

# Methodology for Commercial Buildings Thermal Loads Predictive Models Based on Simulation Performance

Kapetanakis D.S.  
School of Mechanical and Materials  
Engineering  
University College Dublin (UCD)  
+353-1-7161726  
dimitrios.kapetanakis  
@ucdconnect.ie

Mangina E.  
School of Computer Science and  
Informatics  
University College Dublin (UCD)  
+353-1-7162858  
eleni.mangina@ucd.ie

Finn D.P.  
School of Mechanical and Materials  
Engineering  
University College Dublin (UCD)  
+353-1-7161947  
donal.finn@ucd.ie

## ABSTRACT

Commercial buildings incorporate Building Energy Management Systems (BEMS) to monitor indoor environment conditions as well as controlling Heating Ventilation and Air Conditioning (HVAC) systems. Measurements of temperature, humidity and energy consumption are typically stored within BEMS. These measurements include underlying information regarding building thermal response, which is crucial for the calculation of heating and cooling loads. Forecasting of building thermal loads can be achieved using data records from BEMS. Accurate predictions can be produced when introducing these data records to data-mining predictive models. Incomplete datasets are often acquired when extracting data from the BEMS; hence detailed representations of commercial buildings can be implemented using EnergyPlus. For the purposes of the research described in this paper, different types of commercial buildings in various climates are examined to investigate the scalability of the predictive models.

## Categories and Subject Descriptors

I.6.3 [Simulation and Modeling]: Applications

## General Terms

Measurement, Performance, Design, Reliability.

## Keywords

Commercial Buildings, Predictive Models, Thermal Loads, Simulation Data

## 1. PROBLEM DOMAIN

The building sector consumes 35% of global final energy use and is responsible for about 17% of total direct energy-related CO<sub>2</sub> emissions from final energy consumers [1]. Space heating, water heating and space cooling account for nearly 55% of global building energy use and represent the largest opportunity to reduce buildings energy consumption [1]. The use of efficient

energy management systems in buildings is estimated to be able of saving up to 8% of the energy consumption in the entire EU [2]. In order to decrease the energy usage and increase compliance with the European Directives on the energy performance of buildings [3], it is of fundamental importance to control efficiently the existing HVAC systems. Predicting the load of HVAC systems is important for energy management especially during peak energy demand hours [4].

Predictions of building thermal load can be estimated using appropriate simulation software [5] when detailed data such as building geometry, occupancy as well as environmental variables are available. In reality, such data are often unknown, especially for older buildings, where uncertainty arising from parameter and occupancy estimation can lead to significant additional modelling efforts [6]. An alternative way to forecast these loads is to take advantage of BEMS recorded data. These data records include underlying information regarding building thermal response and can be introduced to machine learning predictive models, which utilise extensive assessment of input and output variables, in order to produce accurate predictions [7].

This research project focuses on a novel approach for cost-effective modelling of measured data from commercial buildings, utilising simulation tools such as EnergyPlus [8] and IBM SPSS Modeler [9], with the implementation of machine learning prediction methods that can be assembled rapidly and deployed easily. This approach will constitute a practical research testbed to optimise multiple objectives related to the buildings' energy modelling research area, such as the development of a predictive model for heating and cooling loads of commercial buildings; the generation of highly accurate predictions; the scalability of the new approach to any commercial building and minimum commissioning and maintenance effort requirements.

The main objective of the current paper is to describe the development of research regarding the construction of predictive models able to forecast thermal loads of commercial buildings.

## 2. STATE OF THE ART

The most common methods used in the literature to achieve forecasting of building thermal loads are Regression, Artificial Neural Network (ANN) and Support Vector Machine (SVM).

Aranda et al. [10] used linear regression models to predict the annual energy consumption in the Spanish banking sector. The energy consumption of a single building was predicted as a function of its construction characteristics, climatic area and energy performance. Catalina et al. [11] worked on the development of regression models to predict monthly heating demand for single-family residential sector. The inputs for the

regression models were building shape factor, envelope U-value, window to floor area ratio, building time constant and climate. An update to the aforementioned work [12], in an attempt to simplify the model to obtain fast predictions, used as inputs the building global heat loss, south equivalent surface and difference between indoor and ambient temperature.

ANNs have been applied to analyse various types of building energy consumption as well as heating loads. Kalogirou et al. [13,14] implemented ANN at an early design stage to predict the required heating load of buildings. Input data included the areas of windows, walls and floors, the type of windows and walls, roof classification and the room temperature. Gonzalez and Zamarreno [15] used an ANN approach to predict the hourly energy consumption in buildings. The inputs of the network were current and forecasted values of temperature, the current load, the hour and the day.

SVM models have been used more recently for predicting energy consumption in buildings. Dong et al. [16] were the first to introduce the use of SVM for the prediction of four commercial buildings energy consumption. The input variables were mean outdoor dry-bulb temperature, relative humidity and global solar radiation. Li et al. [17] used a SVM model in regression to predict hourly building cooling load for an office building. The outdoor dry-bulb temperature and solar radiation intensity were the input parameters for this model. A comparison of the developed SVM model against a back propagation neural network, a radial basis function neural network and the general regression neural network was published by Li et al. [18]. Simulation results indicated that the SVM and general regression neural network methods achieved better accuracy.

The research field related to building thermal loads forecast has been very productive, involving various scientific domains. It is a fact though that based on the literature none of the research groups has introduced the predictive models to the entire commercial sector. Hence, the investigation of the existence of a predictive model able to forecast thermal loads of any given commercial building is the ultimate objective of this research. Within the next section, the development of the methodology that will be followed is described.

### 3. METHODOLOGY

The methodology developed is based on the primary objective of this research, which is the determination of a predictive model able to forecast thermal loads of any given commercial building. In order to achieve this, the following steps took place:

- i. Selection of representative commercial buildings to form a reliable set of testbeds;
- ii. Selection of representative climates to cover all possible weather conditions;
- iii. Generation of 1 year simulated data at 15-minute intervals to create a synthetic database, to avoid sparse data;
- iv. Data analysis of simulated data for the selection of input variables to the predictive models;
- v. Development of predictive models using Regression, ANN and SVM algorithms using one of the testbed buildings;
- vi. Evaluation of predictive models based on their accuracy;
- vii. Selection of the most accurate predictive model;
- viii. Examination of scalability by applying the model to the rest testbed buildings;
- ix. Determination of commissioning and maintenance effort for the implementation of the model.

The methodology described in this sequence is illustrated in Figure 1.

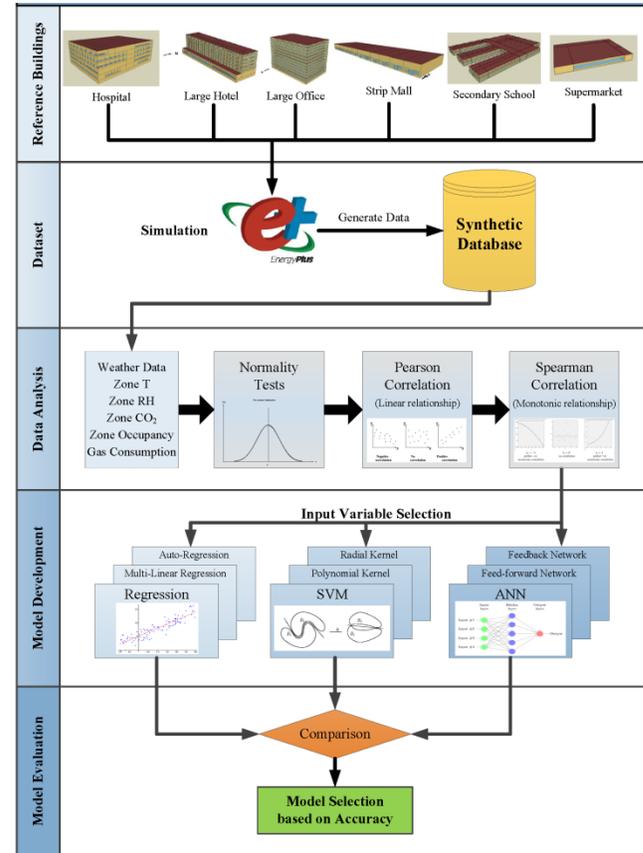


Figure 1. Development of methodology.

#### 3.1 Testbed Buildings and Climates

The first step of the methodology is the creation of a synthetic database that includes data from representative commercial buildings in various climates. Considering that it is impractical to model every commercial building, or even to represent every building sub-sector [19], a small number of prevalent building types is selected. This database is obtained with the use of reference commercial building models created by the Department of Energy (D.O.E.) of the U.S. [20]. The repository of D.O.E. covers building types that directly characterize more than 60% of the commercial building stock and are very similar to other commercial building types. The six reference buildings implemented in this research are summarised in Table 1 and their EnergyPlus models have been obtained from D.O.E. [20].

Table 1. Reference buildings [20].

Building Type	Floor Area (m <sup>2</sup> )	No. of Floors
Large Office	46,320	12
Secondary School	19,590	2
Strip Mall	2,090	1
Supermarket	4,180	1
Hospital	22,420	5
Large Hotel	11,345	6

Investigation of the existence of a global solution for a predictive model of thermal loads of commercial buildings requires the investigation of all possible weather conditions. Therefore, representative cities were selected to capture all type of climates. The selection was based on the American National Standard Institute (ANSI) and the American Society of Heating, Refrigerating and Air-Conditioning (ASHRAE) distinction of climate zones [21]. Table 2 displays the representative cities of each zone and their climates.

**Table 2. Representative cities and climates.**

Zone No.	Representative City	Climate
1A	Miami	Very Hot, Humid
1B	Riyadh	Very Hot, Dry
2A	Houston	Hot, Humid
2B	Phoenix	Hot, Dry
3A	Memphis	Warm, Humid
3B	El Paso	Warm, Dry
3C	San Francisco	Warm, Marine
4A	Baltimore	Mild, Humid
4B	Albuquerque	Mild, Dry
4C	Salem	Mild, Marine
5A	Chicago	Cold, Humid
5B	Boise	Cold, Dry
5C	Vancouver	Cold, Marine
6A	Burlington	Cold, Humid
6B	Helena	Cold, Dry
7	Duluth	Very Cold
8	Fairbanks	Subarctic

The combination of the selected reference buildings and the different climate zones produces a sample of 102 case studies that are simulated with EnergyPlus. The generated results of the simulations form the synthetic database that is utilised for the data analysis process.

### 3.2 Data Analysis

Data analysis is utilised to explore the data, to search for consistent patterns and/or systematic relationships between variables. The process of data analysis to detect interrelationship between variables within the current research consists of three stages: (1) the examination of the distribution of each variable (e.g. adhering to a Gaussian distribution), (2) investigation of linear correlation by calculating Pearson correlation coefficient, and (3) investigation of monotonic correlation by calculating Spearman correlation coefficient.

The initial task of the data analysis process is the selection of BEMS variables to be assessed. All of the variables are selected based on the most common sensors installed in commercial buildings. These variables are divided into two categories, input and output variables. Inputs are the ones introduced to the predictive model and outputs are the thermal loads, which will be forecasted from the model. Data analysis is based on simulated

data obtained by the EnergyPlus models of the testbed buildings. Furthermore, the use of simulated data precludes the possibility of dealing with incomplete data sets that are often acquired when extracting recorded data in real-life. The EnergyPlus models provide one year simulated data at 15-minute intervals.

The procedure to be followed for analysing the data and examining the existence of a correlation between input and output variables is based on statistical techniques. Initially the selected input and output variables are tested to check if they follow a Gaussian distribution. This is achieved with the use of statistical tests [22]. The linear correlation between input and output variables is investigated afterwards by performing a Pearson correlation [23], which measures the linearity between paired data. In a sample of data, it is denoted by  $r_p$  and is by design constrained between -1 and 1, where positive values denote positive linear correlation, while negative values denote negative linear correlation and a value of zero denotes no linear correlation. The strength of correlation can be verbally described using the following guide for the absolute value of  $r_p$  where: 0.00 – 0.19 represents “very weak”, 0.20 – 0.39 represents “weak”, 0.40 – 0.59 represents “moderate”, 0.60 – 0.79 represents “strong” and 0.80 – 1.00 represents “very strong” correlation.

Furthermore, the existence of monotonic relationships can be identified as well through the Spearman correlation coefficient [23]. The Spearman correlation coefficient is a statistical measure of the strength of a monotonic relationship between paired data and is denoted by  $r_s$ . The principles of Spearman correlation are the same as the ones of Pearson correlation coefficient.

Input variables are grouped according to weather data and indoor variables, as given in Table 3. Heating and cooling loads of the testbed buildings are the output variables.

The statistical analysis is carried out using the IBM SPSS Statistics 20 software [24], which is able of addressing the entire analytical process, from planning to data collection to analysis, reporting and deployment.

**Table 3. Input variables under examination.**

Weather Data	Indoor Data
Ambient Temperature	Zone Air Temperature
Ambient Relative Humidity	Zone Relative Humidity
Wind Speed	Zone Occupancy
Solar Radiation	Zone CO <sub>2</sub> level
Sky Clearness	

The selection process of the input variables that will be introduced to the predictive model is based on the calculated Pearson and Spearman correlations between input and output variables. Absolute values of the calculated Pearson and Spearman coefficients are used to simplify this process. Only “moderate”, “strong” and “very strong” relationships are of interest, hence the threshold value for introducing an input variable to the predictive model is 0.5. It is important to select carefully the input variables in order to avoid unnecessary increase in the complexity during the development of the model.

### 3.3 Simulation Tools

EnergyPlus is implemented to generate the simulated results in order to form a reliable synthetic database that will be utilised at the data analysis part of the project. It is a whole building energy

simulation program that engineers and researchers use to model energy use in buildings. It is used to model heating, cooling, lighting, ventilation, other energy flows, and water use. Moreover, it includes many innovative simulation capabilities such as, time-steps less than an hour, modular systems and plant integrated with heat balance-based zone simulation, multizone air flow, thermal comfort, water use, natural ventilation, and photovoltaic systems.

Each reference building type was simulated, for every one of the seventeen climate zones proposed by ASHRAE, for a year at a 15-minute interval. This combination of reference buildings and climate zones (consisting of 102 case studies) covers all possible scenarios regarding weather conditions. The EnergyPlus models of the reference building types can be displayed using SketchUp software [25].

The development of the predictive models will be produced with the use of the IBM SPSS Modeler 14.2 software [9], which is an extensive predictive analytics platform. It is designed to bring predictive intelligence to decisions made by individuals, groups, systems and the enterprise. Additionally, it provides a range of advanced algorithms and techniques that include data analytics, entity analytics, decision management and optimization.

All applicable regression, SVM and ANN algorithms will be developed and applied to all six reference commercial buildings. Investigation of the performance of the various models will be achieved with the use of the same software, since it is possible to extract it as an outcome when the predictions are generated. Once the performance is calculated, assessment of the overall findings will follow.

### 3.4 Predictive Models

In order to determine the new model, the selection of the most suitable predictive model is required. Different multiple regression models alongside numerous SVM models and several architectures of ANN will be developed. All of the developed predictive models will be tested and evaluated regarding their ability to forecast thermal loads in order to identify the most accurate. The developed models will be applied to all six reference commercial buildings, where the accuracy will be calculated and compared between the various models.

The scalability of each model will be the next thing under examination. Evaluation of the performance of each model when applied to all reference commercial buildings in various climates will indicate the ability of each model to be transferred from one building type to another.

Finally, commissioning and maintenance effort for the implementation of the new model will be determined. Hence, the model will be evaluated based on its ability to meet necessary requirements.

## 4. DISCUSSION

The expected outcome of this project is the development of a novel whole-building energy predictive model. The model will take advantage of historical data of commercial buildings in order to generate accurate prediction of heating and cooling load. Simulation software will be one of the milestones of this project, since the generation of the necessary database for the development of the predictive models is achieved with the usage of EnergyPlus. Furthermore, the implementation of IBM SPSS Statistics and Modeler will contribute to the data analysis and the model development respectively.

Once a comprehensive database is obtained, the most suitable methodology for this application is going to be selected between Regression, SVM and ANN predictive models. Evaluation of the developed models will be performed based on the accuracy of each model. After the selection process, the selected model will be evaluated regarding its scalability. The ability of forecasting heating and cooling loads of different given commercial buildings in various climates within the same level of accuracy will be the criterion. Finally, the effort required for commissioning and maintenance of the model should be as little as possible. The final outcome should be a scalable model with minimum commissioning and maintenance requirements.

Currently, the research project is at the development stage of the predictive models. The generation of the simulated database from the 102 case studies has been completed. Moreover, the data analysis of the database, in order to select the input variables for the predictive models, has been performed.

Ideally, this novel approach of estimating the thermal and cooling load of commercial buildings could be implemented to the control of the BEMS. In this way, the efficiency of the HVAC systems of the building will be improved reducing the energy consumption at the same time. This will lead to a reduction of the energy cost of commercial buildings as well.

## 5. ACKNOWLEDGMENTS

This research is conducted with financial support from the Irish Research Council and United Technologies Research Centre (UTRC). The research was also financially assisted by the Electricity Research Centre, University College Dublin, Ireland, which is supported by the Commission for Energy Regulation, Bord Gas Energy, Bord na Mona Energy, Cylon Controls, EirGrid, Electric Ireland, Energia, EPRI, ESB International, ESB Networks, Gaelectric, Intel, SSE Renewables and UTRC.

## 6. REFERENCES

- [1] International Energy Agencies. 2013. Transition to Sustainable Buildings Paris: IEA.
- [2] Ferreira, P., Ruano, A., Silva, S. and Conceicao, E. 2012. Neural networks based predictive control for thermal comfort and energy savings in public buildings. *Energy and Buildings*; 55: p. 238-251.
- [3] European Commission 2010. Directive 2010/31/EU of the European Parliament and the Council on the Energy Performance of Buildings (recast). Brussels.
- [4] Kudiak, A., Li, M. and Zhang, Z. 2010. A data-driven approach for steam load prediction in buildings. *Applied Energy*; 87: p. 925-933.
- [5] U.S. D.O.E. Building Energy Software Tools Directory. 2011. Accessed 8 January 2015. Available: [http://apps1.eere.energy.gov/buildings/tools\\_directory](http://apps1.eere.energy.gov/buildings/tools_directory).
- [6] Kwok, S. and Lee, E. 2011. A study of the importance of occupancy to building cooling load in prediction by intelligent approach. *Energy Conversion and Management*; 52: p. 2555-2564.
- [7] Fouquier, A., Robert, S., Suard, F., Stéphane, L. and Jay, A. 2013. State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*; 23: p. 272-288.

- [8] U.S. D.O.E. EnergyPlus Energy Simulation Software. 2015. Accessed 8 January 2015. Available: <http://apps1.eere.energy.gov/buildings/energyplus/>.
- [9] IBM Corp. IBM SPSS Modeler 14.2. 2011. Accessed 04 February 2015. Available: <http://www-01.ibm.com/software/analytics/spss/products/modeler>.
- [10] Aranda, A., Ferreira, G., Mainar-Toledo, M., Scarpellini, S. and Sastresa, EL. 2012. Multiple regression models to predict the annual energy consumption in the Spanish banking sector. *Energy and Buildings*; 49: p. 380-387.
- [11] Catalina, T., Virgone, J. and Blanco, E. 2008. Development and validation of regression models to predict monthly heating demand for residential buildings. *Energy and Buildings*; 40: p. 1825-1832.
- [12] Catalina, T., Iordache, V. and Caracaleanu, B. 2013. Multiple regression model for fast prediction of the heating energy demand. *Energy and Buildings*; 57: p. 302-312.
- [13] Kalogirou, SA., Neocleous, CC. and Schizas, CN. 1997. Building heating load estimation using artificial neural networks. In *17th International Conference on parallel architectures and compilation techniques* (San Francisco, California, US) p. 1-8.
- [14] Kalogirou, S., Florides, G., Neocleous, C. and Schizas, C. 2001. Estimation of Daily Heating and Cooling Loads Using Artificial Neural Networks. In *7th REHVA World Congress* (Napoli, Italy): Clima 2000
- [15] González, PA. and Zamarreño, JM. 2005. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*; 37(6): p. 595-601.
- [16] Dong, B., Cao, C. and Lee, SE. 2005. Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*; 37(5): p. 545-553.
- [17] Li, Q., Meng, Q., Cai, J., Yoshino, H. and Mochida, A. 2009. Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*; 86(10): p. 2249-2256.
- [18] Li, Q., Meng, Q., Cai, J., Yoshino, H. and Mochida, A. 2009. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management*; 50(1): p. 90-96.
- [19] Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., et al. 2011. United States Department of Energy Commercial Reference Building Models of the National Building Stock. Technical Report. Colorado, National Renewable Energy Laboratory. Report No.: NREL/TP-5500-46861.
- [20] U.S. D.O.E. Commercial Prototype Building Models. 2014. Accessed 04 February 2015. Available: <https://www.energycodes.gov/development/commercial/90.1/models>.
- [21] ANSI / ASHRAE 2007. Energy Standard for Buildings Except Low-Rise Residential Buildings. ANSI/ASHRAE Standard 90.1 -2007. Atlanta, GA, USA ASHRAE.
- [22] Justel, A., Peña, D. and Zamar, R. 1997. A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statist. Probab. Lett.*; 35(3): p. 251-9.
- [23] Spiegel, M., Schiller, J. and Srinivasan, A. 2013. Probability and Statistics. 4th ed.: McGraw Hill Professional.
- [24] IBM Corp. IBM SPSS Statistics for Windows, Version 22.0. 2013. Accessed 12 January 2015. Available: <http://www-01.ibm.com/software/analytics/spss/products/statistics/>.
- [25] Trimble Navigation Limited. SketchUp 8. 2010. Accessed 09 February 2015. Available: <http://www.sketchup.com/>.
- [26] Justel, A., Peña, D. and Zamar, R. 1997. A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statist. Probab. Lett.*; 35(3): p. 251-259.