYB model performance can be improved by adding more neurons. However, the model reaches a steady point where adding more neuron would not improve performance. Rather, it will complicate its structure.
  - Four months of hourly data is enough to train the HYB model
  - Has moderate AIC values and relatively moderate training time.

- The use of any of the proposed short-term forecasting models will provide a significant improvement to the current reactive method at UWSS and other water utilities. Improvements will include:
    - Increase in operational efficiencies and system reliability
    - Avoiding sudden demand peaks
    - Reduce in system mechanical loads and maintenance requirements
    - Better utilization of system storage units; such as: reservoirs, towers, and elevated tanks.
    - Better assessment of the utility's capacity of accepting additional demand.

Overall, the proposed short-term water demand forecasting models will help utilities in water and energy conservancy, more reliable grid, and more sustainable system. This is attained though flattening of pressure peaks. Anticipating future demand helps in avoiding demand peaks which leads to less flow and pressure in the system. As a result, friction losses are reduced, and system leaks are minimized.

Recommendations

This work represents developing short-term water demand forecasting models. Models aimed to study both the end-user and the water utility. Data collected during this study was from a single end-user greenhouse operation and a single utility's water distribution network. It would be sagacious to analyze other datasets from different end-users and water utilities to determine if the same methods and models can be used. It would also help if different scales were used to generalize the findings of this study. These scales include, but not limited to, different utility size, different consumer breakdown, different data span (i.e. 15 minutes or 5 minutes data).

Since this study focused on investigating different models performance, applications in the MATLAB platform were used to develop the models. The architecture of these models allows further functionalities, different data pre-processing, and different data divisions that were not optimized. Further optimization would help in verifying the performance of these models. It is also recommended for water utilities to use more than one flow meter in their water distribution network. This would allow to granularly investigate the network and assign proper models for each main. In addition, post evaluation of model accuracy should be undertaken as it is difficult determine the reliability of the proposed methods when they are only being compared with the hold back data.

Finally, a water utility can determine which model to implement based on the utility's interests and capacity. Small size utilities may focus more on model simplicity, where large size utilities should prioritize model accuracy. This model selection criterion should take in consideration the monetary aspect where savings through the model application is compared to implementation costs.

APPENDICES

Appendix A: Permissions for Previously Published Work

CHAPTER 2: Evaluation of Crop to Crop Water Demand Forecasting: Tomatoes and Bell Peppers Grown in a Commercial Greenhouse

As specified by MDPI Publishing Rights and Permissions[1], the author retains the right to include the article in a thesis if proper attribution and credit is given to the published work.

CHAPTER 3: Short-term Water Demand Forecasting Using Nonlinear Autoregressive Artificial Neural Networks (ANN)

The article has not yet been accepted for publication, so no permission is required. Upon acceptance, the author retains rights to include the article in a thesis, as per American Society of Civil Engineers (ASCE) Reuse Authors Own Material[2].

CHAPTER 4: Hybrid Self Organizing Map and Regression Tree Short-term Water Demand Forecasting Model

The article has not yet been accepted for publication, so no permission is required. Upon acceptance, the author retains rights to include the article in a thesis, as per American Society of Civil Engineers (ASCE) Reuse Authors Own Material[2].

---

[1] https://www.mdpi.com/authors/rights (Accessed Mar 19, 2019)

[2] https://ascelibrary.org/page/informationforasceauthorsreusingyourownmaterial (Accessed Mar 19, 2019)

VITA AUCTORIS

NAME:	Mo’tamad H. M. Bata
PLACE OF BIRTH:	Saudi Arabia
YEAR OF BIRTH:	1991
EDUCATION:	Hajjah Secondary School, Qalqiliah, Palestine, 2009
	An-Najah Nation University, B. A. Sc., Nablus, Palestine, 2013
	University of Windsor, Honor Certificate, Windsor, ON, 2016