Validation of artificial neural networking using household energy consumption in a Toronto high-rise multi-unit residential building

CCTC 2013 Paper Number 1569716737

M. Roque[1,*], S. Prez[2], A. S. Fung[1] and V. Straka[1]

[1] Ryerson University (350 Victoria Street, Toronto, Ontario Canada M5B 2K3)
[2] Icam (6 rue Auber, 59000 Lille, France)
* Author to whom correspondence should be addressed; Tel.: 416-979-5000 ext. 4917; Fax: 416-979-5265

Abstract

An ANN model has been developed in order to predict monthly energy consumption in a Toronto high-rise multi-unit residential building (MURB). The ANN results obtained a R-squared of 0.89 for all phases (training, validation, and testing. A validation set was created (19% of the total dataset) and showed a R^2 (correlation coefficient) of 0.94 using the trained model. The household energy consumption predictions were compared to the 1152 cases (whole dataset) used for developing the model. Comparing predicted and actual household energy consumption, the results of this study identified an error of 0.37% and a R^2 of 0.90.

Keywords: Multi-unit residential building (MURB), artificial neural networking (ANN), validation.

Résumé

Un réseau de neurones artificiels a été développé pour prédire la consommation énergétique mensuelle dans un bâtiment à logement multiple de la ville de Toronto. Le modèle a obtenu un R^2 (coefficient de corrélation) de 0,89 pour toutes les phases (entrainement, validation, test). Une partie validation a été créée (19% de toutes les données) and a décelé un R^2 de 0,94 en utilisant le modèle entrainé. La consommation énergétique des locataires a été comparée au 1152 cas du réseau utilisé pour développer le modèle. Cette comparaison a donné une erreur de 0,37% et un R^2 de 0,90.

Mots clés: Habitat à logement multiple (HLM), réseau de neurones artificiels, validation.

1. Introduction

A way of evaluating the interrelated effects of occupant's household energy use in high-rise MURBs is by mathematical modeling. Mathematical modeling creates a function in order to represent and understand the observations [1]. According to [1], the process of developing these mathematical models has the following steps:

1. Present the problem as simply as possible (*e.g.,* data analysis).
2. Derive reasonable models.
3. Identify the optimal model that represents the observations.
4. Demonstrate the advantage of the model by deriving valuable conclusions.

Canadian residential energy consumption is increasing as well as greenhouse gas emissions. These increasing conditions are highly associated with energy supply and global warming. As these conditions continue to increase, mathematical modelling is highly beneficial in order to further understand household energy use in the residential sector. Household energy use and behaviour is highly complex and differs between occupants [2]. Therefore, creating a model representing household's energy use in high-rise MURBs is very useful in order to explain the residential energy consumption.

2. Overview of Energy Modeling

According to [3], energy modeling is categorized in two ways - top-down and bottom-up. Top-down models predict various factors to represent the entire residential sector or the entire building. These models are at an aggregated level and typically represent energy consumption at a national scale [4]. On the other hand, bottom-up models are at a smaller scale where the model estimates various factors, including energy consumption, but represents individual occupants and/or households. Figure 1 is taken from [3], which illustrates modeling techniques used in residential energy consumption modeling.

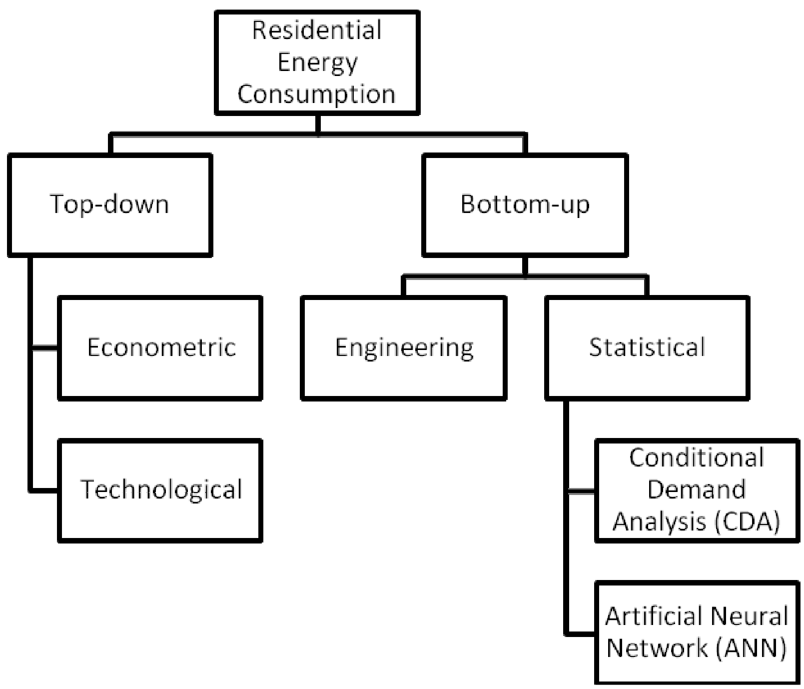

Figure 1: Top-down and bottom-up modeling techniques used for residential sector energy consumption modeling [3].

Top-down models include econometric and technological models. Econometric models include models based on price and income, and technological models include characteristic models representing the entire housing stock [3]. On the other hand, the bottom-up model approach is made up of disaggregated data, such as individual end-use household data and can be categorized further into two bottom-up model methods, statistical and engineering modelling (Figure 1). Engineering models account for energy consumption of end-uses dependant on explicit thermodynamic relationships, heat transfer, equipment use and power ratings [3]. Statistical bottom-up modeling includes historical information and regression analysis. Statistical modeling can also characterize one's household energy use by end-uses and demographics [3].

The bottom-up modelling technique can develop energy consumption models deriving from various factors such as weather conditions, occupant's energy behaviours, occupant's demographics, and building envelope. There are two bottom-up modeling approaches used to evaluate Canadian's energy consumption using end-use, geographical, and demographical data; they are conditional demand analysis (CDA) and artificial neural network (ANN). There are many advantages of bottom-up statistical modeling. First, bottom-up statistical modelling, CDA and ANN, have the capabilities to predict end-use energy consumption [4]. Second, these modeling techniques can include demographics, technical, or energy behaviour elements [3]. Lastly, a model can be created using just energy billing and basic survey information [4].

Some disadvantages, however, are in order to get energy consumption predictions, both methods require a large sample size/dataset. These models rely on historical energy consumption data and information in order to correlate relationships between dependant and/or independent variables [4]. Acquiring a large amount of data is significant in order to increase the confidence of the model's predictive capability [5]. Data (*e.g.*, survey data or energy consumption), however, is usually protected by privacy laws or can only be released by household's consent [5]. A disadvantage, therefore, is modeling aggregated data to represent a small enough region while preserving household's privacy rights and creating meaningful links between household's energy behaviour and their consumption [5].

The objective of this study is to develop a model using the bottom-up method called ANN. The model is then validated for its effectiveness of predicting household energy consumption by using two methods. First, a validation dataset is created and processed through the developed model. Second, predicted household energy consumption from the model is compared to the actual energy consumption and presented percent difference and graphically.

3. Energy Modeling using ANN

The concept of ANN was initially introduced by McCulloch and Pitts in 1943 alongside their McCulloch-Pitts model. The ANN approach is inspired by biological neurons and contains multiple layers of computing nodes [6]. To date, [7-10] are the most comprehensive ANN modeling research on Canadian residential energy use. Two sources of data for developing the ANN model [7-10]: the data from the 1993 Survey of Household Energy Use (SHEU) database and weather conditions for 1993 from Environment Canada. Similar to SHEU 2007, SHEU 1993 consists of detailed information about Canadian households, including house construction, space heating and cooling equipment, household appliance types and usage for 8767 households in Canada. Electricity billing data were also available for the entire year for 2050 households.

An assessment of the effects of socio-economic factors on household energy consumption using neural networks was investigated [8-9]. General trends are found in all three cases. Firstly, as income increases, so does energy consumption. Secondly, owners consumed more energy than renters. Thirdly, dense populations (urban) consumed less energy than less dense areas. In the ALC model, single detached houses consumed more energy than single attached houses [8].

In developing an ANN model, the dataset is divided into training, validation, and testing datasets; where more than two-thirds used for training [11]. Afterwards, the datasets are preprocessed and scaled. Numeric inputs are left numerically. Categorical inputs can be scaled numerically or nominally. Activation functions are then used for both the hidden layers and output layer. Activation functions are needed to introduce nonlinearity into the network [7].

Nonlinearity makes ANN MLPs more effective in discovering interconnected relationships between the inputs. The most common activation functions are logistic and tangent functions. Activation and scaling intervals should be chosen after testing various combinations [12].

To determine the architecture of the network (number of hidden layers), there are no rules to establish the number of hidden layers for a particular application [12]. This is decided by trial and error. Determining the best architecture of the network can be done in two ways. First, the architecture can be done manually by selecting a number of hidden layers and comparing its performance to other architectures. The other way is to do an exhaustive search, which seeks for all possible architecture within the given inputs and hidden layers. This method is very time consuming; however, all possible architectures are presented.

During the training phase of the neural network, weights are distributed randomly and learning algorithms are applied. During the training phase, a great deal of trial and error is required to find the best performing learning algorithm and least sum of square of errors (SSE). Training stops until the least SSE or a high R-square (R^2) is achieved. R-squared is a statistical measure of how well the network outputs actual target values; a R-squared is to 1 indicates it is a perfect fit [11]. A validation set, a dataset that has never been introduced to the network, is used to reinforce the prediction performance of the network.

To further test its prediction performance, a third dataset, called the testing dataset, is applied. Similar to the validation set, the testing dataset is foreign to the trained network. The testing dataset is used after training and validation and its network output is then scaled to get its original units [12].

In order to assess the ANN model prediction performance, [7] and [13] used fraction of variance, which is symbolized as R^2. Alyuda NeuroIntelligence 2.2 is used to develop the ANN model and for simulation in this work [6]. This software has a friendly user interface and a number of options for the different stages of the analysis.

4. Methodology

4.1. Overview

This section begins with the information on the sources of data used to develop the ANN model. It continues with the procedure used to predict household energy consumption.

4.2. Sources of Data

Three sources of data were used for the development of the input units of the ANN models: First, a survey of household energy use in the Toronto MURB [14]; Second, household energy consumption of each unit; lastly, weather conditions obtained from Environment Canada's National climate data and information archive [15]. The output unit of the model was monthly energy consumption per unit for a duration of 24 months.

The survey data consisted of 48 survey responses, completed by occupants residing in a Toronto MURB. The survey consisted of a total of 51 questions related to occupant's space heating/cooling equipment, household appliances, and socio-economic characteristics. With approval from Research Ethics and the Toronto MURB occupants, the survey was conducted for a duration of one month.

Monthly energy consumption for all of the units residing in the Toronto MURB was also obtained from 2010 to 2012. The units for the energy consumption are the amount of electricity used in a month (kWh/month) - first day of the month to the last day of the month.

Weather data conditions include monthly mean temperature (℃), the monthly total of precipitation (mm), the mean direction (10's deg) of gust, and the mean speed of gust (km/h). A total of 24 months of weather data was used to develop the dataset.

4.3. ANN Model

The ANN Model developed in this study consists of one network. This network is used to predict a single output, which is the monthly end-use electrical energy consumption. The development of the model is depicted in Figure 2.

4.3.1. Development of the Network Dataset

In total, 1152 cases were made - 24 months of energy billing data and 48 completed surveys. These cases were then partitioned into three parts:

1. 65% for training the network - 750 cases
2. 19% for validating the network - 218 cases
3. 16% for testing the network - 184 cases

4.3.2. Development of Input and Output Units

The different inputs and outputs used for the ANN model are:

Inputs:

24 categorical columns: Month, Year, Orientation, Gender, Age, Grow Up, Residency, Hours per day, Income, TV – Use, TV – Type, Cable Box Use, Stove Use, Oven Use, Microwave Use, Computer Use, Computer – Age, Internet Use, On an average day, how many light bulbs are turned on longer than 3 hours or more?, Winter – Hours per day light bulbs turned on, Summer - Hours per day light bulbs turned on, Winter temps, Summer temps, How satisfied are you with the appliances in your apartment (i.e. stove, refrigerator, etc.).

10 numeric columns: Number of people per household, Mean Temp (℃), Sum Total Precipitation (mm), Mean Dir of Max Gust (10's deg), Mean Speed of Max Gust (km/h), Electronic ratio, CFLs, Incandescent, Number of light fixtures, Energy consumption behaviour ratio.

Output: energy consumption (kWh/month)

4.3.3. Analysis and Preprocessing

Once the dataset was developed, it was then loaded into the ANN software. The data were then analyzed for any missing and wrong entered data, and/or outliers. All data inputs, whether numerical or categorical, were numerically scaled (and encoded) between -1 and 1. This means that a column with N distinct categories (values) was encoded into one numeric column, with one integer value assigned for each category. Orientation column with values "East" and "West", for example, will be represented as -1 (East) and 1 (West).

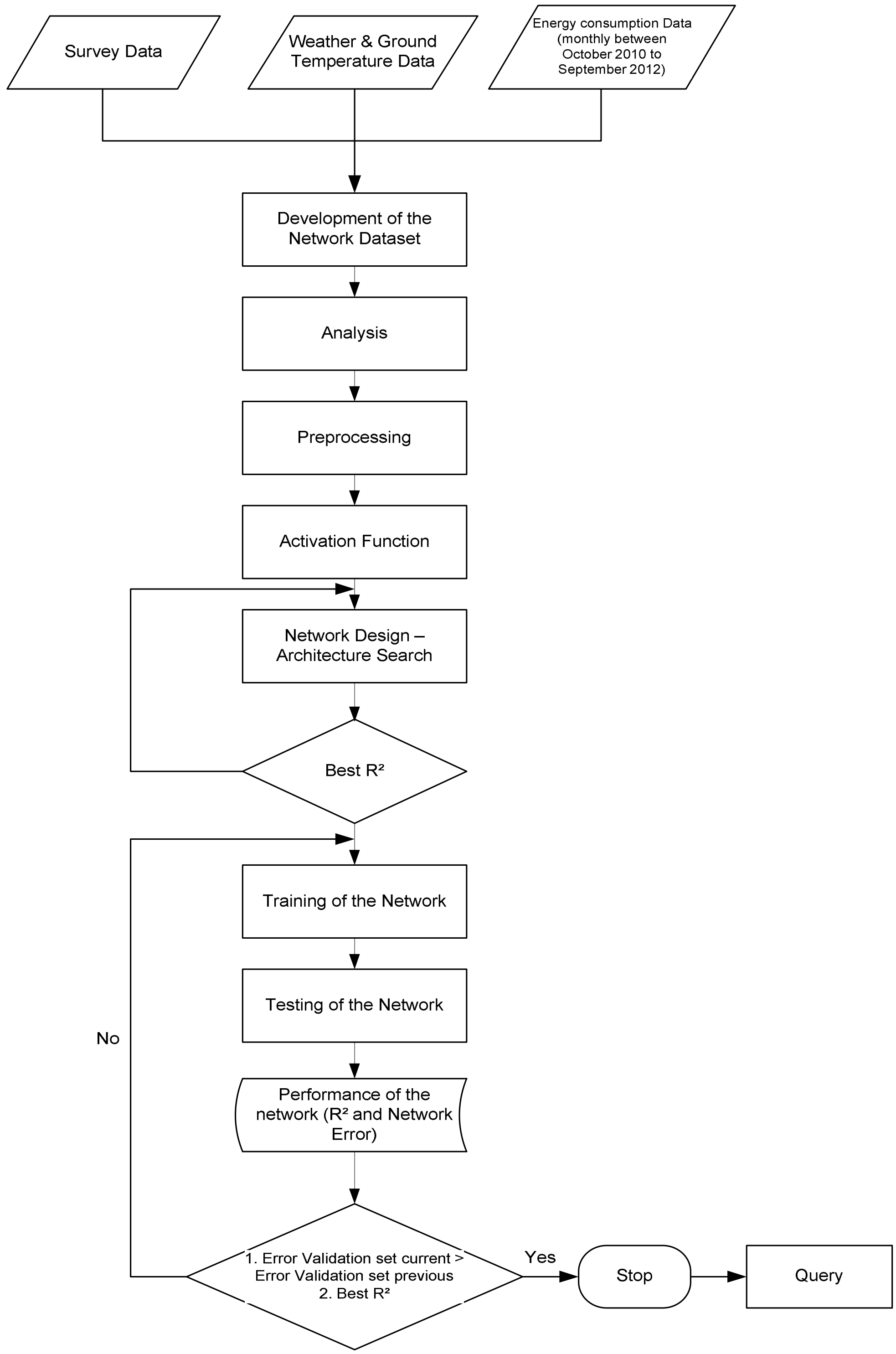

Figure 2: Flowchart depicting the methodology used for the development of the ANN model.

Multiple preprocessing attempts were conducted in order to find the best dataset. This was done by training the model and comparing the R^2 predictor performance values. All data were partitioned randomly. Data were partitioned in the following way: 750 records to Training set (65.1%), 218 records to Validation set (18.92%), 184 records to Test set (15.97%).

4.3.4. Activation Function and Network Design

To design the network, a network architecture (number of hidden layers and units in each layer) and network properties (error and activation functions) are needed. Network properties are defined automatically in the software but it is possible to change them manually, which in some cases improves network performance. In this study, an architecture search was launched with two (2) hidden layers and by changing the activation function. Table 1 shows the results of these searches.

Table 1: Architecture search results

Hidden layer activation function	**Output activation function**	**Architecture**	**Number of weight**	**R^2**	**Correlation**
Logistic	Logistic	[34-85-54-1]	7674	0.993	0.997
Tangent	Logistic	[34-70-39-1]	5259	0.992	0.996
Tangent	Logistic	[34-68-1]	2449	0.974	0.987
Logistic	Logistic	[34-73-1]	2629	0.975	0.987
Logistic	Tangent	[34-77-1]	2773	0.961	0.981
Logistic	Linear	[34-72-1]	2593	0.935	0.967
Linear	Logistic	[34-8-1]	289	0.593	0.771

The best retained network architecture is [34-85-54-1] with a R^2 of 0.993.

4.3.5. Training of the Network

The following algorithms supported by NeuroIntelligence were applied to the training dataset: Quick Propagation (QP), Conjugate Gradient Descent (CGD), Limited Memory Quasi-Newton (LMQ-N), and Online Back Propagation (OBP).

The termination criterion (when to stop training) used was when an increase of the network error occurred or the number of iterations (1,000 iterations). Table 2 shows the different training algorithms and iterations applied to the dataset. Correlation coefficient (R^2) is also shown in Table 2, which is the performance predictor for all datasets. The best network found is using the Quick Propagation training algorithm with 151 iterations and an overall R^2 of 0.895.

Table 2: Performance of the ANN model

Number of iterations	Training algorithm	Coefficient correlation (R^2)			
		Training	Validation	Testing	All
151	QP	0.857	0.942	0.937	0.895
415	LMQ-N	0.873	0.762	0.784	0.835
173	CGD	0.817	0.608	0.690	0.756
1000	OBP	0.778	0.568	0.679	0.756

5. Results

The best network found is using the Quick Propagation training algorithm with 151 iterations and an overall R^2 of 0.895. On average, the energy consumption, using selected household characteristics, was 171 kWh/month. Table 3 displays the average household monthly energy consumption in the Toronto MURB.

In sections 5.1 and 5.2, two methods to validate the neural network are explained.

Table 3: Selected household's energy consumption (kWh): October 2010 to September 2012

Year	Month	Household's energy consumption (kWh)
2010	October	159
	November	159
	December	206
2011	January	227
	February	215
	March	194
	April	165
	May	163
	June	175
	July	173
	August	161
	September	155
	October	162
	November	158
	December	205
2012	January	183
	February	187
	March	172
	April	173
	May	140
	June	140
	July	142
	August	141
	September	148
Average:		171

5.1. Validation using the Validation Set

As stated earlier, 218 cases (18.92%) were randomly used for validate the network. The R^2 found by the ANN software is 0.94 (Table 2).

5.2. Validation using the Whole Dataset

The ANN software allows querying every case in the dataset. This query dataset was compared to the actual energy consumption. Every month was summed and the percentage of error found is 0.37%. Figure 3 shows the comparison between the predicted and the actual energy consumption.

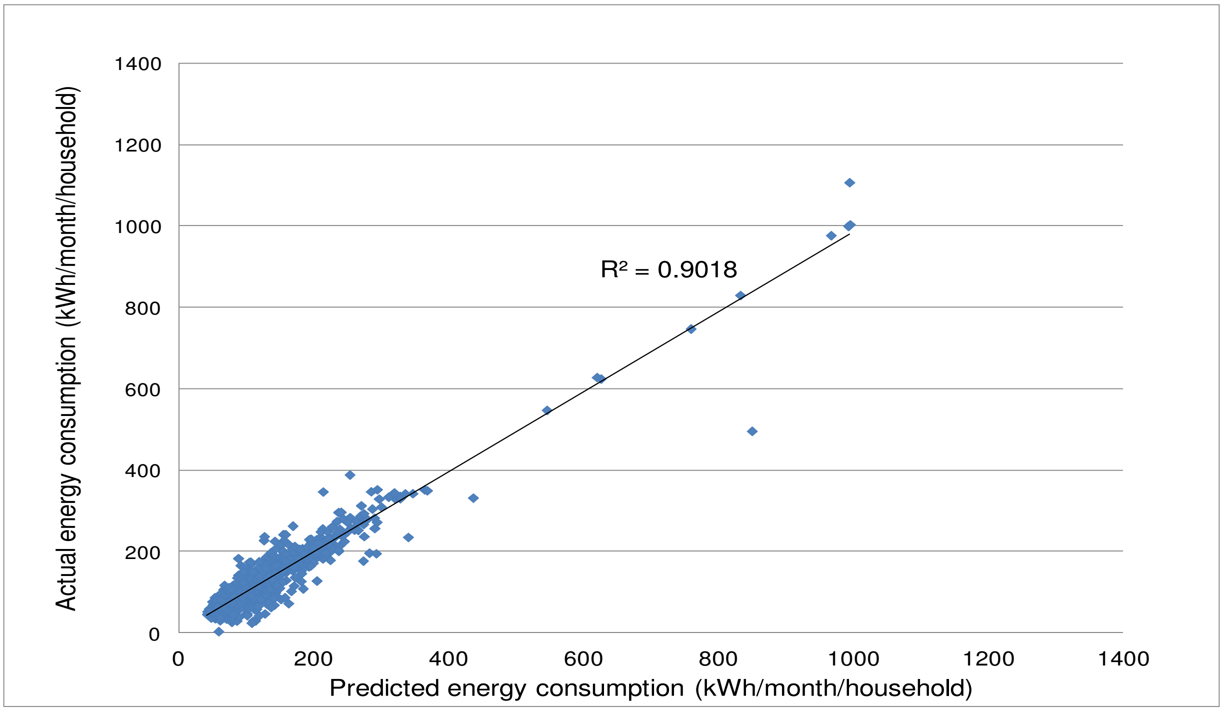

Figure 3: Comparison between the predicted and the actual energy consumption (kWh/month/household).

6. Conclusions

Artificial neural networking has been receiving great interest over recent years and has been applied in many applications [16]. It has been shown to be an effective prediction tool by introducing a foreign dataset (validation set) to the model and having a significantly high correlation coefficient of 0.94. Furthermore, ANN has shown its prediction effectiveness by comparing predicted energy consumption to the actual energy consumption dataset of 1152 cases and resulting in correlation coefficient of 0.90. It is recommended to further investigate the effectiveness of neural networking to other residential energy consumption modeling techniques such as conditional demand analysis or engineering modeling.

7. Acknowledgements

The authors would like to thank all of our sponsors of the project: Canada Mortgage and Housing Corporation (CMHC), Ontario Ministry of Municipal Affairs and Housing (MAH), City of Toronto, Enbridge Gas Distribution Inc., and MITACS.

8. References

[1] Heinz, S. (2011). Mathematical Modelling. *Published by Springer Heidelberg Dordrecht*. New York. pp. 1.

[2] Yohanis, Y. G. (2012) Domestic Energy Use and Householders' energy behaviour. *Journal of Energy Policy*. 41: pp.654-665.

[3] Swan, L. G., Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*. 13: pp. 1819-1835.

[4] Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A, Stevanovic, Z., Djurovic-Petrovic, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Journal of Building and Environment*. 45: pp. 1683-1697.

[5] Boulaire, F., Higgins, A., Foliente, G., McNamara, C. (2013). Statistical modelling of district-level residential electricity use in NSW, Austrailia. *In Proceedings of Journal of Sustainability Science*.

[6] Dayhoff, J. E., DeLeo, J. M. (2001). Artificial Neural Networks: Opening the Black Box. *Conference on Prognostic Factors and Staging in Cancer Management: Contributions of Artificial Neural Networks and Other Statistical Methods*: pp. 1615-1635.

[7] Aydinalp, M., Ugursal, V. I., Fung, A. S. (2003). Modelling of residential energy consumption at the national level. International Journal of Energy Research. 27: pp. 441-453.

[8] Aydinalp, M., Ugursal, V., & Fung, A. (2003). Effects of socioeconomic factors on household appliance, lighting, and space cooling electricity consumption. *International Journal of Global Energy Issues* , 20 (3): pp. 302-315.

[9] Aydinalp, M., Ismet Ugursal, V., Fung, A. S. (2004). Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. *Journal of Applied Energy.* 79: pp. 159-178.

[10] Aydinalp-Koksal, M., Ugursal, V. I. (2008). Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Journal of Applied Energy*. 85: pp. 271-296.

[11] Alyuda Research Company (2003). Alyuda NeuroIntelligence User Manual Version 2.1.

[12] Aydinalp, M. (2002). A New Approach for Modeling of Residential Energy Consumption. ProQuest Dissertations & Theses (PQDT). Dissertation/thesis number NQ77588 - Dalhousie University.

[13] Anstett, M., and Kreider, J. F. (1993) Application of Neural Networking Models to Predict Energy Use. ASHRAE Transactions. 99 (1): pp.505-517.

[14] Roque, M, Straka, V., Fung, A. (2012) Survey of Household Energy Use in a Toronto Rental High-rise Multi-Unit Residential Building (MURB). *In Proceedings of the 2nd World Sustain. Forum, 1-30 November 2012*; Sciforum Electronic Conferences Series.

[15] Environment Canada (2012). National Climate Data and Information Archive. Retrieved on web link: http://www.climate.weatheroffice.gc.ca.

[16] Michelis-Tzanakou, E. (2011). Artificial Neural Networks: An Overview. *Network: Computation in Neural Systems*. 22 (1-4): pp. 208-230.

9. Biography

Miles Roque is a Masters candidate in the Environmental Applied Science and Management program at Ryerson University in Toronto. Miles completed his undergraduate degree at University of Toronto in 2010 and will be completing his Masters in spring 2013. His research interests include energy efficiency, community development, and occupant's household energy use. More specifically, his work examines occupant's household energy use in multi-unit residential buildings.

Sylvain Prez is doing a co-op program at Icam (French engineering school) and Demathieu & Bard (French construction company) since 2010. He will graduate at the end of 2013 and hold a degree in general engineering.

Dr. Alan Fung, P.Eng. (Ontario, Nova Scotia), an Associate Professor in the Department of Mechanical and Industrial Engineering, Ryerson University, oversees a vigorous research program on sustainable building integrated energy systems/“Net Zero” energy buildings. He participates in the NSERC Smart Net-zero Energy Buildings Research Network (SNEBRN) and works closely with public and private sectors in promoting sustainable technology development. He is also the faculty adviser of Ryerson ASHRAE Student Chapter.

Vera Straka is an Associate Professor in the Department of Architectural Science. She is an active member of PEO, the chair of Ontario Division of the Institution of Structural Engineers, member of the Ontario Building Envelope Council and its BSSO designation committee, a member of CSA Committee on Sustainability. Vera is the fellow of the CSCE. Research interests: sustainable buildings, life cycle assessment, post-occupancy evaluation and reuse of components and materials.