Short-term forecasting for the electrical demand of Heating, Ventilation, and Air Conditioning systems

Jason Runge

A Thesis
In the Department
of
Building, Civil, and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctorate of Philosophy (Building Engineering) at
Concordia University
Montreal, Quebec, Canada

April 2021

© Jason Runge 2021

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Jason Runge

Entitled: Short-term forecasting for the electrical demand of Heating, Ventilation, and Air Conditioning systems

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Building Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________ Chair
Dr. Youmin Zhang

______________________________ External Examiner
Dr. Benjamin C.M. Fung

______________________________ External to Program
Dr. Pragasen Pillay

______________________________ Examiner
Dr. Andreas K. Athienitis

______________________________ Examiner
Dr. Hua Ge

______________________________ Thesis Supervisor(s)
Dr. Radu Grigore Zmeureanu

Approved by ______________________________

Dr. Michelle Nokken, Graduate Program director

Wednesday April 7th, 2021

____

Dr. Mourad Debbabi

Faculty of Engineering and Computer Science

Abstract

Short-term forecasting for the electrical demand of Heating, Ventilation, and Air Conditioning systems

Jason Runge, Ph.D.
Concordia University, 2021

The heating, ventilation, and air conditioning systems (HVAC) of large scale commercial and institutional buildings can have significant contributions to the buildings overall electric demand. During periods of peak demand, utilities are faced with a challenge of balancing supply and demand while the system is under stress. As such, utility companies began to operate demand response programs for large scale consumers. Participation in such programs requires the participant to shift their electric demand to off-peak hours in exchange for monetary compensation. In such a context, it is beneficial for large scale commercial and institutional buildings to participate in such programs. In order to effectively plan demand response based strategies, building energy managers and operators require accurate tools for the short-term forecasting of large scale components and systems within the building. This thesis contributes to the field of demand response research by proposing a method for the short-term forecasting for the electric demand of an HVAC system in an institutional building.

Two machine learning based approaches are proposed in this work: a component method and a system based method. The component-level approach forecasts the electric demand of a component within the HVAC system (e.g. air supply fans) using an autoregressive neural network coupled with a physics based equation. The system-level approach uses deep learning models to forecast the overall electric demand of the HVAC system through forecasting the electric demand of the primary and secondary system. Both approaches leverage available data from the building automation system (BAS) without the need for additional sensors. The system based forecasting method is validated through a case study for a single building with two data sources: measurement data obtained from the BAS and from an eQuest simulation of the building. The building used as the case study for the work herein consists of the Genomic building of Concordia University Loyola campus.

Acknowledgements

First and foremost, I wish to thank my supervisor, Dr. Radu Zmeureanu, for his great advice and support throughout my research and studies. This research would not have been possible without him. His invaluable guidance, comments, and support have allowed me to increase my potential as a researcher and person. I am forever grateful.

I would also like to thank Daniel Gauthier with Concordia Facilities Management. I really appreciated you sharing your knowledge of the building and systems and volunteering your time to walk me through them.

I would also like to thank my colleagues of the 16^{th} floor. Thank you all for the laughs and discussions throughout this journey.

Finally, I would like to acknowledge the financial support of Natural Sciences and Engineering Research Canada and the Faculty of Engineering and Computer Science of Concordia University

For my family, without whom this would not be possible.

Table of Contents

List of figures

List of tables

List of acronyms

Name	Definition
AE	Autoencoder
AHU	Air handling unit
AI	Artificial intelligence
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
ARIMAX	Autoregressive integrated moving average with exogenous input
ASHRAE	American Society of Heating, Refrigeration and Air Conditioning Engineers
BAS	Building automation systems
CI	Commercial and institutional buildings
CNN	Convolutional neural network
COP	Coefficient of performance
CP	Central plant
CV(RMSE)	Coefficient of variation of root-mean-square-error
DBN	Deep belief neural network
D-FFNN	Deep feed forward neural network
DL	Deep learning
DR	Demand response
DNN	Deep neural networks
EN-LSTM	Encoder coupled to a long short term memory network
FFNN	Feed forward neural network
GE	Genomic research center
GRU	Gated recurrent unit
HVAC	Heating, ventilation, and air conditioning systems
LSTM	Long short term memory network
LTF	Long term load forecasting
MAE	Mean absolute error
MAPE	Mean absolute percent error
MBE	Mean bias error
ML	Machine learning
MSE	Mean squared error
MTF	Medium term load forecasting
PIR	Power input ratio
PLR	Part load ratio
$\mathbf{R^2}$	Coefficient of determination
RBM	Restricted Boltzmann machine
RF	Random forest
RMSE	Root mean square error
RNN	Recurrent neural network
SFA	Simple forecasting approach
STF	Short term load forecasting
SVM	Support vector machines
SVR	Support vector regression

Chapter 1: Introduction

1.1 Problem Statement

Utility companies are faced with a constant challenge of balancing supply with demand. For periods of high demand, there is a greater potential for disruptions to the electric grid. High demand periods may be caused by such things as: extreme weather days (high system demands), power station outages (low system capacity), and/or reduced delivery capacity (damaged transmission lines or scheduled maintenance) etc. With the purpose of reducing the pressures on the electric grid during high demand periods, utility companies operate demand response (DR) based programs. The aim of such programs is for large-scale consumers to reduce and/or shift their electric demand to off peak hours, thus, reducing the pressure on the electric grid [1]. Companies and participants which participate in such programs are offered monetary incentives for their shifts/or reductions in electric demand [1].

Large scale commercial and institutional buildings (CI), can benefit from participating in such DR based programs due to their large scale heating, ventilation, and air conditioning systems (HVAC). Such HVAC systems may have a significant contribution to the buildings overall electrical demand (19-76% depending on the type of commercial building [2]); additionally such buildings typically have automated monitoring and control systems that facilitate the implementation of different control strategies. Furthermore, due to the inertia of such large scale CI buildings and their thermal mass, the building loads may be reduced or even temporarily shut off at key locations without a compromise to the health and safety of the occupants.

In such a context, accurate and fast tools are required for building energy managers to forecast the future electric demand of the HVAC system and its components. Such a tool is required for the assessment of different demand-response based strategies.

1.2 Thesis Scope

The scope of this thesis is focused on the short-term forecasting of the electric demand for demand response-based programs. Artificial intelligence, specifically machine learning based models can be useful in leveraging available data from existing buildings in order to provide both fast and accurate forecasting models. In addition, such machine learning based models can be automated once applied and therefore, require little human effort for updating.

The main purpose of this thesis is to provide a forecasting tool for building energy managers to assess different demand response based scenarios. This forecasting tool will target future estimations for the electric demand of the HVAC system and components within. Machine learning algorithms will be applied in this thesis leveraging available data from currently installed sensors without the need for additional installations.

As forecasting underpins many different approaches to the management, optimization, and control of energy related fields; there are many tasks which could be incorporated into this work, or applications of such approaches developed herein. While such work may be relevant and useful, it is beyond the scope of this thesis which is limited to the development of the forecasting models.

Chapter 2: Literature Review

This chapter reviews studies applied for the forecasting of energy use and demand in buildings. Section 2.1 begins with an overview for energy models and a categorical breakdown for such models. Section 2.2 provides a literature review of artificial neural network (ANN) models for forecasting energy use in buildings. Next, section 2.3 provides a literature review for deep learning (DL) models applied for forecasting energy use in buildings. Section 2.4 presents the objectives for this thesis. Section 2.5 concludes this chapter with an overview of the thesis.

2.1 Forecasting and energy prediction in buildings

The challenge of forecasting both building and HVAC energy usage can be relatively complex due to a large number of diverse factors which influence the system. Such factors can include HVAC system operations, thermal properties of the building envelop, weather and climate conditions, occupants use and their behaviors, usage of the building, efficiency or COP of equipment, etc.

Merriam-Webster defines the action, to forecast as "to predict (some future event or condition) usually as a result of study and analysis of available and pertinent data" [3]. In addition, Merriam Webster defines the action, to predict as "to declare or indicate in advance, foretell on the basis of observation, experience or scientific reason" [4]. Comparing the two terms, an overlap can be seen. This is indicative for illustrating a problem of the two terms within industry. Often, both words are used interchangeably and/or as synonyms with each other further adding to the confusion and a lack of standardization. In order to address this, the following definitions will be used:

(a) Prediction is the estimation of a current value(s)
(b) Forecast is the estimation of a future value(s)

Herein, to estimate will refer to Oxford's definition, "to roughly calculate the value, number, quantity or extent of" [5]. The focus of this review and thesis is on forecasting models.

2.1.1 Classifications of forecasting models based on forecast horizon

In this work, the forecast horizon is defined as the length of time into the future for which the forecasts are estimated. The forecast horizon can vary depending on the application for the model and the data available. There is no set standard for the classification of forecasting models, however, this work will follow published research, references [6, 7, 8] , which classify forecasting models based on the time range of the forecast horizon. The three main categories are short, medium, and long term models and are based on the time ranges presented in Table 1.

Long term forecasts are used by electric utilities and building managers to manage the reserves of energy, and generate plans over longer periods of time. Such forecasts are typically applied for policy making goals. Medium term forecasting consists of horizons from a few weeks to a few years ahead and are used for scheduling maintenance, negotiating contracts, construction, etc. Short term forecasting consists of forecast horizons ranging from sub hourly to a couple weeks ahead. The applications of such forecast are more important for the day to day operations and include: demand side management, demand response, system control, fault detection, and system optimization.

Table 1: Forecast range classification

Classification	Time Range
Short term load forecasting (STF)	0 - 2 weeks
Medium term load forecasting (MTF)	2 weeks - 3 years
Long term load forecasting (LTF)	+3 years

The scope of this thesis is on the short-term forecasting for demand response based applications. Therefore, the subsequent sections will discuss two main types of energy forecasting models for building and HVAC systems as specified by reference [9]: physics based and data-driven models.

2.1.2 Physics based models

Physics-based models, often referred to as forward/classical or white-box models, are those based off of a comprehensive set of governing mathematical and physics-based equations [9]. These models offer insights into the relationship between the load/forecast along with its driving factors. Physics-based models can estimate the behavior of the whole building energy consumption, the thermal comfort, thermal dynamics of the building, and similarly to the HVAC system. The strong detail to the governing laws, equations, and parameters allow them to precisely describe the dynamics of a system.

Physics-based models have widely been used for several decades and are still in use today. These types of models have been implemented into several building simulation software such as DOE-2, eQuest, TRNSYS, EnergyPlus, and ESP-r among others. Simulations of buildings (with different architectures, windows, HVAC systems, etc.) can be performed at various locations, each with its own dedicated weather files. The simulation outputs are often detailed reports about the predicted hourly and yearly building energy consumption. As such, these models are often used in the sizing and design of a building's HVAC system.

The main advantage of physics-based models is that they offer insight into the relationships and variables which govern the system. In addition, they can be accurate, can precisely describe the model, and address both weakness and strengths within the model. The disadvantage of these type of models is that they often require a lot more information (for calibration) and statistical analytical skills. Thus, these models can be extremely time-consuming and expensive to both develop and solve. A significant amount of time is required for users to obtain all the details and the necessary parameters in order to compute the results. In addition, a significant amount of detailed information may be needed about the building's structure, HVAC equipment installed, its current conditions, weather conditions, HVAC schedules, etc. In a lot of cases, obtaining all the necessary parameters cannot be accomplished, or is extremely difficult to achieve. In addition, equipment parts and components degrade over time, thus stressing the need for on-site measurements whenever applicable.

The aim of this work is to forecast the electric demand of an existing HVAC building under current operating conditions. The interest of this work is not for the design or retrofit, rather the analysis of HVAC performance. In the context of demand response, calibrated physics-based models are not used due to time and monetary constraints, extensive detailed knowledge requirements, and their limited flexibility in allowing for adaptive changes in HVAC operation. In contrast, data-driven models and artificial intelligence are continually being developed and present an alternative for forecasting of the electric demand of existing buildings. Therefore, the literature review from this point focuses on such data-driven models.

2.1.3 Data-driven models

Data-driven models, in contrast to physics based models, apply a mathematical model which is calibrated or tuned with measurement data [9]. Consequently, data-driven models do not require extensive detailed knowledge about the internal parameters and settings for the components within the building or system as they are mostly based on measured data obtained. Furthermore, such data is typically easier to obtain from their monitoring and control systems (e.g. building automation systems or BAS, building energy management systems or BEMS). Therefore, such readily available data can be leveraged for energy improving strategies and may potentially hold a plethora of information regarding the current and past operating conditions for the building and systems within. There are several factors which may influence the performance of the data-driven models including: (i) the quality and quantity of the data obtained, (ii) the data-preprocessing steps applied, (iii) the data-driven model selected, and (iv) the tuning/optimization of the data-driven models. Despite the challenges in their development, data driven models have become quite popular in recent years due to the increasing available data, their ease of development compared to physics based-models, and performance results for energy forecasting in buildings [10]. Furthermore, some data-driven models are quite adaptive to new data and modeling nonlinear data. Despite their advantages, data driven models do have some disadvantages as well. Purely data driven models do not derive a relationship between the target variable and the driving factors for it as seen in physics based models. Furthermore, such models are reliant on the quantity and quality of the data obtained.

The American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE), provides a classification for the two main types of data-driven models as: (i) the black-box model, or strictly mathematical model, and (ii), the grey box model coupling physics-based equations with a mathematical model(s) [9].

Black-box models are a type of data-driven model based solely on measurements. The models can also be viewed in terms of inputs and outputs. Therefore, the relationship between the load and its driving physical law-based functions are not explored. Such models are often used to estimate the energy consumption at various levels of a building including: sectors, districts, whole buildings, component level, etc. Black-box models can model non-linear phenomena and adapt well to changes in the data by training and learning based off of new observations.

Grey-box models, typically apply a mathematical model based on measurements to a physics based model in order to provide its output forecasts. Such models have the advantage of the black-box models accuracy and the ability to see the governing relations behind the output forecast with the physics based equation(s). Furthermore, such models may be able to provide multiple output forecasts from a single model (e.g. supply air flow rate and electric demand). However, such models typically have a slightly larger forecasting error when compared to a strictly black-box based approach.

Typically, the mathematical model applied in a data-driven model can be broken into two sub-categories based on the type of mathematical approach selected: (i) statistical models, and (ii) machine learning based models. Statistical models can include regression and time series-based models. Machine learning (ML) models can include: Fuzzy systems, state vector machines, group method of handling data, random forests, and artificial neural networks, etc. While both statistical and ML approaches are similar, there are a couple clear distinctions between the two. Firstly, statistical models typically apply a pre-set mathematical function calibrated with measurements to provide their forecasts e.g. regression models. In contrast, ML models typically apply an algorithm approach. Secondly, ML models often non-linearly transform their data in comparison to statistical models which do not [11].

2.1.3.1 Statistical models

2.1.3.1.1 Regression

An overview of regression techniques applied for prediction of residential energy consumption can be found in reference [12]. Regression based models have been widely used in forecasting and prediction for energy loads in buildings [13-15]. Regression is a common technique due to its relative ease of implementation, less computational time, satisfactory results, and it does not require detailed information about the building structure/HVAC system [12]. Regression based models identify a relationship between the dependent variable(s) and independent variable(s). Within this work, the dependent variables will be termed the target variables, and the independent variables will be termed regressors. There exist multiple regression-based techniques with the most commonly used including: simple linear regression, multiple linear regression, and non-linear regression. Regressors selected are important factors for the forecasting of any model, such things can represent outdoor air temperature, outdoor humidity, historical load, climate conditions, building/HVAC conditions, etc. However, in terms of short term forecasting for both HVAC and building energy, regression models often provide less accuracy than AI-based models [16-18]. This is due to regressions inability to model non-linear information as accurately as AI-based methods. In addition, regression-based models are more sensitive to outliers than AI-based models.

2.1.3.1.2 Time series

Time series forecasting is one of the most prominent and popular methods. These methods have been widely applied in the business, economic, and financial sectors. A review of time series forecasting models, dating back to 1980, can be found in reference [19]. In addition, a review of time series forecasting techniques applied for building energy use can be found in reference [20]. Beginning with a definition, a time series is an ordered sequence of values recorded in time and over equal intervals [20]. Time series models have many different models and approaches; the following paragraphs will briefly discuss a couple of the most prominent techniques.

The moving average technique is one of the most common time series techniques available, similar to exponential smoothing. Both techniques assume that the time series is locally stationary (constant average and constant variance) within the past few time steps. Forecasts are calculated using averages of weighted values from previous time steps. The difference between the two techniques is that the moving average assigns equal weights to the coefficients within the average. In contrast, exponential smoothing gives greater weights to the most recent values. Grant et al. [21] compared the application of a few time series techniques with artificial neural networks in order to forecast the peak demand in a large government building. The results showed that the ANN models contained smaller forecasting errors than both linear regressions and simple moving average techniques.

The autoregressive integrated moving average (ARIMA) models are one of the most popular forecasting models applied throughout many different industries. The models are applied to stationary time series, that is, a time series with a constant mean and constant variance over time. The ARIMA forecasts values of a time series through weighted values of previous lags and lagged errors. The ARIMAX is an extended version of the ARIMA model which incorporates the use of regressor variables. Newsham and Brit developed an ARIMAX model in order to forecast the power demand of an office building [22]. The author's main contribution was the exploration of including occupancy as an independent variable. Occupancy information was obtained by: (i) using login information or (ii) via motion sensor. The authors noted that the addition of occupancy did not improve the model's accuracy significantly. By the addition of occupancy into the model, the error decreased to 1.217% (mean absolute percent error) from 1.224% without the use of occupancy as a regressor. Thus, a small increase was achieved using occupancy data as a regressor for the model.

2.1.3.1.3 Kalman filter

The Kalman filter is named after its founder, Rudolph Kalman, who introduced the algorithm in 1960 [23]. The Kalman filter is a linear model based on discreet state-space representation that performs its forecast in a recursive one-step ahead method. It is an explicit mathematical model which estimates the next steps value of the system and minimizes the covariance at each step during the iterations. These filters have widely been applied in navigation, navigation systems as well as signal processing and econometrics. However, they have not been widely applied in forecasting energy consumption in buildings.

2.1.3.2 Artificial Intelligence and Machine Learning

Artificial intelligence (AI) models are present in an increasingly large number of fields, and their applications can be seen to be growing throughout our daily lives. A few of the scientific fields include: computer science, linguistics, mathematics, psychology, neuroscience, business, sciences, and engineering. A few literature reviews of AI-based models and their applications to energy related studies within buildings is presented in papers [24-27]. Based on these papers, a few of the most popular algorithms are briefly presented within the following subsections along with their applications for forecasting of energy in buildings.

Machine learning (ML) is a subfield within the overall field of AI. To date, there are numerous definitions for both AI and ML as there has been no yet standard and agreed upon definition. Therefore, for the sake of this work the definition of ML which will be followed is based on reference book [28]: “a program is said to learn from experience E, with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E”. An addendum to this definition is based on reference [11], which states that typically ML provides an algorithmic approach which nonlinearly transforms the data.

A main advantage to ML algorithms is that they are adaptive, non-linear, and can learn by being presented data to improve on a certain task. To date, there are numerous ML algorithms and hybrid (combined) ML algorithms. A few of the most prominent algorithms include: k-nearest neighbors, k-means, decision trees, random forests, support vector machines, hidden Markov models, group method of handling data, artificial neural networks, etc. A few of the most prominent ML algorithms applied to forecasting building energy will be discussed in the following subsections.

2.1.3.2.1 Support vector machines

Support vector machines (SVM) are a supervised machine learning algorithm first developed by Vapnik in 1995 and are used for classification and/or regression analyses [29]. SVM are developed on the concept of distributing datasets by a decision hyperplane. The idea underlying the SVM model is based on finding the hyperplane which contains the largest margin between datasets. Details to the governing equations and theory of SVM can be found in reference [29]. SVMs were originally developed for classification problems, however, they extended to regression problems and time series-based forecasting after they showed promising results. In the case of regression, the input data is transformed using a non-linear kernel function, which maps the inputs into a high-dimensional feature space. As the model depends on the transformation to a nonlinear feature space, the general performance of SVM models depends on selecting the optimal kernel parameters. SVM models are among the most popular AI-based algorithms applied to forecasting and prediction in buildings. Examples of published work related to SVM and the forecasting of energy in buildings are presented in the subsequent paragraphs.

In references [30, 31], SVM models were applied in order to predict the hourly cooling load of an office building using data obtained from simulation. In both papers, the SVM showed slightly better performance than the ANN models applied (SVM 1.15-1.18%, ANN 1.19-2.36%). The authors noted that the architecture of an ANN was difficult to design. However, the neural network architecture was designed in a trial and error approach. Nevertheless, the papers did show the successful application of SVM for prediction with promising results.

Le Cam et al. applied an SVM in a cascaded based approach for various components of an HVAC in an institutional building [32]. The SVM model was coupled with physics-based equations to provide a hybrid grey box model. Target variables included: the supply air flow rate of an air handling unit, cooling coil load, building cooling load, and the electric demand of the fan, chiller, and cooling tower up to six hours in advance. The models presented within the paper showed good forecasting results. As such the originality of this paper is the overall cascade-based forecasting for the electric demand of an HVAC system incorporating an SVM as a hybrid grey box.

2.1.3.2.2 Random forest

Random forest (RF) is a supervised ensemble machine learning method used for both classification, regression, and feature selection. In essence, a random forest is a collection (ensemble) of multiple decision trees working in parallel where the result of each tree is then combined for a final output. RF have successfully been applied for a variety of forecasting applications within buildings.

Chae et al. [33] applied the RF algorithm for feature selection through permutation importance. The selected features were then applied to an ANN model in order to forecast the electricity usage and peak demand of a commercial building up to a day ahead. The selected variables were validated with the correlation coefficient. The overall forecasting results of the ANN over the 24-hour forecast horizon were promising.

Wang et al. applied RF models in order to predict the hourly building energy consumption of two educational buildings [34]. Models were compared to a regression tree technique along with a support vector regression technique. Within this work, RF was found to contain the lowest prediction errors among the forecasting methods. In addition, features were selected through permutation importance. Therefore, the contribution of this paper was the application of RF for both feature selection and prediction of building energy.

2.1.3.2.3 Artificial neural networks

Artificial neural networks (ANN) models are one of the most popular AI based methods. Such models are inspired by biological systems and the interconnection of neurons. McCulloch and Pitts first hypothesized how neurons may work and presented models based on simple electrical circuits in 1943 [35]. Next, in 1958 Rosenblatt modeled a simple single-layer perceptron for the classification of continuous values [36]. Since then, ANN models have obtained many breakthroughs which has aided their growth and popularity. One of the primary reasons for their continued application and growth is their abilities. ANN algorithms have the ability to capture and model nonlinear relationships between inputs and outputs learnt from data. Thus, such models are quite useful and adaptable for modeling complex systems. A more in depth look and review into how have ANN models been applied to research to date is presented in section 2.2.

2.1.3.2.4 Deep learning

Deep learning (DL) models are an emerging trend within AI. Deep learning has been defined as "representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level" [37]. The classification of DL models is based on the levels of nonlinear transformations/operations within the overall structure. Shallow architectures, or traditional architectures, have been the majority of the research to date and refer to such models with one to three levels of non-linear operations. In contrast, deep architectures contain four of more levels of such nonlinear operations [38].

Compared to traditional methods, deep learning based models offer some potential advantages and disadvantages. Starting with the advantages, firstly, it is common practice for the developer to select and extract good features between the input regressors and the output target values. The efficacy of the developed models depends on selecting good quality inputs. Thus, the selection of inputs requires domain expertise and engineering skill in order to develop effective models. In contrast, deep learning based methods can apply a general learning procedure and automatic learning, thus, not requiring such domain expertise. Therefore, a key advantage of DL based models in comparison to traditional models is that feature extraction may be automatically learnt [37]. Secondly, DL models can better handle and learn with large amounts of data; thus, as big data has become a problem in recent years, DL based models offer a potential solution to set problem. Finally, DL models can hold and store significantly more information within their models compared to conventional and traditional ML models. Therefore, this allows for more learning of distribution representations (learning many-to-many relationships between types of representations) and thus enables generalization to new values not explicitly shown in the learning data. However, DL models do have disadvantages as well. Such models may be more difficult and slower to train. This is a result of a larger number of hyperparameters needed for tuning and larger datasets. A more in depth review into how DL models have been applied to energy forecasting in buildings is presented in section 2.3.

2.1.4 Summary forecasting energy models

All energy forecasting models have their respective merits. Some are more applicable than others based on the data available, current conditions of the building (design stage or currently in operation), application, etc. A summary is provided for the different categories of energy forecasting models based on review papers [10, 39] and is presented in Table 2. The scope of this work is for the short term forecasting of the electric demand of an HVAC system. Due to their respective advantages, data driven models are to be applied within this work.

Table 2: Summary breakdown for categories of energy forecasting models

Approach	Required information	Software	Model complexity	Processing speed	Accuracy
Physics based	Detailed physical information	DOE-2, EnergyPlus, TRYSYS, etc.	Fairly high	Low	High
Data driven (Grey box)	Simplified physical information and historical data	Matlab, Python, C++, etc.	High	Low	Fairly high
Data driven (Black box)	Historical data	Matlab, Python, C++, etc.	High to fairly high	High	High to very high

2.1.5 Summarizing review papers for the applications of forecasting and prediction of building energy models

As data driven models have begun to increase in popularity in recent years, there have been a number of review papers published each focusing on a different aspect of building energy models. Therefore, this section is meant to provide a summary of the literature review papers over the past decade (2010 to 2020) and highlight the specific focus of each journal. It should be noted, that within such review papers, the distinction between forecasting and prediction was not stated; rather both words were used interchangeably. However, the lessons learnt and the overview of both data driven forecasting and prediction models is beneficial for understanding the effectiveness and limitations of such models.

The methodology for this review consisted of keyword based searches over available publication sources. Papers were screened to ensure that the review paper discussed forecasting and prediction energy models for buildings. As the primary focus of this review was directed towards forecasting building energy use; papers reviewing forecasting for electricity grids, electricity generation, grid management, etc. were omitted from the review. After the relevant review papers were found and screened, they were recorded over a standard set of criteria including: purpose/focus of literature review, summary results/ findings within, research gaps identified, and suggestions for future research. A chronological summary of the review journals is presented below discussing the focus of each review and some of the key points the authors recommended for future work.

In 2012, Zhao and Magoules provided a review for energy prediction models for buildings [10]. The authors compared the main categories of energy models including physics based, statistical, and AI based models. Future research directions proposed by the authors included: (i) the development of more accurate and effective prediction models, (ii) improved AI based hyperparameters searches, and (iii) the establishment of a database for various buildings and cases to help future researchers.

In 2013, Kumar et al. reviewed ANNs for building energy modeling and prediction [40]. Papers were reviewed for their applications throughout the building including: thermal loads, energy loads, indoor air temperature prediction, etc. Based on their review, the authors suggested that future work should focus on testing the performance and adaptability of ANN models for changing environments.

In 2014, Li et al. reviewed the integration of energy models with building operations and control [14]. The authors concluded that future work should focus on the reduction of computational costs while maintaining accuracy. Amhad et al. (2014) reviewed SVM, ANNs, and hybrid models for forecasting building electrical energy use [25]. Amhad et al. noted that each of the algorithms have their merits; therefore, it is difficult to decide which in fact may be better. However, combining models may help improve the forecasting performance.

In 2017, Wang and Srinivasan investigated AI based building energy prediction models comparing single point and ensemble forecasting models (multiple single point forecasting models integrated together into an overall forecasting model) [24]. The authors recommended that future work should include: (i) further applications of ensemble models, (ii) exploration of occupants and their impacts for building energy prediction, (iii) residential case studies, and (iv) studies which focus on the selection of optimal training size. Daut et al. (2017) reviewed both conventional and AI based forecasting approaches for building electric energy use [26]. Deb et al. (2017) reviewed time series based forecasting of building energy consumption with a focus on prominent and hybrid techniques [20]. Deb et al. noted that few papers explored: (i) the financial costs of building performance/control and, (ii) studies which cover wireless sensor networks for smart homes integrated with energy saving strategies.

Amasyali and El-Gohary (2018) provided a granular review of machine learning algorithms applied for building energy prediction [41]. Among their conclusions and recommendations for future research, the authors noted that deep learning based models have not yet been sufficiently studied and therefore require future research efforts. Furthermore, in 2018, Wei et al. reviewed data-driven approaches for prediction and classification of building energy use; focusing on mapping, benchmarking, retrofitting and prediction of building energy [39]. Ahmad et al. (2018)

reviewed forecasting, benchmarking, mapping, and profiling of building energy use [42]. Future work recommended by the authors included further studies applied to large scales and districts.

In 2019, Bourdeau et al. reviewed data-driven models for forecasting building energy [43]. Prominent algorithms (time series, statistical, and ML) for data processing and model applications were reviewed along with a breakdown of trends. Furthermore, in 2019 Mohandes, et al. provided a comprehensive review of ANN models for building analysis, HVAC applications (e.g. COP estimation) and indoor air temperature prediction [44]. Among their recommendations the authors proposed that future research work should focus on applying deep learning based techniques and non-typical target variables. In 2020, Sun et al. presented a review for data-driven models applied to energy predictions of buildings with a focus on a review: (i) for feature engineering, (ii) data-driven algorithms (statistical and ML), and (iii) factors considered for outputs (e.g. temporal granularity, scale, updating, etc.) [45]. The suggestions for future work in this review included: (i) industrial and hotel based case studies, (ii) using the same dataset to compare different data-driven models and approaches, and (iii) case studies focusing on the practical application of data-driven models.

To summarize and conclude a few of the key points over the various literature review papers focusing on data driven models for forecasting and prediction of energy in buildings: (i) AI methods are among the highest performing models, (ii) ANN and SVM models are the most popular AI based algorithms due to their high performance and flexibility, (iii) the most commonly applied ANN has been the standard feed forward neural network, and (iv) most case studies have been applied to commercial buildings, with hourly data, and forecast horizons of 1 hour or 24 hours ahead. In addition, it was noted among various literature review papers that future studies should focus on (i) component and occupant driven loads, (ii) apply sub-hourly data, (iii) explore the performance of ensemble based models, (iv) explore different types of ANN models, and (v) explore improved hyperparameter based searches. Due to their good performance, flexibility, and fast computational time, ANN models are selected for this work and further reviews of such models is to follow.

2.2 Literature review for the application of ANN in forecasting energy in buildings

This section conducts a literature review in order to identify relevant trends and gaps for the application of ANN for forecasting of energy within buildings. To date, there have been numerous

ANN models applied in published works; however, no two such papers are the same and there are variations among all papers. To further complicate matters, there are no set standard methods for definitions, creating models, tuning models, classifications, performance metric to use, etc. This makes understanding the work accomplished a challenging endeavor. In order to overcome this, a table was created in order to record the publications over a standardized set of criteria to systematically capture, quantify and process the data related to publications. Within the table each column corresponded to a specific section of relevant information regarding publication information, forecasting model, case study, and model performance. Examples of the columns used include: year of publication, application of the model, type of forecasting model(s), target variable(s), forecast horizon, features used, feature selection method, data size, ANN architecture selection method, and performance. After the construction of the table, a search for papers was conducted over the time range of 1990 until 2019 over available publication sources. Relevant papers were then cataloged within the table. The processing and analysis allowed for the objective questions of this literature review to be met, and highlighted relevant research trends, gaps and emerging techniques. Most of the contents of this chapter was published as J. Runge and R. Zmeureanu, "Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review," Energies, vol. 12, no. 17, 2019.

2.2.1 Objectives of ANN literature review

The objective of this literature review aimed to answer the question, “How have ANNs been applied to forecasting the energy use and demand in buildings?” This was accomplished by answering the following questions: How and where have ANNs been deployed? How have such models been developed before deployment? What are the performances of such models? What new trends are emerging?

2.2.2 Methodology for literature review

In order to answer the objective question of this literature review, a methodology was created to systematically capture information within each publication over a standard set of criteria. Each column would record information within the publication that was a common element in overall forecasting models (e.g. what forecast horizons were applied). From this point, each specific criteria could then be analyzed to provide a look at the big picture for how ANN models have been used at this specific element. The methodology used in order to conduct the literature review was composed of five main steps presented in Figure 1 and is outlined below.

Step 1: Conducting key-word based searches

Keyword(s) based searches of relevant articles was conducted through sites related to scientific publications of buildings and energy. Examples of keywords include: forecasting, prediction, neural networks, buildings, energy, data-driven models, electricity, heating, cooling, artificial intelligence, deep neural networks, etc. Keywords were combined in order to create overall keyword-based searches. Examples of keyword-based searches include: neural network forecasting prediction energy buildings, data-driven building energy, buildings forecast, deep learning forecasting, etc.

Step 2: Screening articles

Articles were downloaded and filtered based on the following criteria: (i) they contained a forecasting model targeting an energy load for a building, (ii) one or more forecasting models was an ANN, (iii) they contained sufficient information at a high granularity about the ANN forecasting approach, and (iv) the results of the forecast are presented within the paper. If a paper did not pass through the filters and meet all the criteria, it was removed from the cataloging process.

Step 3: Identifying and screening of additional articles

Articles which cited relevant publications were also used in order to help accumulate potential candidate articles. The selected articles were screened as well as per the process in Step 2 to ensure they were appropriate for cataloging.

Step 4: Reviewing all relevant articles

All relevant forecasting articles filtered were cataloged with high granularity in order to define their: purpose, application, data characteristics, forecasting model properties, model optimization, and performance.

Step 5: Analyzing the results of the articles

The results of the catalog were analyzed to find research trends of published work, limitations of models, and future research directions. A summary of set findings is presented in section 2.2.4.

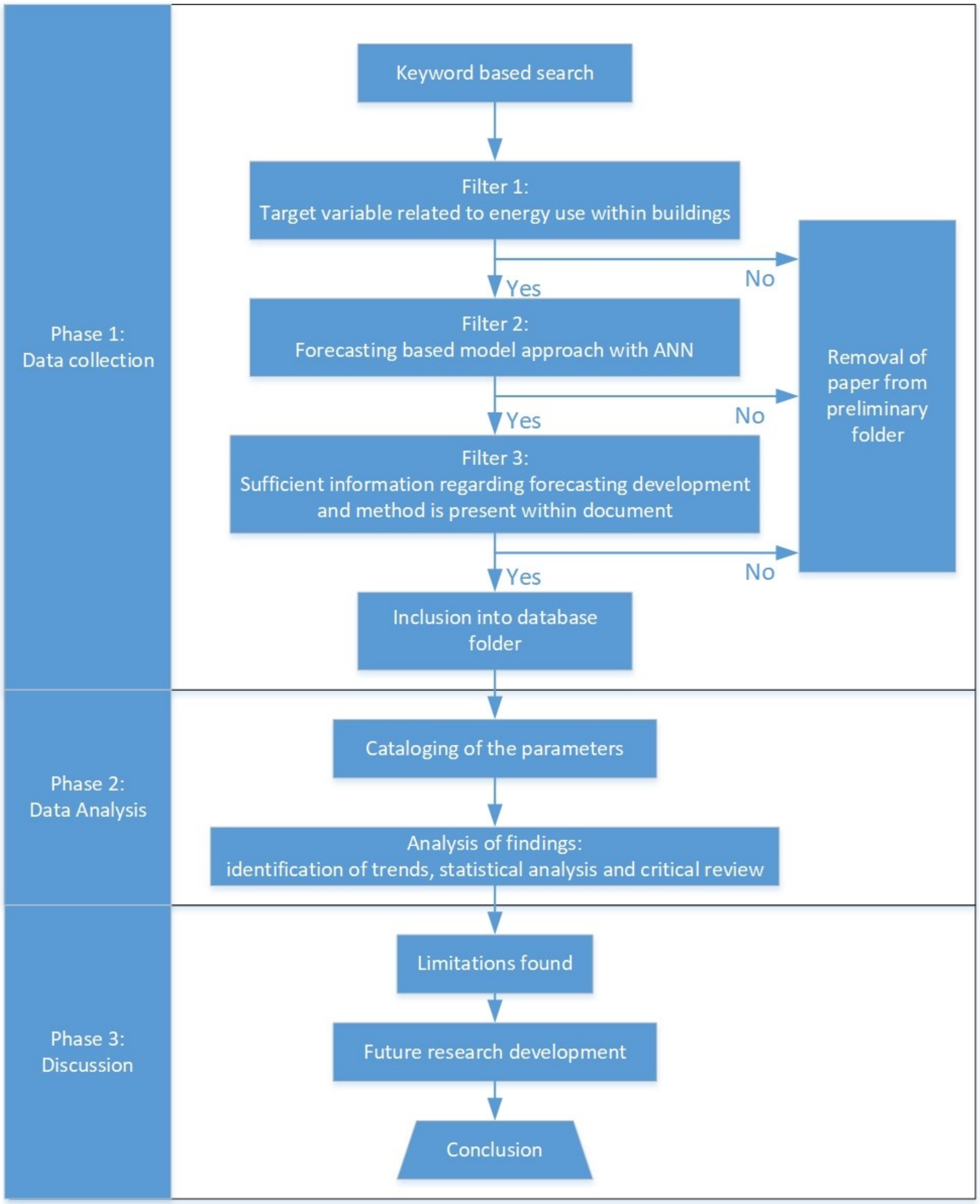

Figure 1: Literature review methodology

2.2.3 Limitations of literature review

Limitations for the state-of-the-art literature review include the allowed publication sources via Concordia University library's website. As such, only a few well-known publication sources were utilized and included: Google Scholar, Elsevier, Taylor and Francis, IEEE Xplore, ASHRAE transactions, and IBSPA.

2.2.4 Analysis of trends for ANN energy forecasting in buildings

The results of the data collection and filtering process found 91 papers over the specified time range. This section presents the results and analysis of trends found for the field. The references used for this analysis are [33, 47-136].

2.2.4.1 Timeline

The amount of publications per year is shown in Figure 2. An increase in forecasting publications can be seen to begin to occur in 2011, this may be attributed in part due to the breakthroughs in deep learning in 2010-2012 which began to re-popularize AI. Please note, that while the overall search for this review was conducted over the time range of 1990 until 2019; all papers found in the 1990s were prediction-based models. Thus, Figure 2 omitted the years 1990 to 1999.

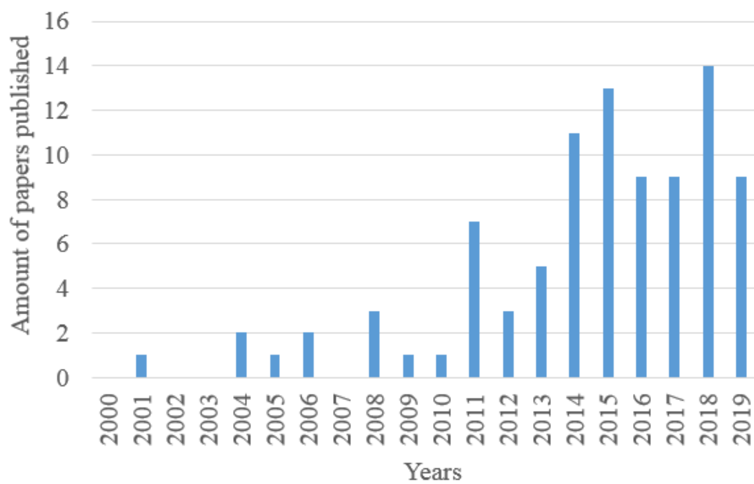

Figure 2: Timeline of ANN publications

2.2.4.2 Forecasting model types

ANNs are a data-driven model, which are typically broken into black box and hybrid model types as previously defined by ASHRAE in section 2.1.3. The majority of ANN forecasting models have been applied as black-box based models (84%), followed by hybrid-ensemble models (12%) and hybrid-grey-box models (4%) as shown in Figure 3. Despite their lack of publications, grey-box models should be used more, due to their flexibility in applications and their incorporation of the physical laws governing the system(s).

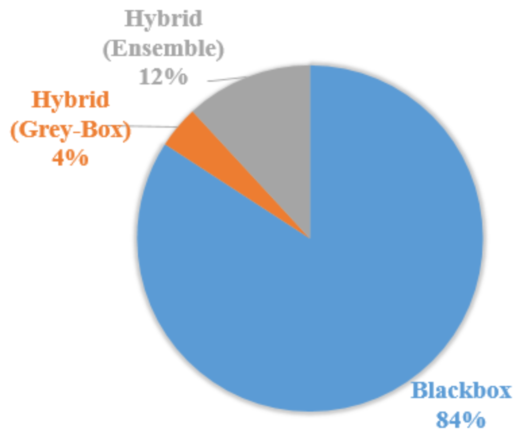

Figure 3: Application of ANN in forecasting model types

Further exploration of hybrid-ensemble based ANN models, revealed an approximate 75% to 25% breakdown between homogenous and heterogeneous models. This could be a result of the ease of development between the two. After creating an initial forecasting model, it is quicker to make a similar model rather than create a new model using a different technique. However, it should be noted, that the overwhelming majority of hybrid homogenous ensemble models applied a standard feed-forward neural network approach. Thus, there is a lack of heterogeneous ensemble models and homogenous ensemble models utilizing different artificial neural networks.

2.2.4.3 Application levels

The application level of the applied forecasting model refers to the level at which the target variable(s) were applied in relation to a building. The levels used in this study consist of: territory, building, sub-meter, and component. The territory level refers to a group of two or more buildings. As such, this can refer to a district system, neighborhood, or the full residential sector of a country. The building level refers to a target variable applied to a single building's load. Examples of these include a buildings overall heating, cooling, or electricity consumption. The sub-meter level refers to forecasting models were the target variable was applied on a sub-meter(s) within a building. For example, on a single circuit breaker for the kitchen or laundry room or mechanical room. The application is less than that of a whole building, however, contains multiple components or appliances together in a single dataset. The final level refers to a component-based application. These occur when the target variable was a single appliance or component within an overall building. Examples include a chillers electricity demand, the electric demand for an air handling unit fan, etc.

From the selected papers, most of the applications were found to focus on the building level (81%). This is perhaps due to the easier access to whole building meters and their data. The applications for the territory level accounted for 13%, component level 5%, and the remaining 1% accounted for the sub-meter level. Focusing within the 81% applied to the whole building, a breakdown of 83% to 17% was found between commercial and residential buildings. This higher prevalence towards commercial buildings may be a result of leveraging existing data from Building Automation Systems (BAS) already installed in commercial buildings. It can be more difficult and costlier to obtain data from residential building, sub-meters, and components without incurring extra costs associated with the installation of sensors and recording systems.

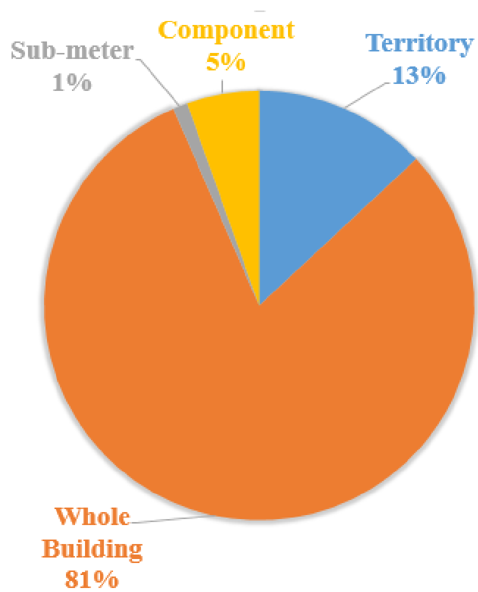

Figure 4: Building application level of ANN forecasting in buildings

Forecasting at the component level could be more beneficial for building energy management rather than at the whole building level. This could help identify where the largest consumers are and apply energy saving applications to such components. Hence, future research should focus on applications of components within buildings which consume a large amount of energy demand and usage.

2.2.4.4 Target variables

The target variables of the ANN forecasting models were classified into six main categories: whole building, heating, cooling, lighting, natural gas and components (Figure 5). The whole building category refers to such target variables applied to the demand/consumption of the: (i) buildings electricity, (ii) buildings overall energy, and (iii) buildings yearly energy. Next, the heating loads, cooling loads, and lighting load categories refer to published work in which the target variable was either a thermal or lighting load to be forecasted by the models. Finally, the component category refers to target variables which consisted of component(s) within the overall building. Examples include: ground source heat pump, electricity consumption of the chiller, plant, fans, etc.

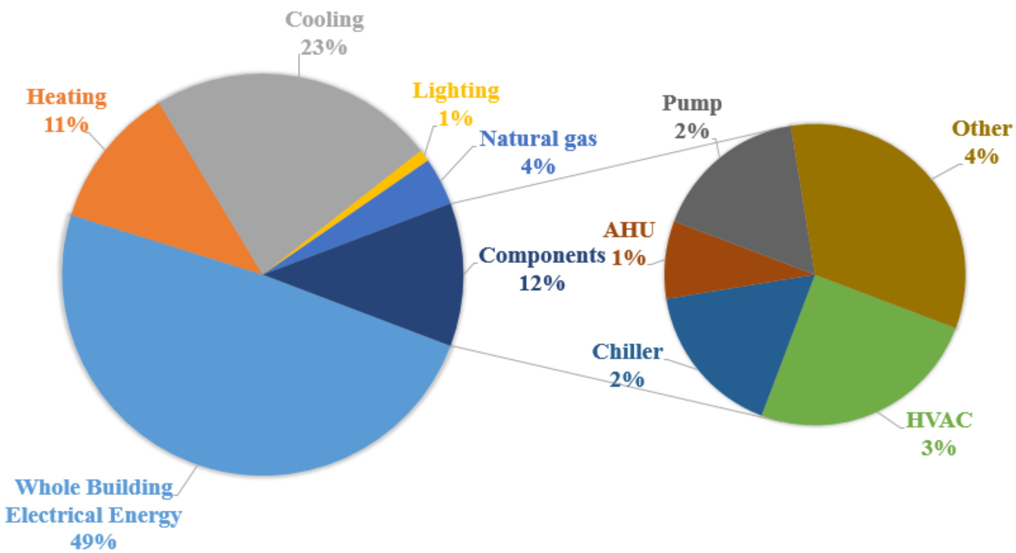

Figure 5: Categorical breakdown of ANN forecasting model target variables

Whole building target variables accounted for 49%, heating 11%, cooling 23%, lighting 1%, natural gas 4%, and components 12%. Within the component target variables, pumps accounted for 2%, AHU 1%, chiller 2%, HVAC 3% and other 4% (Figure 5). Other refers to devices such as a ground source heat pump, reheat, sub-meter, fans.

2.2.4.5 Forecast horizons

The forecast horizon is the length of time into the future over which the forecast is made. This can be composed of a single step model or multiple step ahead model. Short term forecasting (sub-hourly to a day ahead) were found to constitute the majority of case studies applied to date. Sub-hourly here refers to forecasts below a one-hour horizon (e.g. 1, 15, 30, and 40 minutes). Please note, that for the daily horizon, a single value for the entire day energy consumption can be predicted, or a load profile for the day can be forecasted (e.g. using hourly steps of a 24 hr. load profile). Both these are differentiated in Figure 6. The distribution of the number of papers with respect to the forecast horizon is illustrated in Figure 6: sub-hourly 10%, hourly 25%, multiple hours 6%, daily (profile) 29%, daily total load (6%), multiple days 6%, week ahead 5%, multiple weeks 1%, monthly 5%, multiple months 1%, and 6% for yearly horizons.

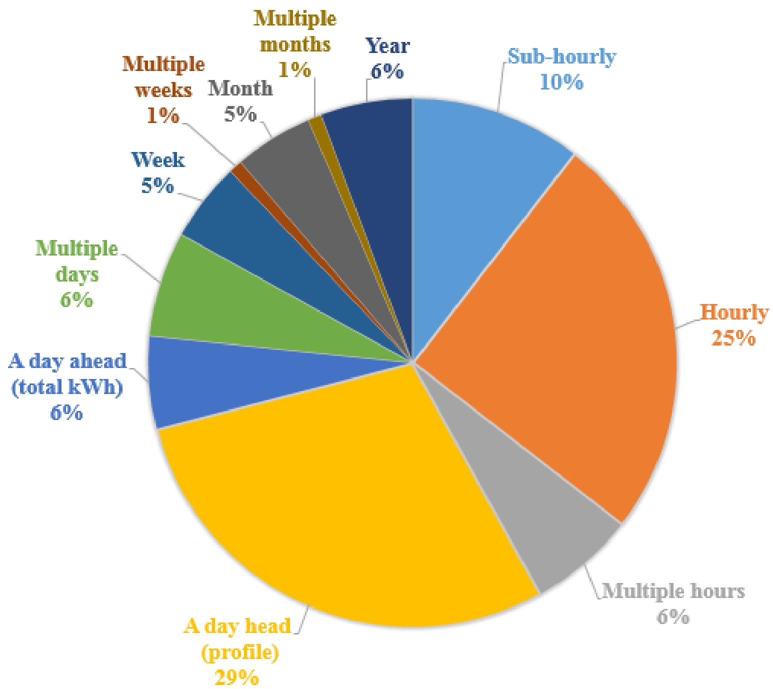

Figure 6: Forecast horizons for ANN models

Hourly horizons corresponded to the most common forecast horizon. This may be a result of a few reasons. First, the granularity of the data obtained contributes to the overall time step of the final model. Many papers used weather dependent variables as inputs into their models. The weather data was typically obtained off-site and through an online and open source which contained data on an hourly scale. Thus, the forecasting models are typically built with the time step of the data obtained. Secondly, a few papers which presented data on a shorter time steps (e.g. 1 min) and aggregated their data to hourly intervals in order to reduce the error of time and system delay [137]. Thirdly, a factor which could contribute to the prevalence of hourly models is the cost of energy. Typically, utilities bill on an hourly time scale (for electricity) [138].

2.2.4.6 Data applied

Three types of data are available for models: (i) synthetic/simulated, (ii) benchmark, (iii) and real/measurement data. Real data refers to measured data obtained from various BAS, electricity meters, weather/climate stations/utility bills/national reports, and surveys. Synthetic and simulated data was obtained from building simulation software such as: EnergyPlus, eQuest, TRYNSYS, and DeST. Benchmark data is obtained from publically available datasets e.g. ASHRAE great energy shootout. Data obtained from measurements contributed to the majority (85%) of the cases presented with synthetic or simulated data accounting for 14%, and benchmark accounting for 1% of the type of data used (Figure 7).

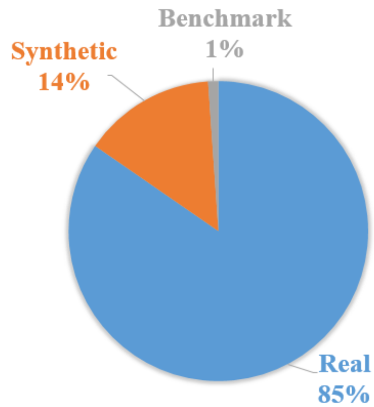

Figure 7: Data type breakdown for ANN models

2.2.4.7 Performance range

The performance metrics found in reviewed papers are the: mean absolute percent error (MAPE), coefficient of variation of root-mean-square-error (CV(RMSE)), mean absolute error (MAE), mean bias error (MBE), mean squared error (MSE), and the coefficient of determination (R^2).

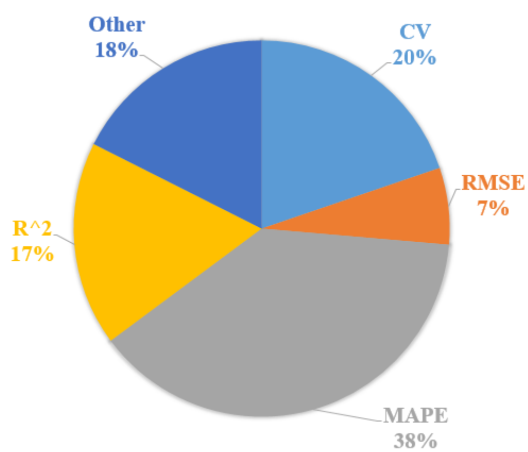

Figure 8: Breakdown of performance measures applied for ANN models

Prior to beginning the literature review, a problem concerning which performance metric to catalog became apparent; many papers presented multiple performance metrices (as there is no set standard for forecasting models). The recording of all metrices would result in the overall cataloged table becoming cumbersome. Therefore, an approach for selecting which performance metrics to record was created. First, the CV(RMSE) / CV % was the main or first performance measure to be selected. If CV(RMSE) was unavailable, the RMSE was recorded. These two measures were selected first as they are the recommended performance measures by ASHRAE [139]. If CV(RMSE) or RMSE was unavailable, then MAPE was selected. If MAPE was unavailable, then R^2 was selected. If R^2 was unavailable, then the most relevant error metrices was selected as indicated by other in Figure 8. Figure 8 presents the breakdown of performance indices used: MAPE (38%), CV(RMSE) (20%), other(18%), R^2 (17%), and RMSE (7%).

Table 3: Performance range of ANN forecasting models

		Sub-hourly	
Paper ID #	***Time step***	***Forecast horizon***	***Error***
[100]	15 min	15 min	0.001-0.059% (MAPE)
[112]	30 min	30 min	0.939-8.34% (MAPE)
[60]	5 min	40 min	13.2-14.4% (MAPE)
		Hourly	
Reference	***Time step***	***Forecast horizon***	***Error***
[112]	1 hr	1 hr	0.95 – 19.1% (MAPE)
[95]	1 hr	1 hr	36.5 % (MAPE)
[127]	15 mins	1 hr	4.5-5.4 % (MAPE)
[96]	5 mins	1 hr	8.59-23.86% (MAPE)
		Multiple hours	
Reference	***Time step***	***Forecast horizon***	***Error***
[106]	12 hrs	12 hrs	5.03-7.4% (MAPE)
[84]	1 hr	1-6 hrs	7.30-8.09% (CV(RMSE))
[67]	15 mins	1-6 hrs	30% (CV(RMSE))
		Daily (profile)	
Reference	***Time step***	***Forecast horizon***	***Error***
[78]	1 hr	24 hrs	1.04-4.64% (MAPE)
[123]	1 hr	24 hrs	11.56% (MAPE)
[56]	15 mins	24 hrs	2.59-5.42% (MAPE)
[111]	15 mins	24 hrs	36.86-42.31% (MAPE)
		Daily (load)	
Reference	***Time step***	***Forecast horizon***	***Error***
[59]	Daily	Day ahead	4.75% (MAPE)
[55]	Daily	Day ahead	6.63-17.64% (MAPE)

Table 3 presents the breakdown of the performance ranges for ANN models. Only the short-term (sub-hourly to daily) forecasting values are presented as this is the main scope of this work. In addition, it should be noted that only errors of the forecasting energy demand/use are presented. A few papers provided the results of target variables different than energy consumption (e.g., air flow rate, temperature, supply fan modulation, relative humidity), the error range of those variables were omitted from Table 3. Furthermore, when selecting the error range with each paper, only the highest and lowest values were recorded (regardless of ANN architecture, type, dataset). Within Table 3, first the single step forecasts (t+1) are provided followed by multistep forecasts (t+1 to t+n).

2.2.5 Summary and conclusions of ANN literature review

Generalizing the results found from the type of ANN models applied; the overwhelming majority of ANN models were applied as black-box based approaches, applied a feed forward neural network (FFNN), and heuristically selected the architecture of the ANN model. Furthermore, generalizing the results of the case studies to which such models have been applied showed: the majority of such models have been applied to CI buildings, used hourly data, targeted a whole building energy load (overall energy, electricity, thermal), and had a forecast horizon of up to 24-hours ahead.

Due to the results of this analysis, the following research gaps can be observed:

- Grey-box or ensemble forecasting models
- ANN models other than FFNN
- Iterative forecasting models
- Target variables of natural-gas, lighting, residential, sub-meters, and components
- Forecasting models applying sub-hourly data
- Forecasting models using sub-hourly, multiple hours, or multiple days in advance
- Case studies of industrial, territorial, and residential buildings

Furthermore, the analysis conducted found the overall performance ranges for ANN energy forecasting in buildings to be: (I) 0.001–36.5% (MAPE) for single step ahead forecasting, and (II) 1.04–42.31% (MAPE) for multistep ahead forecasting. Additionally it can be observed in multistep ahead forecasting that the performance ranges have be 1.04–11.92% (MAPE) with the application of hourly data and 2.59–42.31% (MAPE) with sub-hourly data. Finally, the results of this analysis showed a new emerging trend within recent years for ANN forecasting, the emergence of deep learning and deep neural networks.

2.3 Literature review for the application of DL techniques in forecasting building energy

In the literature review of section 2.2 focusing on ANNs, an emerging trend was observed in recent years with the emergence of deep learning (DL) and deep neural networks (DNN) forecasting models. This section conducts a literature review for the application of DL for forecasting building energy use. Most of the contents of this chapter was published as J. Runge and R. Zmeureanu, "A Review of Deep Learning Techniques for Forecasting Energy Use in Buildings," Energies, vol. 14(3), no. 608, 2021.

2.3.1 Objectives of DL literature review

The objective of this literature review is to answer fundamental questions related to the applications of deep learning techniques for forecasting energy in buildings. Therefore, this literature review aims to answer: (i) How and where have deep learning based techniques been applied for forecasting energy use in buildings? (ii) What are the prevailing DL forecasting model types which have been deployed? (iii) Has there been any performance effects from applying the DL based techniques compared with other ML or data-driven models?

2.3.2 Methodology for literature review

The methodology applied for the collecting, filtering, and cataloging of papers is similar to the literature review methodology presented in section 2.2.2. However, an additional filter was used for screening to ensure that a deep learning based model was applied within the published work. It should be noted, that DL and DNN are not exactly the same. DL refers to ML models which contain multiple levels of nonlinear transformation; whereas DNN models are the application of such strategies to neural networks. To the best of the author's knowledge, there have been three main ways to date in which DNN have been applied as DL based techniques:

I. Increasing the number of hidden layers in feed forward neural networks
II. The application of recurrent neural networks. Such models may have a single or multiple hidden layers, however, even single layer recurrent neural networks may be considered deep learning due to training approaches. Unfolded, which occurs in

training of the network, such models are consider networks with very deep structures as information from previous states is passed to current states [141].

III. Through the sequential coupling of different types of algorithms into overall structures (example, an autoencoder coupled with an SVR forecasting model)

This work applied the same review methodology previously described in section 2.2.2. However, an adjustment was made to the filtering process to ensure DL based techniques were applied within the published paper. Furthermore, the range of dates from this literature review was modified and is from January 2000 until October 2020.

2.3.3 Limitations of literature review

The first limitation of this literature is similar to that of section 2.2.3, and refers to the available publication sources. Only such well known sources available through Concordia University library's website were used. Such publication sources include: Elsevier, Taylor and Francis, Google Scholar, IBSPA, IEEE Xplore, and ASHRAE transactions.

The second limitation for this review is a restriction on the applied DL techniques for forecasting. Specifically this review focuses on DL techniques applied to forecasting energy use in buildings. While the applications, developed models, and approaches applied in other fields may have useful information or approaches, they are beyond the scope of this review as they fail to contribute to answering the underlining questions.

2.3.4 Analysis of trends for DL energy forecasting in buildings

The results of this work found 63 papers over the time period of January 2000 until October 2020 which have applied DL based techniques for the forecasting of building energy. This section presents the results of the analysis of trends found and is based on references [119 - 136] and [142 - 185].

2.3.4.1 Timeline

Figure 9 presents the results for the number of publications each year over the specified time range of January 2000 until October 2020. An increase in publications can be seen beginning in 2016, this may be attributed to breakthroughs in DL achieved in 2010 to 2012 [37] which then began to branch out to various fields including forecasting energy use in buildings.

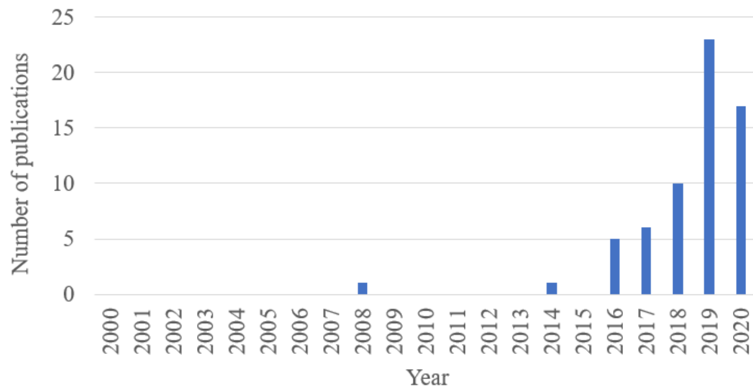

Figure 9: Timeline of DL building energy forecasting publications

2.3.4.2 Forecasting model types

To date, the overwhelming majority of models (91%) have been applied as a black-box based approach while the remaining 9% have been applied as ensemble models. Based on this review, no such DL models have been applied as grey-box models to date.

2.3.4.3 *Application levels*

The breakdown for the application levels of the applied case studies is presented in Figure 10. As previously described, the application level refers to the level at which the DL forecasting model has been applied and consists of: territory (multiple buildings), whole building level (an overall energy load for a single building), sub-meter (an energy load within a building, less than an overall energy load, that consists of a group of components aggregated together), and component (a component or system within the building). Based on this review, the distribution of applications were found to be: whole building (53%) followed by territory (37%), sub-meter 6%, and component (4%). A table for the territory based applications is shown in Appendix A, whole building based in Appendix B, and both subsystem and component in Appendix C.

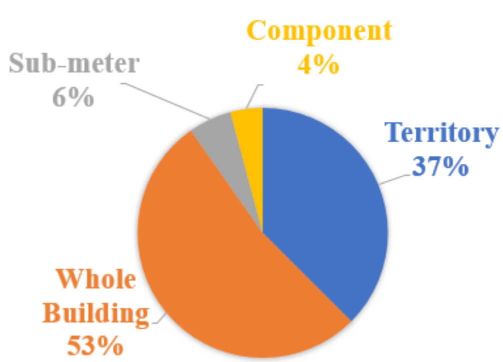

Figure 10: Application level for DL forecasting models

2.3.4.4 Target variables

The target variables for the DL forecasting models were similarly cataloged into six main categories as previously described in section 2.2.4.4 and consist of: whole building, heating cooling, lighting, natural gas, and components. Whole energy, heating, and cooling, lighting, and natural gas refer to an energy load for a whole building. In contrast, component(s) refer to target variables which consists of component(s) and systems within the overall building. Examples of component based models include: HVAC system, ground source heat pump, compressor pump, etc. Figure 11 presents the breakdown for target variables applied for the DL forecasting based models. A distribution of the target variables over published work was found showing: whole building (54%), heating (19%), cooling (15%), lighting (0%), natural gas (6%) and components (6%).

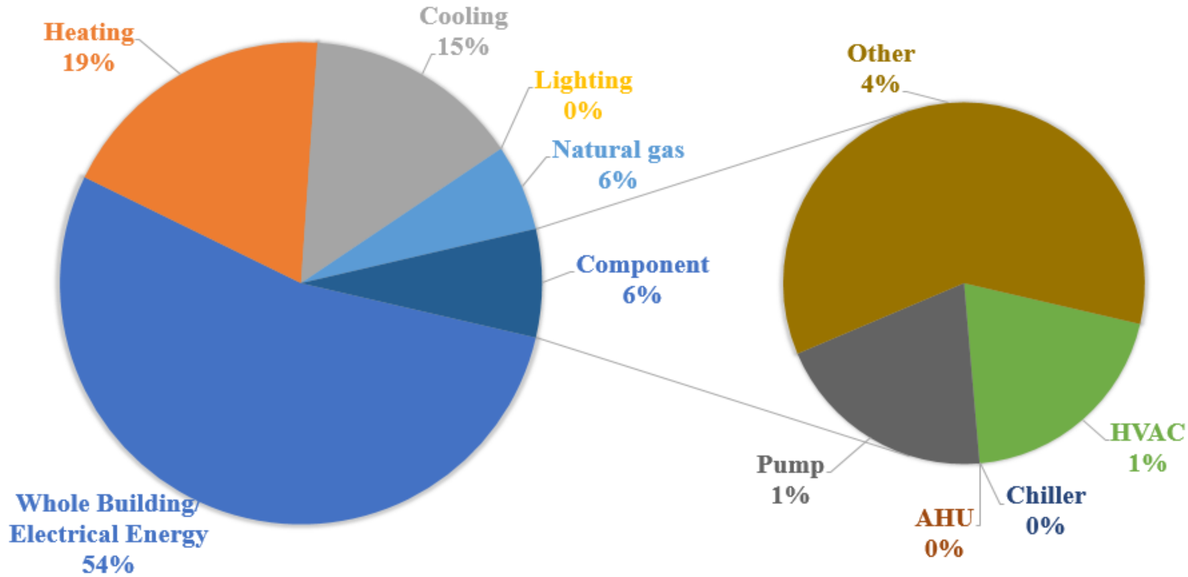

Figure 11: Target variable breakdown for the DL forecasting models

2.3.4.5 Temporal granularities

Forecasting models have two main temporal characteristics to consider: the time step of the data and the forecast horizon. Based on this analysis, the time steps for the applied deep learning models were found to be 1% yearly, 0% monthly, 3% weekly, 6% daily, 41% hourly, and 49% sub-hourly. Focusing on the forecast horizon, this analysis found a breakdown of sub-hourly (19%), hourly (22%), multiple hours ahead (11%), 24-hours ahead (27%), a day ahead (2%), multiple days ahead (7%), a week ahead (5%), a month ahead (2%), and a year ahead (5%). The distribution of the forecast horizon for DL models is presented in Figure 12.

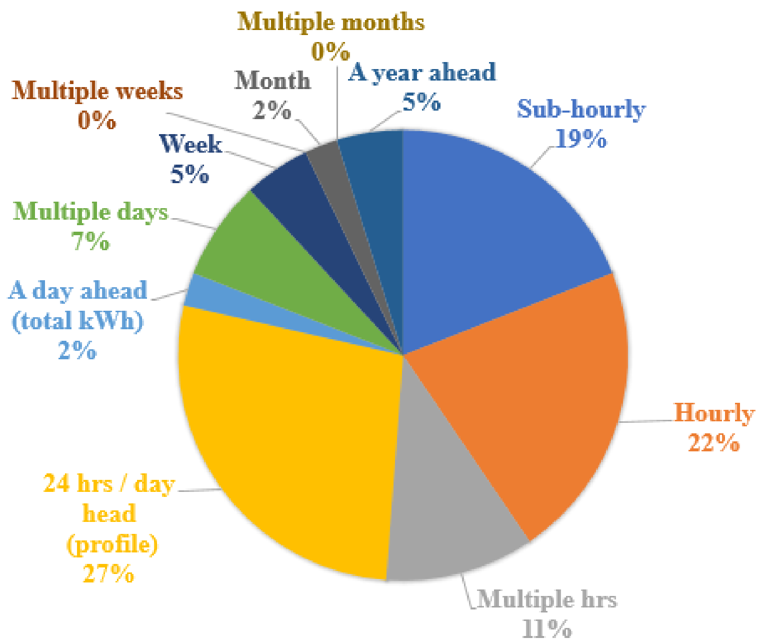

Figure 12: Forecast horizons for DL forecasting models

2.3.4.6 Data properties

Data properties refer to the type of data applied as case studies and the length of data for each case study. Based on the analysis it was observed that the majority (95%) of DL models have been applied to case studies based on measurement data, followed by 3% for synthetic data case studies and 2% for benchmark data case studies. Focusing on the time length of data used in each case study, it was observed that 17% of the models applied under six months of data, 22% used six-months to one-year, 58% applied greater than one year of data, and 3% did not specify the amount of data applied in their case studies.

2.3.4.7 DNN forecasting models applied

The type of DNN forecasting model(s) was recorded in each paper and for each case study. The most common types include: deep feed forward neural network (D-FFNN), Convolutional neural network (CNN), Restricted Boltzmann machine (RBM), Recurrent neural network (RNN), Gated recurrent unit (GRU), Long short term memory neural network (LSTM), Deep belief neural network (DBN), and other. Here other refers to combined models, modified models, or deep deterministic policy gradient. Figure 13 provides a breakdown for the type of the DNN models found through this analysis.

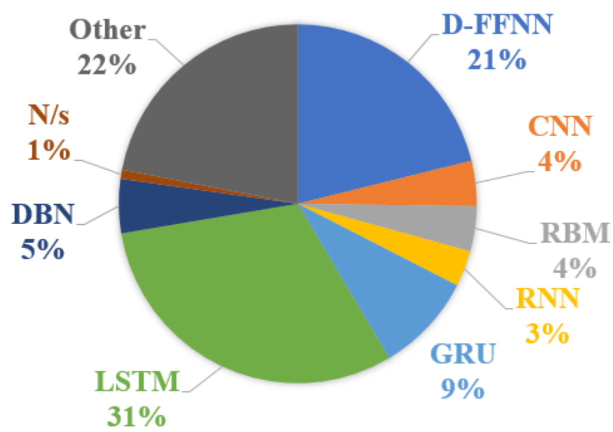

Figure 13: DL forecasting models applied

2.3.5 Summary and conclusions of DL literature review

Generalizing the results found from this review, the majority of DL models applied have been an LSTM or Deep FFNN based model, applied as a black box based approach, used hourly input data from existing buildings, targeted and overall energy load within a building or territory, and applied a forecast horizon of one-hour or 24 hours ahead.

In spite of their short age, DL based techniques have begun to be applied to building energy forecasting based research in increasing amounts. However, there still remain numerous gaps in research related to such models; a few of the key research gaps identified include:

- A lack of DNN models applied to a component or system within a building
- Few papers which have applied DL feature extraction models coupled with DNN forecasting models
- A lack of grey-box and ensemble based DL models
- Few papers which have applied multiple hour forecast horizons, multiple days ahead, and medium to long term forecasts
- The enrichment of DL techniques across a variety of building types, with an emphasis on comparison based papers and studies
- The establishment of guidelines for DL model development; including automation of the hyperparameter selection
- The establishment of scalable DL based models which can be developed and tuned in a timely manner for practical implementations across different buildings and systems

- The development of robust models which can continue to provide accurate forecasts in the event of changes of operation, sensor failure, etc.

Furthermore, the analysis conducted found that a DL based technique applied for feature extraction purposes typically led to an increase in forecasting performance compared to other such techniques. However, there were a few instances observed in which it did not. Moreover, when the DL based techniques were applied as forecasting models, similar observations were observed.

2.4 Thesis objectives

In the domain of demand response, fast and accurate tools are needed in order to forecast the electric demand of the HVAC system. Such models can help building operators and energy managers to plan fast DR-based strategies. This thesis contributes to the DR research field by presenting methods to forecast the electric demand of an HVAC system and components with a short-term forecast horizon. The main objectives of this work include:

1.) Firstly, this thesis proposes a short-term grey box model consisting of an ensemble of nonlinear autoregressive neural networks coupled with a physics-based model to forecast the electric demand of a supply fan (component-level) of an existing HVAC system. This forecasting model will be developed and validated using measurements from a building automation system.

2.) Secondly, this thesis proposes the application of deep learning techniques in order to forecast the electric demand of an HVAC system (system-level) and validation of the forecasting model on a case study building using synthetic data and measurement data.

2.5 Thesis overview

Chapter 3 describes the proposed methodologies applied for the short term forecasts of the HVAC electric demand within the scope of this thesis. First, the overall forecasting methodology is presented. Next, two different methods are presented, the component level and the system-level method. In addition, the governing equations for the system-level forecasting model are presented within Chapter 3 as this approach is applied within two separate chapters (5 and 6) over different case studies. Therefore, to prevent the repetition of the governing equations for the DL models, they were placed with the methodology of the system-based model. Finally, the methodology applied for data conversion is presented at the end of Chapter 3.

Chapter 4 presents the application of the component based method. The component model is applied to forecast the electric demand of a supply fan of an air handling unit for an institutional building.

Chapter 5 and 6 explores the application of the system-based methods applied to a single case study with two different data sources. The case study is the Genomic research center located at the Concordia University Loyola campus. Chapter 5 presents the system based approaches applied with a synthetic data source while chapter 6 applies the same models to measurement data from the building BAS system.

Chapter 7 provides a comparison of the system based monolithic model using airport weather data as an input to the forecasting model through two different approaches. First, the monolithic model is applied substituting current and past historical data from the closest airport as inputs to the forecasting model. Next, a monolithic model incorporating future weather forecasts of airport data as an input is applied.

Chapter 8 concludes this research with a discussion of the contributions, limitations and the potential for future work.

Chapter 3: Proposed methods

The scope of this research is the development of forecasting methods for estimating the future electric demand of the HVAC system. Within this work, two methods have been proposed: the component based method and the sequential based method. The component based method provides forecasts for the electric demand of a component within the overall HVAC system. In contrast, the system based method provides forecasts for the overall HVAC system's electric demand. The governing equations for the system level forecasting model is presented in section 3.3. This is a result of the application of the model in both chapters 5 and 6. Therefore, in order to avoid confusion and repetition, such equations are presented alongside a description of the approach. However, the physics based equations for each case study will be presented within each of its respective chapters.

3.1 Forecasting methodology

The forecasting methodology used within the proposed thesis is shown in Table 4. This method is the generalized forecasting methodology that will be applied to all models within the thesis with minor modifications as necessary (e.g. development of multiple models). The methodology is based on a forecasting methodology originally proposed in reference [186], however, it has been modified in order to meet the objectives for this work.

Table 4: Proposed forecasting methodology

Step Number	Description of Tasks	Definition of Tasks
1	Problem Definition	- Identifying problem - Listing of objectives - Listing of governing equations and assumptions
2	Information Gathering	- Exploration search for data over available data sources - Extracting the data from available data source(s)
3	Preprocessing of data	- Organization and time synchronisation of data - Identification, removal or replacement of missing data, outlier data, and erroneous data
4	Preliminary analysis	- Exploratory analysis; graphing of data, simple statistics - Feature selection - Breakdown of data into training, validation, and testing sets
5	Forecasting model construction	- Feature scaling (normalization) - Hyperparameter selection based on tests in training data - Selection of forecasting architectures
6	Evaluation of forecasting models	- Training, validation, and application of forecasting models selected to the testing data - Recording the performance of the applied models
7	Comparison of approaches	- Comparison of forecasting models and/or approaches

3.2 Component-level forecasting model

The component-level model is applied to target forecasting a component within the overall HVAC system. The model applies the overall forecasting methodology as shown in Table 4. For this approach, the forecasting model is applied to a variable which is monitored and controlled within the overall HVAC system by the BAS. However, the electric demand of that component is not a measured and recorded variable. Therefore, the future forecasts for the electric demand of the component are obtained from the forecasts of the control variable applied to a physics based equation for that component. Consequently, this approach leverages existing BAS controlled variables without the need for additional permanent sensor placement. For this work, the regressors applied to the forecasting model are limited to those of current and past historical measurements.

This work will, extract the controlled variable from the available BAS data and then apply processing steps to the data. Next, an automated hyperparameter search is conducted in order to find the optimal hyperparameter values to apply for the selected model. The application of the top performing architectures will then be applied to the testing data in order to forecast future values of the controlled variable. The output forecasts of the controlled variable will then be passed to a physics based equation in order to forecast the future electric demand of a component within the overall HVAC system. A copy of the Matlab code for the component model is provided in Appendix D.

3.3 System-level forecasting model

This section presents the system based forecasting model proposed. One of the main differences for the system based model compared to that of the component based model is that this method targets forecasting the overall system, whereas, the component based method targets a component within the overall system. In addition, while the component-level model applies a hybrid-grey box approach, the system-level model applies a black-box based approach.

Two different system level forecasting approaches are applied: a monolithic approach and a sequential approach. Furthermore, both approaches apply the same overall forecasting model, the autoencoder and LSTM ensemble. Both system-level approaches are explored over a case study of a building with two different data sources. The first data source is a calibrated eQuest simulation of the building previously completed and shown in reference [187], the output of this simulation thus provides a synthetic data source. The second data source for this building is the BAS currently installed and monitoring the operations within the building. Based on the two system level approaches and the two sources of data; this work explores four different scenarios:

I. Synthetic data case study with a system-level monolithic approach;
II. Synthetic data case study with a system-level sequential approach;
III. Measurement data case study with a system-level monolithic approach; and,
IV. Measurement data case study with a system-level sequential approach.

3.3.1 Monolithic approach

The system level monolithic approach consist of a single large DL forecasting model targeting the electric demand of both the primary and secondary systems over the forecast horizon. Figure 14 presents the monolithic forecasting model. The input regressors for this model rely strictly on current and historical values. This approach is desired to explore the effectiveness of DL for forecasting with a large amount of inputs when applied to the HVAC system.

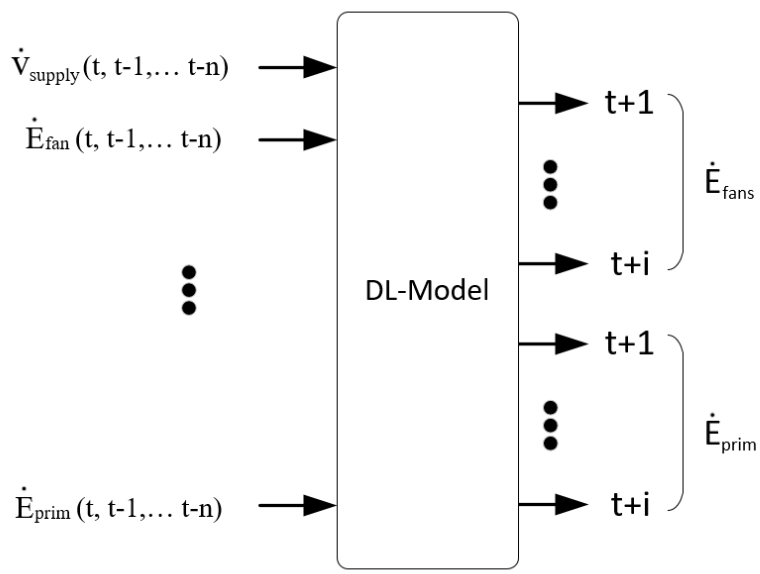

Figure 14: Monolithic approach

3.3.2 Sequential approach

The second system level model consists of the sequential approach. This approach applies multiple DL models each targeting a specific energy load within the HVAC system. Forecasts are generated by a DL model, and then sequentially passed on to be used as an input along with historical data to the next DL forecast model. Thus, this approach sequentially forecasts multiple target variables, and sequentially passes the output forecasts to be used as an input to the subsequent model. Figure

15 presents the sequential model to be applied, the target variables for the models, and a few of the input regressors which will be applied based on reference [188]. The sequential approach is desired in order to apply smaller DL based models and explore the ability of such models in learning more information from historical data.

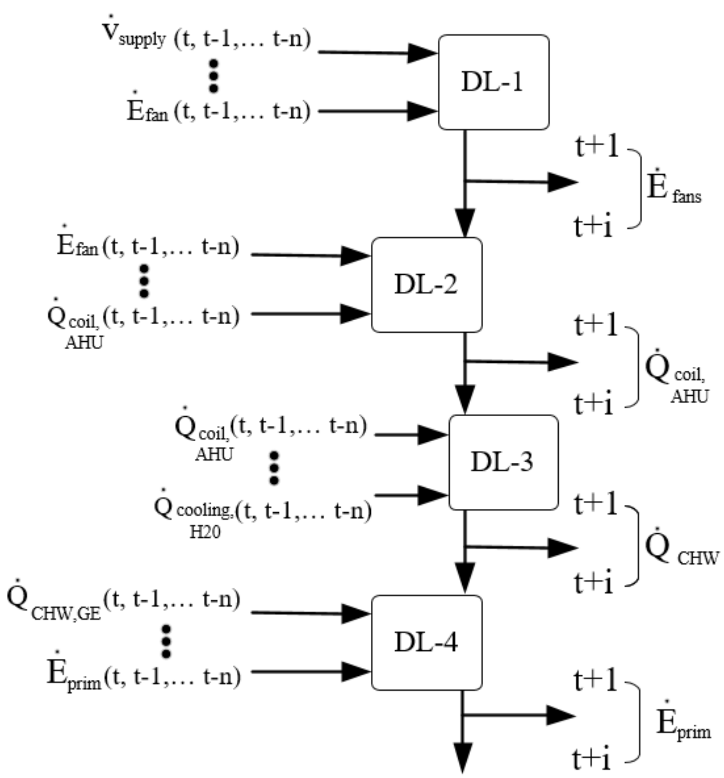

Figure 15: Sequential approach

3.3.3 System based forecasting model: An autoencoder-LSTM ensemble

The system based forecasting models (monolithic and sequential) will consist of an autoencoder coupled with an ensemble of LSTM neural networks as depicted in Figure 16. An ensemble approach was desired due to their ability to increase the stability of the forecasting performance through time. The ensemble will consist of four LSTM models each coupled together through an equal weights approach to provide the final output forecasts for the respective target variable. A copy of the Python code for the system based forecasting model is provided in Appendix E.

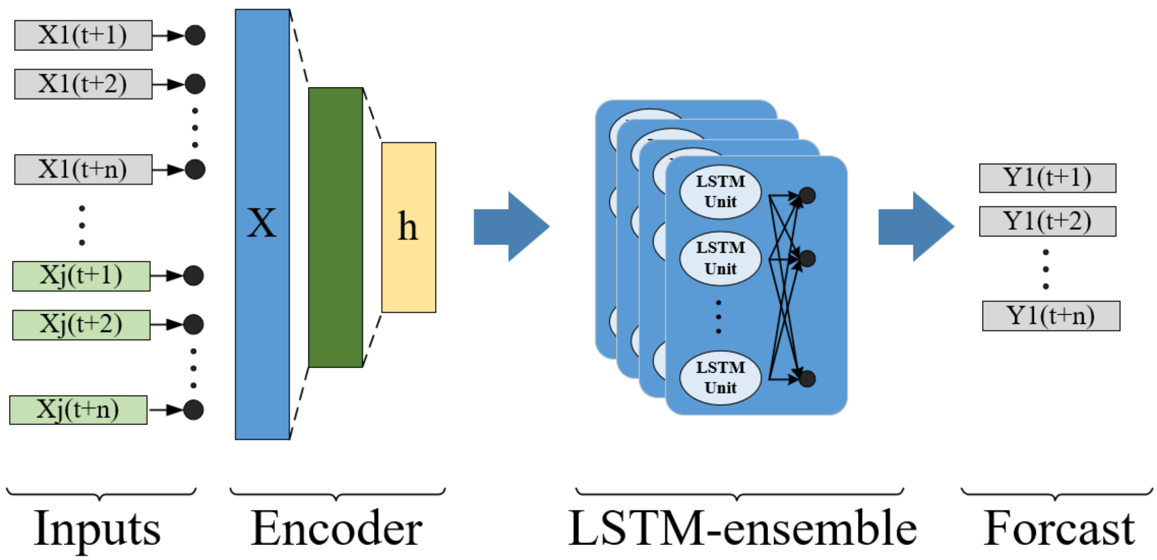

Figure 16: Auto-encoder and LSTM ensemble

3.3.4 Autoencoder applied for feature extraction

Feature extraction is the process of taking the original dataset and reducing the dimensionality to a more manageable group for processing [189]. For this work, the encoder section of an autoencoder is applied in order to reduce the dimensionality of the input data. An autoencoder (AE) is an ANN model which is trained to reconstruct the input of the ANN as an output. Figure 17 presents the overall structure of the AE, which consists of two main parts: an encoder model and a decoder- model [141]. The purpose of the encoder model is to map the input dataset into a hidden representation in a hidden space. The decoder takes the hidden representation and maps it to an output. Therefore, given an input dataset X, the encoder maps the input to a hidden space h=f(x). The decoder then takes the hidden representation in order to provide an output g(h)=X'. Autoencoders can be classified into two main types: (i) under complete and (ii) overcomplete. An under complete AE is where the number of hidden layer neurons is less than the number of input/output neurons. In contrast, and overcomplete AE is where the number of hidden layer neurons is greater than the input/outputs. For this work, an undercomplete AE is applied as an overcomplete AE may fail to learn meaningful information from the data [141]. The principle advantages of the autoencoder are the reduction in dimensionality and denoising. This work will apply the autoencoder for dimensionality reduction purposes.

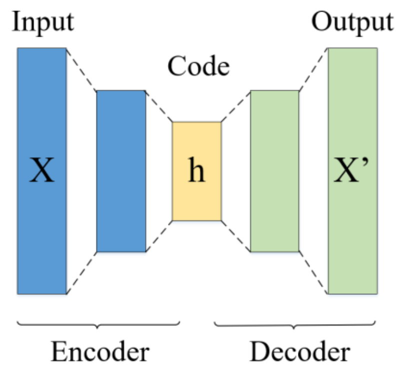

Figure 17: Under-complete autoencoder model

3.3.5 Governing equations for the long short term memory neural networks model

From the literature review, it was concluded that the majority of DL models have been applied to whole building energy loads and the application of such models to HVAC systems still requires further investigation. Due to their ability to learn and model sequential information, the long short term memory (LSTM) neural network models will be applied. The LSTM network is a model first introduced by Hochreiter and Schmidhuber in 1997 [190]. Compared to other types of recurrent neural networks, the LSTM models have the advantage of constant error backpropagation within the memory cell which allows the LSTM model to learn long term dependencies. The LSTM resolve the vanishing gradient problem experienced by other recurrent neural network through the use of gate controllers. Furthermore, the authors specified that LSTM models are able to generalize well, handle distributed presentation of continuous data, and have reduce hyper parameter tuning [190]. The mapping of data, x_t, is accomplished to the output, h_t, by the use of three different gates: the forget gate, input gate, and the output gate. Equations 3-1 to 3-6 provide the governing equations for a LSTM model:

$$f_t = \sigma\,(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3-1}$$

$$i_t = \sigma\,(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3-2}$$

$$\tilde{C}_t = \mathrm{Relu}(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3-3}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3-4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3-5}$$

$$h_t = o_t * \mathrm{Relu}(C_t) \tag{3-6}$$

Where x_t, is the input, f_t corresponds to the forget gate, i_t to the input gate, o_t corresponds to the output gate, and h_t is the hidden state vector (output). In addition, C_t corresponds to the cell state, $\tilde{C}_t$ corresponds to the candidate cell state, W, and b corresponds to parameters of the LSTM cell.

3.4 Inclusion of off-site weather data

For this work, it was desired to explore the performance of using off-site weather data applied to a forecasting model calibrated with on-site data. The objective is to explore the performance effects of: (i) the inclusion of off-site weather data from a publically available source and applied as an input in the event of a failure in the on-site weather station, and (ii) when publically available forecasted weather data is applied as an input.

Both such scenarios are to be applied for the case study involving BAS measurement data from a system level forecasting model. Therefore, based on the case study of the GE building located in Concordia University Loyola campus, the closest weather station with publically available data is the weather station at Montréal-Pierre Elliott Trudeau International Airport located approximately 8 km away. However, local airport data is measured and recorded with hourly time steps, in contrast, the BAS measurement data uses data at 15 minute time steps. Therefore, a methodology was required in order to extract, process, convert and then substitute the airport weather data. The methodology applied for the weather conversion is presented in Table 5.

Table 5: Data conversion methodology

Step Number	Description of Tasks	Definition of Tasks
1	Problem Definition	- Identifying problem - Listing of objectives - Listing of governing equations and assumptions
2	Information Gathering	- Exploration search for data over available data sources - Extracting the data from available data source
3	Preprocessing of data	- Organization and time synchronisation of data - Identification, removal or replacement of missing data, outlier data, and erroneous data
4	Data conversion	- Application of data conversion techniques - Verification of conversion techniques (comparison to on-site data)
5	Substitution of data	- Substitution for local weather sensor data to local airport data

Chapter 4: Component based forecasting method

4.1 Objectives

The overarching scope of this research is the short term forecasting for the electric demand for demand response based programs. The primary objective of this chapter is to present a high performing forecasting model targeting the electric demand of a component within the HVAC system. This forecasting model is developed with sub-hourly data obtained from the buildings BAS, thus, this model leverages currently available data for a component without the need for an additional permanent sensor. The forecast horizon for this work will be up to six hours in advance. The objective of the component model achieves thesis objective number one, listed in section 2.4. Most of the contents of this chapter was published as J. Runge, R. Zmeureanu and M. Le Cam, "Hybrid short-term forecasting of the electric demand of supply fans using machine learning," Journal of Building Engineering, vol. 29, 2020.

4.2 Methods

The forecasting methodology applied for this work is described in section 3.1. This section outlines: the forecasting model, the preprocessing steps applied, the preliminary analysis, the techniques applied for the construction of the forecasting models, techniques for the application of the forecasting models, and the performance metrics applied in this work.

4.2.1 Forecasting method

Forecasting methods typically apply one of two different approaches: a multistep ahead approach or an iterative approach [192, 89, 158]. In multistep ahead forecasting, the forecasting model is developed with multiple outputs, each at a time step over the forecast horizon. For instance, one model may have input regressors at t, to t-n and the outputs of the model would be t+1 to t+k steps ahead. In contrast, the iterative forecasting strategy uses a repetition of single step ahead forecasts with the same forecasting model and continues the repetitions until the forecast horizon has been achieved. For instance, the first forecast estimated by a model is a value for the target variable at t+1 using current and historical values for the regressors. The estimated value at t+1 is then past backwards and used as an input regressor in order to estimate the value at t+2. This iteration repeats until the desired forecast horizon has been achieved. Equations 4-1 to 4-3 provides an example over a three step forecast horizon.

$$\hat{y}(t+1) = f\big(y(t), y(t-1), \dots, y(t-n)\big) \quad 4\text{-}1$$
$$\hat{y}(t+2) = f\big(\hat{y}(t+1), y(t), y(t-1), \dots, y(t-n+1)\big) \quad 4\text{-}2$$
$$\hat{y}(t+3) = f\big(\hat{y}(t+2), \hat{y}(t+1), y(t), y(t-1), \dots, y(t-n+2)\big) \quad 4\text{-}3$$

Where $\hat{y}$ is a forecasted value and y is a measured value, t is the time step, and n is the step number.

In this study, an iterative approach is applied using a single input regressor of the current and previous historical values. Work shown in [32, 188] demonstrated that the target variable is influenced by previous usage and occupants for the building. This work will apply strictly an autoregressive approach and compare the forecasting results to the model developed in references [32, 188]. Furthermore it was shown that between a multistep ahead approach and iterative approach, there is minimal differences in error [158].

4.2.2 Preprocessing of data

The preprocessing of data includes the cleaning and corrective actions of obtained measurement data. This includes: the identification of missing data, outlier values, erroneous data, and the application of corrective actions. The goal of the present research was to obtain a training dataset of no more than 30 days. This would occur at 15-minute time intervals and correspond to 2,880 observations total. Missing data corresponded to “No data” or “0” values occurring within the BAS trend data. Missing data can occur randomly due to faults in the sensors, recording system, and/or periodic maintenance of the HVAC system. All outliers and missing data observations were omitted from the training, validation, and testing data sets based on the following approach: (i) if four or more consecutive observations had missing data, the corresponding rows of observations were removed from the dataset (ii) if the number of missing data was less than four, the values were linearly interpolated from the previous and future time steps. Four consecutive observations corresponds to one-hour of data. In order to compensate for the removal of data observations as a result of preprocessing actions, additional observations were added in order to have a constant overall length of the training dataset.

4.2.3 Preliminary analysis

4.2.3.1 Exploratory analysis

An exploratory analysis is first performed on the target variable in order to help understand the operation of the component and help identify patterns in its operations. It is also beneficial to help view the data itself and see if there are any unusual observations or characteristics. The exploratory analysis consists of: (i) data visualization, and (ii) descriptive statistics. The exploratory analysis helps to identify different operational modes through the use of graphs and plots. Several tools are available such as carpet plots, day plots, and histograms. Carpet plots help provide an overall visualization of the data over time within a compact figure. This is especially effective if viewing lengthy data and helps to identify the operational trends on a component and if they change over time. During the data visualization, all graphs are plotted using the Part Load Ratio (PLR) provided by equation 4-4 [193].

$$\text{Part Load Ratio (PLR)} = \frac{M_s}{M_{design}} \tag{4-4}$$

Where M_s is the air flow rate of the fan and M_{design} is the rated capacity of the fan.

4.2.3.2 Feature selection

Feature selection is the process of selecting relevant features to be applied as input values. Previous work demonstrated that the target variable was highly correlated by historical values and occupants during occupied/unoccupied periods [32, 188]. However, for this work only previous usage was applied as an input regressor in order to compare a multivariate and univariate model. The removal of a regressor may help alleviate model development time and computational time needed to train the forecasting models. Autocorrelation was used to verify that the target variable was highly correlated to the previous usage. The number of previous time lags to be used as input regressors was found through the automated extensive search algorithm which is discussed in section 4.2.4.2.

4.2.3.3 Feature scaling

Prior to constructing the forecasting method, due to large values occurring within the data, normalization of the data is required. If data entering the activation function of the neural network is too large or small, it can drive the output of the neuron to zero or infinity. This in turn can cause significant errors and poor performance. The min-max normalization method was applied to the data, as this is best suited for cases where the bounds are known and not well suited for cases with many and varying outliers [194]. The governing equation for the min-max method is provided in equation 4-5:

$$x'_k = \frac{x_k - x_{min}}{x_{max} - x_{min}} \tag{4-5}$$

Where xk is the specific point in the time series, x_{min} refers to the minimum value within the time series, and x_{max} refers to the maximum value within the time series. This process shifts the time series data to lie within the range of 0 and 1.

4.2.4 Forecasting model construction

4.2.4.1 Forecasting model

The forecast of the target variable (supply air flow rate) is accomplished by using a nonlinear autoregressive neural network. The training of the neural network aims at approximating the unknown function f( ) by using the measurement data to tune the weights within the neural network. The governing equation for a nonlinear autoregressive model is shown in equation 4-6.

$$y(t+1) = f\big(y(t), y(t-1), \ldots, y(t-n)\big) + e_t \qquad \text{4-6}$$

The topology of the ANN is shown in Figure 18.

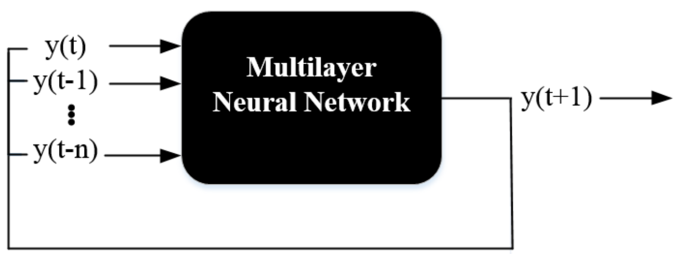

Figure 18: ANN forecasting at time t+1

The general architecture of the ANN is composed of three layers: the input layer, one hidden layer, and the output layer. The input layer corresponds to the time delays of the variable of interest (supply fan air flow rate) at time steps t, t-1,…,t-n. The output layer corresponds to the supply air fan flow rate at future time steps. The ANN is trained in Matlab with measurement data applying the Levenberg-Marquardt algorithm. The algorithm tunes the weights of the ANN to minimize the error between the forecast and the output during training. Training stops for the ANN when one of two criteria have been met: (i) the error falls below the threshold value of 1.0e-7, or (ii) the maximum amount of iterations has been reached. The activation function applied to the hidden layer is a hyperbolic tangent sigmoid function (Equation 4-7), and the output neuron activation function is given by Equation (4-8).

$$\tanh(x) = \frac{\exp(2x) - 1}{\exp(2x) + 1} \qquad \text{4-7}$$

$$\mathrm{id}(x) = x \quad (4\text{-}8)$$

In order to apply the ANN over the forecast horizon, an iterative approach is used. The ANN is provided the input values from y(t), y(t-1) to y(t-n) in order to forecast the value at y(t+1) as illustrated in Figure 18. The forecasted value of y(t+1), is recorded and then fed back as an input to the ANN in order to forecast the next value y(t+2). The ANN keeps a constant number of inputs, thus in forecasting the value at time y(t+2), the input values are: y(t+1), y(t), y(t-1)….y(t-n+1) (Figure 19). This iterative forecasting process continues until the forecast horizon has been achieved.

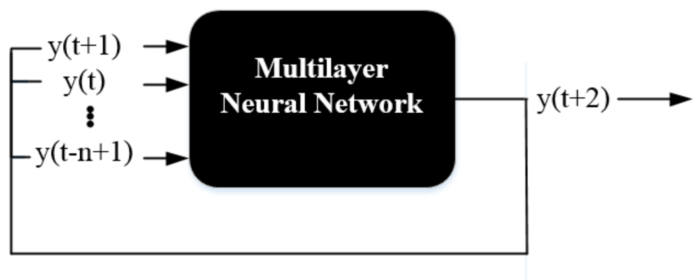

Figure 19: ANN forecasting at time t+2

4.2.4.2 Hyperparameter optimization

The architecture of the ANN is determined by its topological structure [195]. This can constitute a variety of different parameters including: the number of input neurons, number of hidden layers, number of hidden layer neurons, number of output layer neurons, and the transfer function for each node. Given a learning task, too many connections within the ANN model and it may over-fit noise learnt during the training period. Conversely, an ANN with too few connections may not perform well due to its limited learning abilities [195]. Therefore, the selection of an ANN architecture is an essential element to its successful application. Despite the importance, to date, there remains no exact method in order to select the near-optimize architectures. Three main approaches exist to find the optimal ANN architecture: (i) heuristics, (ii) evolutionary algorithms, and (iii) cascade-correlations algorithm.

Heuristics consists of approaches which rely on trial-and-error, or ‘rules of thumb’ based equations. Within the heuristics approach, trial-and-error remains the most popular with examples found in (but not limited to) references [48, 196, 197].

Evolutionary algorithms have been applied to neural networks; however, this has mainly been applied to the optimization of ANN weights rather than ANN architectures as shown in papers [26, 69, 198]. To date, few papers have applied evolutionary algorithms to architecture selection of neural networks. Both genetic algorithm and auto-correlation analysis were applied for finding the near optimal ANN architecture in reference [67].

The cascade-correlation algorithm is another approach for ANN architecture selection [33, 199]. This is an adaptive learning approach which grows the ANN. In the beginning, the ANN consists of a small architecture, typically with a single hidden layer neuron. The ANN is trained, applied to testing data (within the training dataset), and the results are recorded. Another neuron is then added to the hidden layer, and the ANN is re-trained and tested again on the same dataset. If the ANN has a lower error than the previous architecture, the process continues until the error no longer decreases. Disadvantages of this approach are that it can be computationally intensive in the beginning, and the search could stop at a local optima.

The approach to be applied to the component model was based on an automated cascade-correlation algorithm. The optimal ANN architecture is selected through an extensive search which automates the growth of an ANN and explores all combinations of input neurons ranging from 1 to 50 neurons (n_i) and the number of hidden layer neurons from 1 to $n_h = n_i - 1$ [200]. Furthermore, each architecture is trained varying the length of training from 1 to 30 days. Hence, the extensive search is carried out over 36,750 possible architectures. The extensive search algorithm is suitable when computational time is not excessive and input values are discreet numbers [201]. For each architecture and specified length of training data, the training algorithm (Levenberg-Marquardt) and transfer functions were kept constant. Once trained, each specified ANN was tested (within the training dataset) over a six-hour forecasting horizon and the results of the forecast were recorded. The results were recorded using the root mean square error (equation 4-12) between the forecasted values and measurements. The ANN architecture that gives the minimum average RMSE, corresponding to five starting times (01:00, 05:00, 09:00, 13:00 and 17:00) over the day prior to testing is selected as the optimum ANN architecture. By exploring the full solution space, the global minima of RMSE can be found and used as a future reference solution. In contrast to an EV algorithm, which can get trapped in local minima [195].

A drawback of this approach is that it can be computationally intensive; however, this can be reduced by running on several multi-core computers simultaneously, or using cloud computing. With regards to implementation, this method is time intensive in the beginning in order to find and apply an optimum ANN architecture with a sliding window approach (similar to implementation). An additional benefit of this method is that it finds multiple optimum architectures, which is beneficial in creating a homogenous ensemble.

4.2.4.3 Ensemble forecasting approach

A crucial aspect of ensemble forecasting models is the combination of generated forecasts into an overall output forecast. This involves determining appropriate weights to assign to each output forecast or each forecasting model. Currently, several different methods are available, including equal weights, Bayesian, and genetic algorithms. Within the following work, an equal weights approach was implemented as this was demonstrated to provide robust results [202].

4.2.4.4 Sliding window retraining technique

As new data becomes available from the BAS, this can be applied to the ANN for retraining in order to help maintain accuracy through the progression of time. Three main retraining techniques have been applied within literature, static, accumulative and sliding window. With regards to the static technique, the ANN is trained initially and as new data becomes available, the ANN is not retrained. With such a technique, there is a high possibility that the forecasting model becomes invalid and offers poor performance when new patterns begin to emerge.

Accumulative retraining entails the accumulation of data with periodic retraining. When new data becomes available from the control system, it is combined with the initial training dataset. The ANN is then retrained with the larger historical dataset and applied to provide the next forecasts. Thus, for this method the training dataset continually grows larger.

The sliding (receding) window technique holds a constant and fixed length of training data that is shifted in time. As new data becomes available, the newly available data is added to the training dataset. When the new data is added, the oldest data of equal length is removed from the initial dataset in order to keep the overall length constant. The ANN is then retrained periodically with the newest available data to provide its forecasts. A comparison of these three retraining techniques are presented in [33, 203, 204]. Within such case studies, it was concluded that the sliding window technique can provide among the highest performance.

For this work, a sliding window technique was applied. Furthermore, the application of the sliding window technique was explored through two different approaches: batch and iterative. In the case of the batch approach, retraining is applied when a batch of new data becomes available. In the case of the iterative approach, retraining is applied when a single new data point becomes available. In addition to the two approaches, two separate processes for retraining can be applied: (1) the initial weight and biases are randomly selected, and (2) the weights and biases from the previous ANN are used for initialization of the new ANN. Thus, applying both sliding window approaches and the processes, four different scenarios are presented and evaluated within this research:

I. Batch updating with the random initialization of weights and biases;
II. Batch updating starting with the weights and biases of the previous ANN;
III. Iterative updating with the random initialization of weights and biases; and
IV. Iterative updating starting with the weights and biases of the previous ANN.

4.2.5 Forecast for the electric demand of an HVAC component

The forecasted electric demand of the AHU supply fans is calculated using the ANN forecasted supply air flow rate, ($\dot{V}_{for}$), and a physical model Eqs (4-9 to 4-11).

$$E_{AHU} = E_{des} * Y \tag{4-9}$$

$$Y = a_3X^3 + a_2X^2 + a_1X + a_0 \tag{4-10}$$

$$X = \frac{\dot{V}_{for}}{\dot{V}_{des}} \tag{4-11}$$

Where E_{des} is the design electric demand, $\dot{V}_{des}$ is the design supply fan volumetric flow rate, and the coefficients $a_3 = 0.8732$, $a_2 = 0$, $a_1 = 0.1268$, and $a_0 = 0$ of supply fans were found through the analysis of measurements of supply air flow rates from BAS, and of electric current from portables sensors [188].

4.2.6 Performance evaluation

The performance of the forecasting model is presented in terms of the Root Mean Square Error (RMSE) and the Coefficient of Variation of the RMSE or, CV(RMSE) as per ASHRAE standard [139] and the Efficiency Valuation Organization (EVO) [205].

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad \text{4-12}$$

$$\text{CV(RMSE)} = \frac{\text{RMSE}}{\bar{y}} \quad \text{4-13}$$

Where $\hat{y}_i$ is the forecasted value, y_i is the recorded value of the ith observation.

4.3 Case study using real measurements

4.3.1 Case study description

The proposed forecasting model is applied to the Genomic research center (GE) building of Concordia University in Montreal. The building has a total floor area of 5,400 m^2, with four stories and a one level basement. The building has a window to wall ratio of 33%, contains 48 offices, three conference rooms and corridors which account for 53% of the floor area, laboratories with fume hoods account for 30%, and the remaining space occupied by kitchen/lounge and restrooms [187].

Figure 20: Genomic research center [206]

There are two main components of the cooling systems for GE: the primary system feeding multiple buildings, and the secondary/local system supply within GE. The primary system supplies chilled water to several buildings on campus and includes two 900-ton chillers (3,165 kW) with a coefficient of performance of 5.76 at design conditions. Heat is removed from the chillers by two cooling towers with a capacity of 4,750 kW (1,350 tons/each). During the summer months, the operation of the chillers is alternated. Both chillers are operated simultaneously only when the demand cannot be met from a single chiller. The primary system supplies chilled water (6.7^{0}C) to several buildings by two constant speed pumps connected in parallel with each other. The pumps are operated in relation to the chillers, one pump for each chiller operating them both simultaneously. In addition to the supply pumps, two constant speed pumps are used in order to extract heat from the condenser line and send it to the cooling tower. More information regarding the operation of the primary cooling equipment is presented in [207].

The cool water supplied by the primary systems is connected to the cooling coils in Genome buildings' Air Handling Unit (AHU), otherwise known as the secondary/local system (Figure 21).

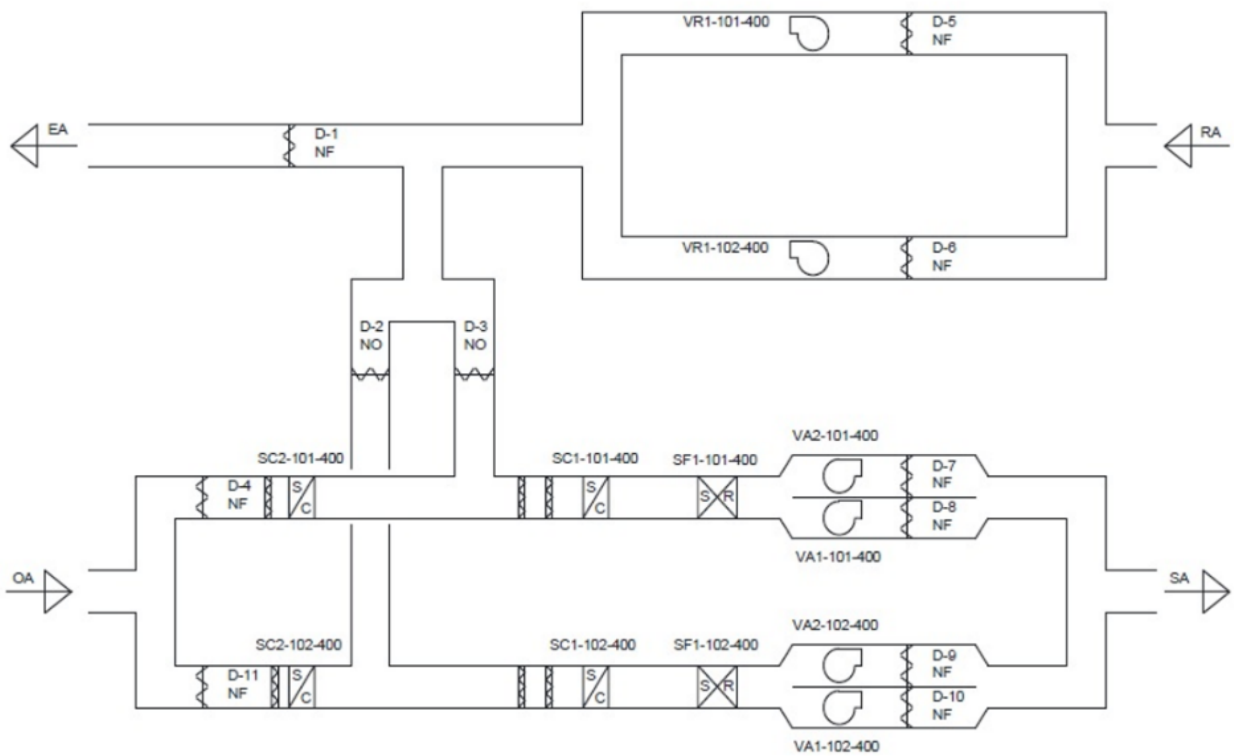

Figure 21: GE secondary system

The secondary system contains four supply fans each with a capacity of 10,618 L/s and an input demand of 29.8 kW. For the following work, the forecasting model will be applied to the total AHU supply air flow rate and electric demand.

4.3.2 Preliminary analysis results

Data for the model was obtained by the building automation system (BAS) of the supply air flow rate recorded at 15-minute intervals from June 1st to August 31st, 2014. The target variable is the total air supply flow rate for the air handling unit and the electric demand of the secondary system AHU. The following section explores the data obtained to help identify trends and patterns with the dataset.

The daily operation of the AHU is presented in Figure 22. The vertical axis corresponds to the day, while the horizontal axis corresponds to the hour of the day. The color of each cell refers to the intensity of the part load ratio. The maximum possible range of the AHU PLR varies from zero to one, the full capacity of the system (42,727 L/s). For Figure 22, a legend is provided in the figure for the PLR, the lowest (blue) corresponds to zero values and the highest (orange) corresponds to highest recorded values (approximately 65% of the total capacity). Figure 22 presents the full dataset, with preprocessing steps applied. As such, this contains missing, erroneous, or outlier data which can appear as zeros (blue). The periodic dark blue lines on the right-hand side within the figure correspond to scheduled maintenance of the AHU and system occurring on Sundays. Figure 22 demonstrates that the AHU has a typical operating range of 0.3 to 0.65 PLR, with increased usage (0.5 to 0.65) occurring from 08:00 to approximately 18:00.

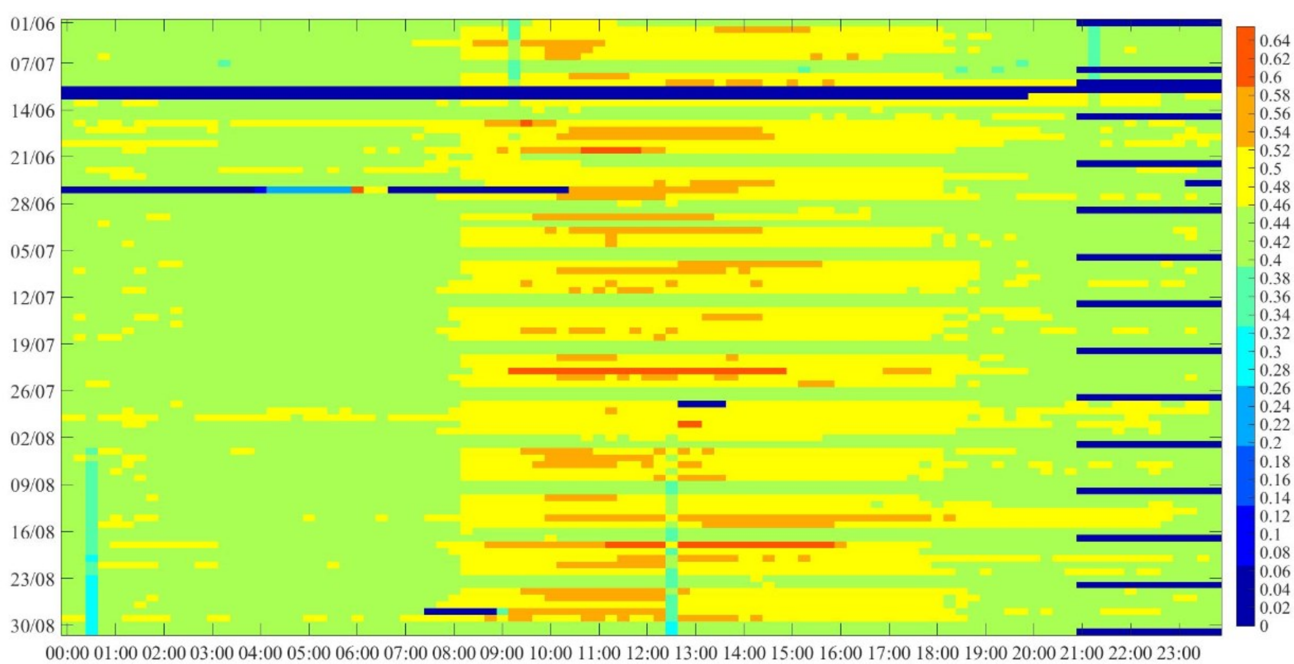

Figure 22: Carpet plot total supply air flow rate

Figure 23 presents the PLR as a day plot. Taking one week of data, 20/07/14 to 26/07/14 or one full week prior to the testing dataset, each day is plotted against the hour of the day. This helps visualize the daily trends occurring within the dataset. The horizontal axis corresponds to the hour of the day and the vertical axis corresponds to the PLR. The daily variations of supply air usage can be seen, with an approximate 0.44 of the total capacity providing the base demand.

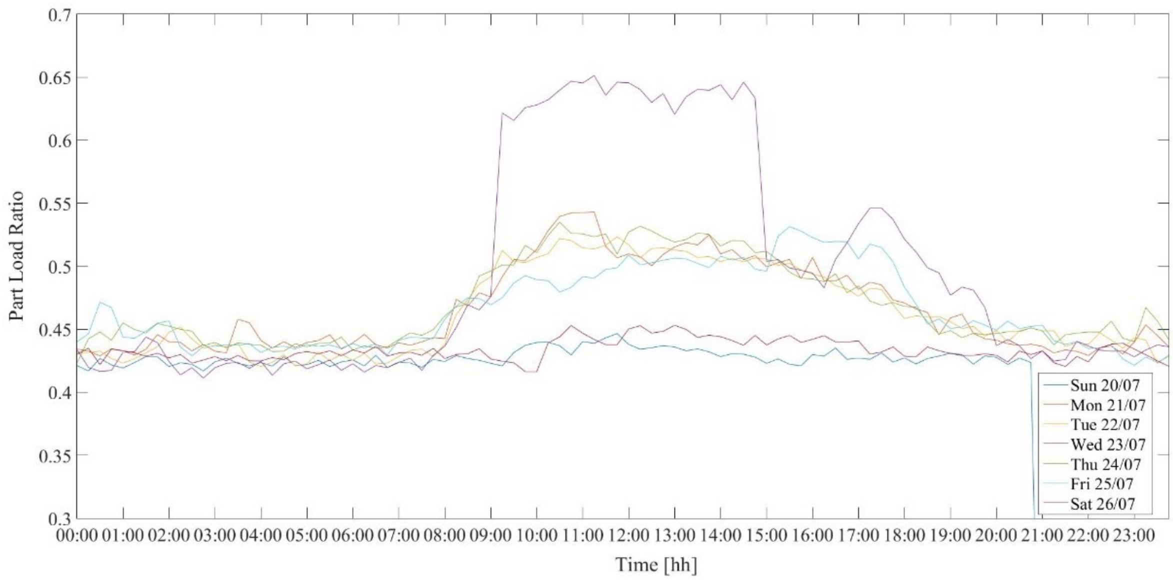

Figure 23: PLR Day Plot

Separating data into occupied (08:00 to 17:45) and unoccupied (00:00 to 07:45 and 18:00 to 23:45) times, the frequency of operational supply occurrences is presented in Figure 24. The vertical axis corresponds to the frequency of occurrence and the horizontal axis corresponds to the PLR. The occupied distribution is presented in the blue color while the unoccupied is presented as orange, the unoccupied contains a larger distribution due to a larger number of samples. The median PLR for the occupied times was 0.49 (20,811 L/s) in contrast to the unoccupied times of 0.44 (18,688 L/s). The most frequent values occur over the range of 0.4 to 0.45% and account for the base supply of the air handling unit. In addition, it can be seen that the maximum demand of the AHU corresponds to 65%.

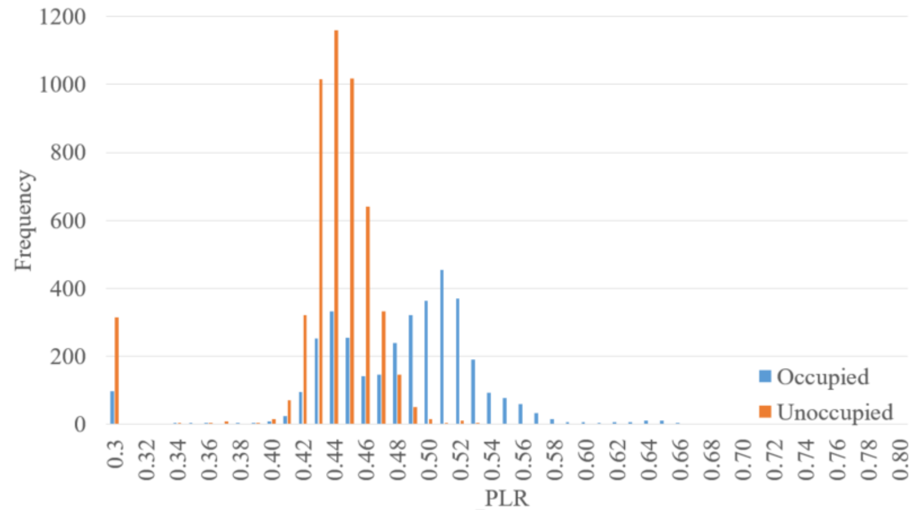

Figure 24: Frequency histogram PLR

4.3.3 Forecasting model construction: Hyperparameter optimization results

As previously mentioned, the architecture of an ANN plays an important role in the performance of the forecasting model. Through the application of the extensive search approach, section 4.2.4.2, 36,750 different possible architectures where explored in order to find the optimal architecture. The optimize architecture is selected (#1 of Table 6), which was observed to have the lowest average of 740 L/s. The selected architecture has 33 inputs consisting of current and past values (over the last 8 hours and 15 mins), a single hidden layer with four neurons, one output layer neuron (33-4-1), and requires 18 days of training data. The RMSE for the selected architecture has the lowest average of 740 L/s, which was calculated by six-hour forecasts within the training dataset.

The top five (#1 to #5) ANN architectures were selected and applied in ensemble-based forecasting. In addition, another ANN architecture that needs only six inputs, one hidden layer with three neurons, one output (6-3-1), and 16 days of training is also selected for comparison (#6 in Table 6). The ANN architecture of (49-47-1) with 17 days of training data provides the worst forecasting performance with the highest RMSE of 24,881 ± 11,188 L/s (Table 6).

Table 6: Ranking of the ANN architectures obtained from the extensive search of the optimum architecture.

Rank	**ANN architecture**	**Training dataset (days)**	**Average RMSE (L/s)**	**Standard deviation of RMSE (L/s)**
1	33-4-1	18	740	175
2	37-4-1	15	746	291
3	29-3-1	29	764	316
4	38-3-1	24	768	372
5	45-4-1	19	775	366
6	6-3-1	16	776	293
7	10-2-1	15	777	211
8	36-8-1	28	788	371
9	50-5-1	24	790	196
10	48-6-1	15	795	398
.	.	.	.	.
.	.	.	.	.
36, 748	50-49-1	30	20,870	6,543
36, 749	50-47-1	15	21,005	5,233
36,750	49-47-1	17	24,881	11,188

Figure 25 presents a small snippet within the overall results obtained from the extensive search. Selecting the architecture with 33 input neurons and a constant amount of training days, Figure 25 presents the error as additional hidden layer neurons are added. With the increasing number of hidden neurons, the results show the increasing forecasting error, expressed by RMSE. The units of RMSE relate to the air supply flow rate (L/s) which contains an overall capacity of 42,472 L/s.

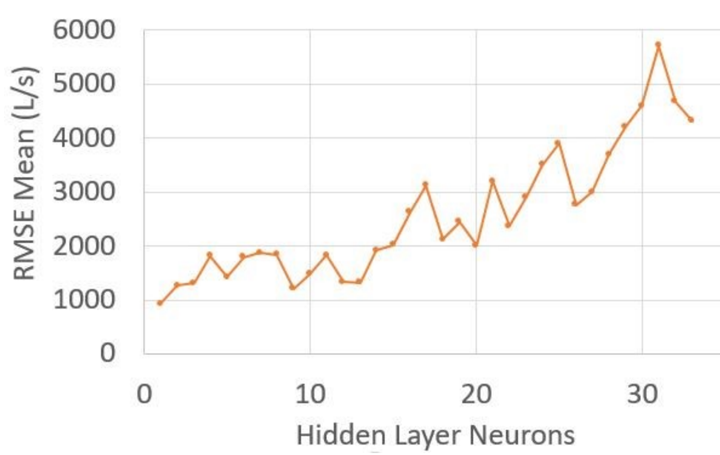

Figure 25: RMSE of the ANN architecture #1 with an increase of hidden layer neurons

4.3.4 Forecasting results

The data is broken into three sections, training, validation and testing datasets. The ANN models (optimized. un-optimized, and ensembles) are trained each with their own required length of training data until July 30th, 2014 at 05:30. The models are then applied to the validation dataset from 05:45 to 08:45 on July 30th, 2014. The testing data begins at 09:00 and forecasts over the next six hours until 14:45.

Table 7: Comparison of the forecasts over the time horizon of six hours starting on 09:00 July 30th, 2014.

Forecasting model	**Supply air flow rate**		**Electric demand**
	RMSE (L/s)	**CV(RMSE) (%)**	**CV(RMSE) (%)**
Optimized ANN #1	386	1.8	5.2
Un-optimized ANN #6	506	2.4	6.9
Optimized SVR	479	2.2	4.8
Un-optimized SVR	724	3.4	7.3
Homogenous ensemble model (ANN #1 to #4)	387	1.8	5.3
Heterogeneous ensemble model (Optimized ANN#1+ optimized SVR)	417	2.0	5.7
Simple Forecasting Approach (SFA)	479	2.2	6.5

For comparison purposes, the forecasting results obtained from two SVR models [32] and a simple forecasting approach (SFA) is also applied. The SFA approach assumes that the supply air flow rate at time (t+1) is equal to the measured value (BAS trend data) from the same time of the previous day [208]. Table 7 presents the results of the forecasting models applied to the testing datasets for both the total air supply flow rate and the electric demand of the AHU.

Various ensemble forecasting models applying an equal weight approach are explored in this case study. The top five (#1 to #5) architectures presented in Table 6 were trained, validated, and then applied to the forecasting dataset. Homogenous ensembles were created by combining forecasts of consecutive architectures into final output forecasts of the target variable. The forecasting performance of the different homogenous ensembles were recorded and is presented in Table 8. The performance results demonstrate a negligible difference between the CV(RMSE) values of the four homogenous ensembles in order to forecast the supply air flow rate. A heterogeneous ensemble consisting of the optimized SVR and optimized ANN were also applied to the testing dataset; the results of the heterogeneous model is presented in Table 8.

Table 8: Comparison of the homogeneous and heterogeneous ensembles over the time horizon of six hours starting on July 30, 2014 at 9:00. Forecasting of supply air flow rate.

	Homogenous ensemble				**Heterogeneous ensemble (ANN and SVR)**
	#1-#2	**#1 -#3**	**#1-#4**	**#1-#5**	
RMSE (L/s)	410	398	387	384	417
CV(RMSE) (%)	1.9	1.9	1.8	1.8	2.0

4.3.5 Forecasting results: Comparison of sliding window retraining technique

This section explored the performance of different scenarios for applying a sliding window technique. The optimized ANN was selected and applied to the testing data based on the four different scenarios presented in section 4.2.4.4. For the purposes of this work, a batch length was set equivalent to the length of the forecast horizon and the testing period is extended from six hours to 12-hours (48 points). The extension of the testing period allows for each scenario to be evaluated over 48 consecutive forecasts (F+1 to F+48).

Table 9 presents the average performance of each scenario over the 48 consecutive forecasts. Scenario (i), batch updating and random weight initialization, has the lowest error with a RMSE of 413 L/s and CV(RMSE) of 2.1%. In addition, scenarios (iii) and (iv) have also low performance errors with a CV(RMSE) value not exceeding 3.7%. Scenario (i) offers the advantage of less computational time as retraining is repeated only after 24-steps; in contrast to retraining iteratively.

Table 9: Comparison of the forecasts of the supply air flow rate by using the sliding window retraining techniques with the forecast horizon of 12 hours over 48 forecasting sets (F+1 to F+48).

	Scenario (i)	**Scenario (ii)**	**Scenario (iii)**	**Scenario (iv)**
Average RMSE (L/s)	413	2,110	576	737
Average CV(RMSE) (%)	2.1	10.7	2.8	3.7

4.3.6 Comparison with other studies

This section compares the performance of the forecasting models applied in this case study, Table 7, with ANN forecasting models in published work and shown in Table 10. All such models presented in Table 10 apply ANN models to forecast various energy loads within a building. The performances presented in Table 10 maybe different in a variety of ways. For instance, the ANN models applied within each published work could be an optimized architecture or un-optimized model. It is beyond the scope of this comparison to discuss each presented paper in detail, as they are presented for comparing a general benchmark for performance.

It is discussed in reference [67] that it is comparatively easier to achieve a good performing forecasting model which targets a whole energy load for a building rather than a model which targets the electric demand of HVAC equipment. This is a consequence of a whole energy load integrating all energy uses into a single profile. In contrast to HVAC profiles which may be more sensitive to changes in operation and/or occupants. The results of Table 10 supports the discussion presented in reference [67].

Table 10: Example of CV(RMSE) values of forecasting using ANN models

Application	CV(RMSE) (%)	Reference
Forecasting of whole building energy performance		
Electricity consumption over a one-hour horizon using hourly benchmark data	2.6-3.1	[54]
Electricity consumption of an institutional building with a horizon of 1 to 6 hours using hourly measurements	7.3-8.5	[84]
Electricity consumption of an office building over the horizon of 24 h using hourly measured data	3.0-16.5	[53]
Residential energy consumption using hourly measurements and a 24-hour horizon	14.3-27.6	[97]
Cooling load over a 24 hr horizon of an institutional building using 30-mn measurements	20.1-25.1	[136]
Electricity consumption of a medical clinic over the horizon of 1 day using 15-mn measurements	7.0-11.1	[58]
Forecasting of performance of HVAC equipment		
Electricity demand of a chiller in an office building over the horizon of 24 h using hourly synthetic data	4.0-40.0	[203]
Electricity demand of a chiller in a research center building over the horizon of 24 h using hourly measurements	23.0-253.0	[203]
Supply fan modulation in an office building over the horizon of six hours using 15-mn measurements	17.6	[67]
Supply fan electric demand in an office building over the horizon of six hours using 15-mn measurements	30.0	[67]

Focusing on the performance of whole building energy loads, when hourly measurements are used the CV(RMSE) performance range is from 2.6-27.6%. Furthermore, when sub-hourly data is used the performance range for whole building energy loads is 7.0-25.1% CV(RMSE). In contrast, when ANN models have been applied to forecasting HVAC equipment the CV(RMSE) results in a larger performance range. For instance, when hourly measurement data is used the ANN models have achieved the performance range of 4.0-253% CV(RMSE) and 17.6-30.0% CV(RMSE) with sub-hourly data.

Therefore, the results obtained in this study which applied sub-hourly data at 15-mn intervals and achieved a performance range of 1.8-3.4% for the supply fan air flow rate and 4.8-7.3% for the electric demand demonstrate an improved performance in comparison to those of other publications.

4.4 Conclusion of the component model

The results shown in section 4.3.4 are obtained from the case study applied. Therefore, they cannot be generalized to all cases and across multiple case studies. The results presented demonstrate that when an optimized ANN or an optimized SVR are applied, they both can obtain good forecasting performance [139]. Overall all models obtained a good performance with performances not exceeding 3.4% CV(RMSE) for the supply air flow rate, and 7% CV(RMSE) for the electric demand of the AHU. When comparing all forecasting models within this case study, both the optimized ANN #1 and the homogenous ensemble model provided the best performance results with a CV(RMSE) of less than 2% for the supply air flow rate and less than 6% CV(RMSE) for the electric demand of the AHU.

It should be noted that due to the similarities between the daily profiles of July 30th and July 29th, the SFA obtained good forecasting results with a CV(RMSE) of 2.2% and 6.5%, respectively. However, when the SFA was applied at 09:00 over 30 days of data, the performance of the SFA obtained a much larger performance range of 1.75% to 24.47% CV(RMSE) solely when forecasting the supply air flow rate. This increased range and error is a result of diverse HVAC patterns in operation. Furthermore, it should be noted that when errors were large, they were typically repeated over multiple days.

In comparison of the ensemble based models, it is observed that the homogenous model obtained a slightly better performance than the heterogeneous model. The major difference between the two models is in the construction time of the forecasting models. The creation time for a homogenous model is less than that of a heterogeneous model. This is a result of the automated hyperparameter approach which can leverage multiple high performing models of a single type (ANN) without the need to then optimize the hyperparameters of a second model (e.g. SVR).

Chapter 5: System-level forecasting method using synthetic data

5.1 Objectives

The overarching scope of this research is a short term forecasting model for demand response based programs. One of the main objectives of this thesis is to present a high performing forecasting model targeting the overall electric demand of a HVAC system. The system based forecasting models are explored on two different case studies. The building with both case studies is the same; however, the difference between the two studies is the source of data. The first case study obtains synthetic data from a calibrated eQuest simulation of the building. The second case study obtains its data from the building BAS system. This chapter explores the system based models applied to the case study using synthetic data.

5.2 A preliminary LSTM forecasting model development

This section presents a preliminary comparison of artificial neural network models (FFNN and LSTM) previously discussed in sections 2.2 and 2.3. In addition to the two previously mentioned single point forecasting models, a forecasting approach consisting of an ensemble of long short term memory neural networks was applied. The forecasting model and approaches in this work are applied to the case study of a Research Center for Structural Genomics in Concordia University. The data source for this study is obtained from a calibrated eQuest simulation of this building completed in reference [187]. Synthetic hourly data from the eQuest simulation was extracted for the components of the HVAC system. These values were then summed, based on equation 1, in order to generate the dataset of the overall HVAC systems electric demand. While a full year of data was obtained from the eQuest's simulation, this work is limited to applying the models during the summer period. The dataset consisting of the total electric load for the HVAC systems during the cooling season was applied to train, validate, and test the forecasting models presented in this case study.

$$\dot{E}^{t+i}_{GE,HVAC} = \dot{E}^{t+i}_{fan,supply} + \dot{E}^{t+i}_{pump,CHWS} + \dot{E}^{t+i}_{Chiller} + \dot{E}^{t+i}_{pump,CDS} + \dot{E}^{t+i}_{CT} \qquad 5\text{-}1$$

5.2.1 Preliminary analysis

A preliminary analysis is conducted on the generated dataset and presented in Figure 26 through a carpet plot. The dataset extracted for this work consisted of two full months of data, starting from June 1st, 2014 at 00:00 and ending on August 1st, 2014. The color of each cell within Figure 26 represents the total electric demand of the HVAC system. From the figure, periods of low demand can be seen in the color blue, and operate over the range of 60 kW to 90 kW during unoccupied times. The high demand times are shown with the color red and occur during occupied periods (08:00 to 19:00) and operate over a range of 100 kW to 170 kW.

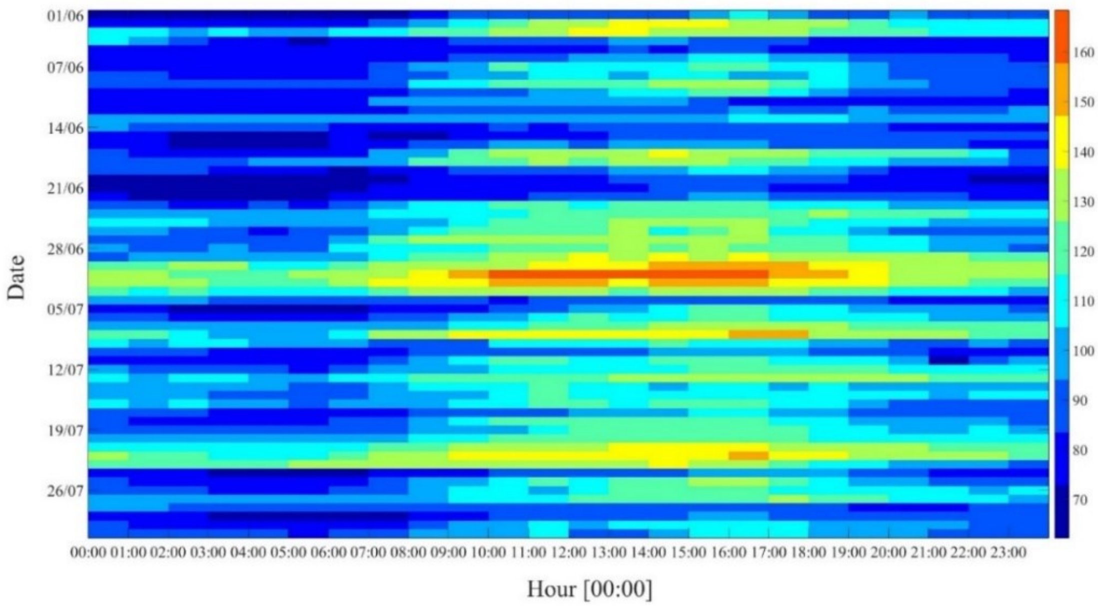

Figure 26: Carpet plot of the Total HVAC's electric demand

5.2.2 Forecasting models

The forecasting models and approaches to be applied within this work consists of:

1. Single point LSTM forecasting model with tuned hyperparameters
2. Single point FFNN forecasting model with tuned hyperparameters
3. Simple forecasting approach
4. Ensemble of LSTM models with randomized hyperparameters based on simple heuristics

5.2.3 Hyperparameter selection

For this work, the hyperparameter searches of the machine learning models were conducted through a grid search technique. The computer used was an Intel Core 2.8 GHz CPU with 8 GB RAM and operating Windows 10 with a 64 bit operating system. All models constructed in this work were built in Python. For this work, H# refers to the hidden layer which will be used in order to describe the number of hidden layer neurons within. Thus, H1 refers to the first hidden layer within the overall neural network architecture. The hyperparameter search was conducted within the training dataset, splitting the data into 90% training and 10% validation (of the training dataset).

In addition, each architecture was evaluated multiple times and the performance results were then averaged. Finally, the computational time for the evaluation of each model was recorded.

For a single layered neural network (LSTM and FFNN), the number of hidden layer neurons was varied from 1 to 10, then 10, 15, 25, 50, 75, 100, 125, 150, 175, and 200. For the single layered LSTM model, it was found that architectures over a range of 75 to 125 units within the hidden layer performed with the least error. Thus, such heuristics will be applied as boundary conditions within the overall LSTM ensemble. After the single point forecasting models were searched over the aforementioned range, a localized search was conducted in order to find the optimal number of neurons within the single hidden layer for both the LSTM and FFNN network models.

Furthermore, a search of deeper LSTM models was conducted with 2 or more hidden layers. With regards to two layered LSTM models, the search was varied from the range of 5, 10, 25, 50, 75, 100, 125, 150, and 175 for H1. For each value of H1, H2 was searched over the range of 5, 10, 15, 20, and 25. The reduction of search range for H2 compared to H1 arose from the result of the increasing computational time in training the models without a significant performance change.

For LSTM models with two or more hidden layers, the range of step sizes was reduced further due to increasing computational time. The range of values consisted of 5, 10, and 15 neurons each. A sample of the hyperparameter search results are provided in Table 11 showing the performance and computational time required to train the models. From Table 11 it can be observed that with the addition of hidden layers, the performance did not drastically change. However, the training time significantly increased with the addition of more hidden layers and deeper architectures. Furthermore, for this work it was observed that increasing the amount of hidden layer neurons did not lead to a significant performance reduction in error. The result appear consistent with those from other fields (such as natural language processing, time series, etc.). For instance, during the M4 time series competition it was observed that increasing the number of hidden layer neurons did not lower the result of performance and rather led to non-convergence of such models in training [209]. Furthermore, Reimers and Gurevych (2017), explored various hyperparameter adjustments for LSTM models [210]. From this, it was found that LSTM models are more sensitive to hyperparameters such as epochs, activation function and the random weight initialization. In addition, it was observed with their work that the number of hidden units are of less importance.

Table 11: Hyperparameter results for LSTM

H1	H2	H3	H4	RMSE (kW)	CV(RMSE) (%)	Training Time (sec)
15	0	0	0	5.75	5.66	198
15	15	0	0	5.68	5.58	309
15	15	15	0	13.72	13.39	481
15	15	15	15	8.57	8.39	860

5.2.4 Ensemble approach

For this work, an ensemble consisting of four LSTM single point forecasting models is applied. The hyperparameter of the models within the overall ensemble consists of randomized values based on the heuristics found during the hyperparameter search. This overall forecasting approach is inspired by a competitor in an ASHRAE Great Energy Shootout competition. Within the competition, the competitor forgone hyperparameter tuning for FFNN and leveraged a large homogenous ensemble model to provide the output prediction [211]. The outcome of the approach was third place within the overall competition, however, where the top two performers spent long periods of time on the development and tuning of their models, this approach spent half a day. However, the work herein differs from the competitor in that a quick automated grid search is applied to learn some heuristics of the model. Such heuristics are then used as boundary conditions to create randomized architectures in a homogeneous ensemble. The single point forecasts within the ensemble are then combined in an equal weights approach.

5.2.5 Retraining approach

For this work, the accumulative window retraining approach was applied as new data became available for the models. With the accumulative approach, the forecasting models are trained on the initial dataset. As new data becomes available, the new data is added to the original dataset and the model is retrained on the new larger dataset. This process continues through time, continually adding to the initial dataset in batches and retraining the models. For this work, the addition of newly acquired data and the retraining of the machine learning models occurs every 12-hours.

5.2.6 Error Indices

This work applies the root mean squared error (RMSE) and the coefficient of variation of the root mean square error CV(RMSE) as the forecasting error indices (eq. 4-12 and 4-13). These error indices were selected as they are recommended by ASHRAE Guideline 14 [139] and are shown in section 4.2.6.

5.2.7 Results and discussions

This section presents the results of the single point and ensemble models applied to forecast the future electric demand for the HVAC with a horizon of six-hours in advance. The models applied historical lags as inputs to the machine learning model with current and past usages from t to t-n hours behind. The full datasets were broken into training, validation, and testing data. The validation set consists of the 12-hours prior to the testing dataset. At each hour over the testing dataset, the forecasts for the next six hours were generated and compared with that of the synthetic eQuest data in order to calculate and record the error of each forecasting model. Retraining occurs for the machine learning models every 12 hours. After each model has been applied to the full testing dataset, the average of each model is then calculated. Table 12 presents the average performance results for each model. For this work, the testing dataset is from 08:00 07/30/2014 to 08:00 07/31/2014, or one full day.

The results can be observed that the LSTM models provided the top two performing models over the testing data. Furthermore, the LSTM ensemble model obtained the lowest forecasting error with 5.73% CV(RMSE) while the optimized single point LSTM model was a close second with and error of 5.77% CV(RMSE).

Table 12: Performance results over testing dataset for the preliminary study

Forecasting Model	**RMSE (kW)**	**CV(RMSE) (%)**
LSTM ensemble	5.58	5.73
LSTM (24-100-6)	5.55	5.77
FFNN (24-50-6)	6.66	6.91
SFA	11.33	11.60

5.2.8 Verification of code development

In order to verify the code developed for this work, a single point LSTM forecasting model was applied to generated datasets. These generated datasets consisted of two daily profiles with a larger and smaller operational time. The first generated dataset will consist of a high/low period over an eight-hour duration, while the second generated dataset will consist of a high/low period over a five-hour duration. Both daily profiles were applied over a total length of two months, similar of the previous work (June 1st to August 1st, 2014). Furthermore, the single point LSTM model was applied in an accumulative window approach, used only historical lags as inputs, and consisted of the same overall LSTM architecture (24-90-6). Similarly, the goal is to forecast up to six-hours ahead.

Figure 27 presents the operational profile for both generated datasets constructed. The longer (eight-hour) set begins an operational high (300 kW) at 10:00 which remains until 18:00 when an operational low (100 kW) begins and continues until the following day. The shorter generated dataset begins its operational high at 11:00 which remains until the low period beginning at 16:00.

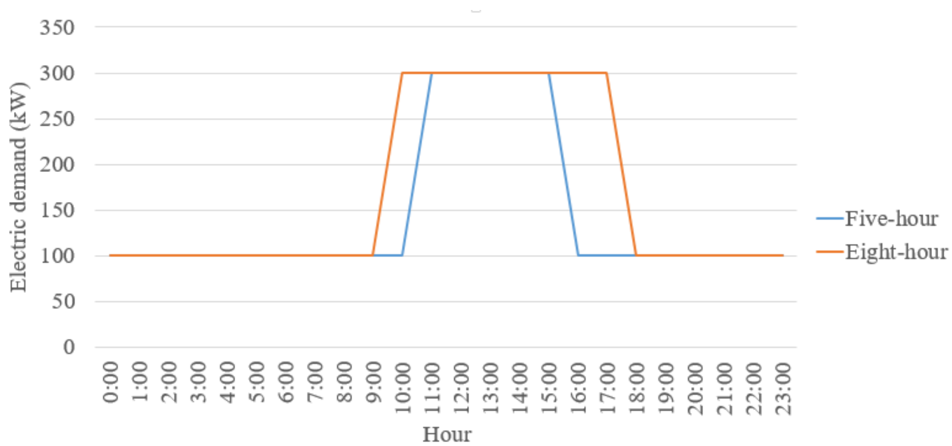

Figure 27: Five and eight-hour generated datasets for verification

Table 13 presents the average forecasting errors as the LSTM model was applied through the testing dataset. It can be seen that the forecasting models provided an average CV(RMSE) of 1.49% when applied to the five-hour operational profile and an average CV(RMSE) of 1.81% with the longer eight-hour operation load profile.

Table 13: Performance Results for generated datasets

	Five-hour duration		**Eight-hour duration**	
	RMSE (kW)	**CV(RMSE) (%)**	**RMSE (kW)**	**CV(RMSE) (%)**
Error	2.30	1.49	2.53	1.81
STD of Error	3.02	1.75	1.82	1.48

5.2.9 Preliminary LSTM furcating model conclusions

The models demonstrate an ability to capture the general shape of the electric demand profiles, however, some phenomena remain difficult to forecast. It was observed that LSTM models are more sensitive to hyperparameters such as epochs and activation functions rather than additional hidden neurons. Furthermore, the additional layers drastically increased the training time required for the LSTM models without a significant performance effect. The models developed and presented in this work showed acceptable performance with a CV(RMSE) of less than 7% for the machine learning models over the forecast horizons when applied to the testing dataset.

From the literature review of artificial neural networks applied to forecast energy in buildings, it has been shown that the majority of applications have been applied to whole building based energy loads [24, 41, 44] . There are few papers which applied ANN models and fewer which have applied DL models for forecasting internal electric loads within buildings. Despite this, the forecasting for the electric demand of the HVAC system remains an area of high interest for the application of demand response based research and programs. Thus, additional research efforts are required [21, 188, 212].

5.3 Methods

Section 5.3 outlines the methods and techniques within the development of the forecasting models applied in the synthetic data case study. The headings of section 5.3 follow the methodology for developing forecasting models presented in section 3.1, however, a more in-depth overview of the specific techniques applied are described in each sub-section.

5.3.1 Information Gathering

This work is based off data obtained from an eQuest model for the building previously published in [187]. For this work, it is assumed that the eQuest model is sufficiently calibrated. The original eQuest file completed was calibrated for 2011. As such, minor modifications were required, however, it was desired to keep the number of modifications to a minimal amount. The first modification applied was the substitution of weather data for 2014. For this, a weather file was obtained from Environment Canada for the Montréal-Pierre Elliott Trudeau International Airport located in Dorval over the year 2014. This airport was selected as it was the closest to the university campus and building and substituted into the simulation. The second modification completed was the alteration of the hourly output report. New variables were required as output measurements from the building and HVAC system based loads. The output variables selected from the eQuest simulation were chosen as they were similar to those of the measured values of the existing building. Upon completion of the modifications, the eQuest simulation was run, and data was collected for the system.

5.3.2 Preprocessing of data

5.3.2.1 Outlier detection, missing and replacement of data approach

As previously stated, this case study is based on data from an eQuest simulation model of a building. Therefore, the overall dataset is whole and without missing data points. Furthermore, all points are calculated and based on physics based equations. As a result, it is assumed that the data obtained from the eQuest file is noise and outlier free. Thus, the data obtained from this case study does not require preprocessing steps which normally would be applied to data obtained from sensor measurements.

5.3.3 Preliminary analysis

5.3.3.1 Exploratory analysis of the target variables

An exploratory analysis is applied to the target variables in order to provide a summary of their main characteristics. The methods applied within the exploratory analysis include: (i) carpet plots of the target variable over the full dataset, (ii) day plots applying a single week of data over a 24 hour axis, (iii) a probability distribution graph for the target variable isolating weekdays and weekends, and (iv) summary statistics for each target variable over the full dataset, weekdays, and weekends.

5.3.3.2 Feature selection

For the synthetic case study, feature selection was primarily based on available eQuest data from the hourly reports. It is desired to be able and compare the system based models on both case studies while keeping similar as much as possible (applied dates, forecasting approach, features/regressors, hyperparameter tuning approach, horizon, etc.). Therefore, the dataset generated for this work was accomplished by comparing, identifying and then selecting similar variables between the eQuest software and the BAS system. However, there were a few minor differences for certain variables. Firstly, the electric demand of the chiller is a single point in eQuest, however, there are multiple loads in the BAS software. Nevertheless, such loads in the measurement case study will be summed into a single overall point. Secondly, there was a difference in the electric demand for the pumps. In eQuest, the electrical demand for the pumps is provided as an hourly output value [kW]. In contrast, in the BAS data, the recorded value for the pumps is based on their state of operation [ON/OFF]. However, such pumps are operated at a constant speed and can easily be converted to a numerical value. A final difference are the regressors which are readily available within the eQuest software, however, within the BAS measurement data they require a physics based equation. For example, the electric demand of the cooling tower is provided as a numerical output in the eQuest simulation [kW]. In contrast, the BAS output provides the measured fan modulation [%] which requires a physics based equation in order to calculate the electric demand.

Based on available outputs and a synchronization of variables with BAS software, regressors were then selected for each target variable based on reference [188]. The feature selection analysis conducted in reference [188], selected the regressors for the same building and was based on a cross correlation analysis and physics based equations.

5.3.4 Forecasting model construction

5.3.4.1 Feature scaling

The first step in the development of the forecasting models is the normalization of the data. For this work, the min-max method was applied in order to normalize both the input and target data. The governing equation for the min-max method is provided in equation 4-5 in section 4.2.3.3.

5.3.4.2 Hyperparameter tuning

The DL model selected for the system level approach consists of a large number of hyperparameters required for tuning. Such hyperparameters include: the number of hidden layers, the number of hidden layer neurons in each layer, the number of epochs, and the amount of training days. Such hyperparameters are needed for both AE and LSTM models. Additionally, there is no standard procedure for the selection of hyperparameters and as such, model development has typically been achieved through heuristic means in previous research. It was desired to explore all possible architectures, however, such a search is computationally expensive and requires a long period of time. Therefore, for the work herein a standard procedure was followed based on a combined grid search technique. The following provides the steps completed in the tuning of the hyperparameters for the EN-LSTM models:

Step 1: Select the number of inputs and consecutively adjust the time lags for all regressors

1.1 Separate full dataset into training, validation, and testing.

1.2 Select the number of lags for each model. The search range conducted for the monolithic model is 5,6,7,8,9,10,11,12,18, and 24 lags. While the search range for the sequential models is 5,6,7,8,9,10,11,12,18,24,30, and 36 lags.

1.3 An additional step is required for the sequential models to couple the forecasted values, t+1..t+24, to the lagged values (t,…t-n) as an overall input structure for each sequential model.

Step 2: Finding the encoder architecture

2.1 Select the training data and separate it into two sections: training (85%) and validation (15%).

2.2 Select the number of hidden layer neurons in hidden layers one and five based on the number of inputs and lags.

2.3 Select the number of neurons in hidden layers two and four. Boundaries of the search range of these layers are from 0.5*H1 to 0.95*H1 with step sizes of five neurons. The boundary conditions are such that hidden layers two are less than hidden layer one (H1>H2) to ensure data compression.

2.4 Select the number of neurons in the third hidden layer. Boundaries of the search are set to 0.5*H2 to 0.95*H2 with step sizes of five neurons.

2.5 Training the AE with the ninety percent data and validate the model with the ten percent. Average the performance of the AE model and record the result.

2.6 Adjust the architecture of the AE through points 2.1 to 2.5 over the boundary conditions.

2.7 Based on the search results for the AE architectures, select the architecture which obtained the lowest error and extract the encoder architecture for each given number given lags.

Step 3: Finding the EN-LSTM architectures

3.1 Select the training data and validation datasets.

3.2 Select the number of input regressors and the EN architecture found for each.

3.3 Select the number of LSTM units. The search range conducted was 75, 100, 125, 150, 175, and 200 units in the LSTM layer and the output layer of the LSTM consisted of a fully connected layer equal to the number of output forecasts.

3.4 Couple the EN to the LSTM, train the model and apply to the validation dataset. Average the forecasting performance of the EN-LSTM single point model and repeat.

3.5 Record the average performance of each architecture.

3.6 Select the EN-LSTM architecture with the lowest error. Adjust the length of the training dataset over the steps of 59, 50, 40, 30, and 20 days. Train the EN-LSTM with each varied length of training data, apply the train models to the validation set (remains unchanged) and record the average forecasting performance of each model with a varied length of training data.

Step 4: Selecting EN-LSTM ensemble hyperparameters

4.1 Average the performance ranges of EN-LSTM hyperparameters

4.2 Select the range of EN-LSTM models which achieves the lowest forecasting error

5.3.4.3 Ensemble forecasting approach

For this work, an ensemble forecasting model is applied to the testing dataset. The ensemble model consists of a homogenous ensemble of the EN-LSTM forecasting model. An encoder compresses the input data which is then applied as inputs to four LSTM models. The output forecasts of each LSTM model are then combined in an equal weights approach.

5.3.4.4 Performance metrices

For this case study, the performance metrics selected include the root mean squared error (RMSE) and the coefficient of variation of the root mean square error CV(RMSE). Such performance metrics are calculated at each time step for the case study over the forecast horizon. The equations for both are presented in equations 4-12 and 4-13 and were selected as they are recommended by ASHRAE Guideline 14 [139].

5.4 Synthetic data case study

5.4.1 Case study description

The data source for this study is obtained from an eQuest simulation of the Genomics building located at Concordia University, Loyola campus in Montreal Canada. The eQuest model for this work was previously completed and is published in reference [187]. Synthetic hourly data is extracted from the simulation for both regressors and target variables, and then used in order to tune and apply the proposed system based forecasting models. The proposed models are applied over the cooling season and the range of dates extracted are from June 1st to August 31st 2014. All data preprocessing, model tuning and model application for the synthetic case study were completed on an Intel Core 2.8 GHz CPU operating Windows 10 on a 64 bit operating system and with 8 GB RAM. Furthermore, all forecasting models were simulated in Python using Keras.

5.4.2 Governing equations

The system level models are applied to forecast the electric demand of the HVAC cooling system during the summer operation. The following details the governing equations for the case study with respect to the data obtained from the eQuest model for the building. The forecast for the electric demand of the cooling system over the forecast horizon is shown in equation 5-2.

$$\dot{E}^{t+i}_{GE,HVAC,ED} = \dot{E}^{t+i}_{secondary,sys} + \dot{E}^{t+i}_{primary,sys} \qquad 5\text{-}2$$

Where t refers to time and i refers to the specific time step. For example, the forecast generated at six hours ahead would be (t+6). The first term in equation 5-2 ($\dot{E}^{t+i}_{secondary,sys}$) refers to the electric demand of the secondary system. The secondary system is defined by ASHRAE as the elements of the HVAC between the central heating/cooling system and those of the terminal or zone units [9]. For this case study, this refers to the electric demand of the AHU supply fans. The second term of equation 5-2 refers to the electric demand of the primary system ($\dot{E}^{t+i}_{primary,sys}$). The primary system is defined by ASHRAE as the part of the HVAC system which consumes energy and delivers the heating/cooling to a building(s) through the secondary system [9]. Equipment of the primary system may include: chillers, boilers, cooling towers, pumps, and other co-generation/thermal storage equipment [9]. For this case study, the secondary and primary system equations are provided in equation 5-3 and 5-4 respectively.

$$\dot{E}^{t+i}_{secondary,sys} = \dot{E}^{t+i}_{fan,supply} \qquad 5\text{-}3$$

$$\dot{E}^{t+i}_{primary,sys} = \dot{E}^{t+i}_{pump,CHWS} + \dot{E}^{t+i}_{Chiller} + \dot{E}^{t+i}_{pump,CDS} + \dot{E}^{t+i}_{Cooling,Tower} \qquad 5\text{-}4$$

Within the eQuest file, the hourly report was adjusted to output all the variables within equations 5-3 and 5-4. The variables were then combined prior to the preliminary analysis in order to generate the data for this case study.

In addition to the primary and secondary system electric demands, the system level sequential approach outputs forecasts for two thermal energy loads within the HVAC system: the air-side cooling load and the water-side cooling load. The equation for such models is provided in equation 5-5 for the air side cooling load and in equation 5-6 for the water-side cooling load. It should be noted, that while eQuest does provide hourly data for each variable in equation 5-5 and 5-6, it additionally provides the cooling load as outputs. Therefore, such datasets were obtained, and the equations were only applied to verify.

$$\dot{Q}^{t+i}_{GE,air\ side} = \dot{V}^{t+i}_{supply,air} * \rho_{air} * 10^{-3} * (h_{mixing,air} - h_{cold,deck}) \quad 5\text{-}5$$

$$\dot{Q}^{t+i}_{GE,water\ side} = \dot{V}^{t+i}_{CHW} * \rho_{water} * 10^{-3} * c_{p,water}(T_{CHW,sup} - T_{CHW,ret}) \quad 5\text{-}6$$

Where $\dot{V}$ refers to the volumetric flow rate (L/s); ρ refers to the density (kg/m^3), h refers to the enthalpy of the mixing air and cold deck (kJ/kg), c_p refers to the specific heat capacity of water (kJ/kg. ^{o}C), and $T_{CHW,sup}$ refers to the chilled water supply temperature (^{o}C) and $T_{CHW,ret}$ refers to the chilled water return temperature (^{o}C).

5.4.3 Preliminary analysis of the target variables

This section presents the preliminary analysis for each target variable for the system based forecasting approaches. The preliminary analysis conducted on each target variable includes (i) an exploratory analysis graphing the variables and identifying trends, and (ii) summary statistics of each variable for weekday, weekend, and full datasets.

5.4.3.1 Preliminary analysis secondary system electric demand

The daily operation of the supply fans for the HVAC system is outputted from the eQuest simulation with hourly time steps. The data is extracted for a full year, however, for this case study the dates from June 1st to August 31st, 2014 are applied. The restriction of dates is for comparison purposes between case studies. The daily operation over the full dataset is presented in Figure 28 were the horizontal axis refers to the hour of the day and the vertical axis refers to the day of the year. The color of each cell corresponds to the supply fans electric power demand, with higher demand shown in red and lower demand shown in blue. Within Figure 28, we can see a ramp up of operation (low to high) typically occurring at approximately 08:00 with a ramp down beginning at approximately 17:00.

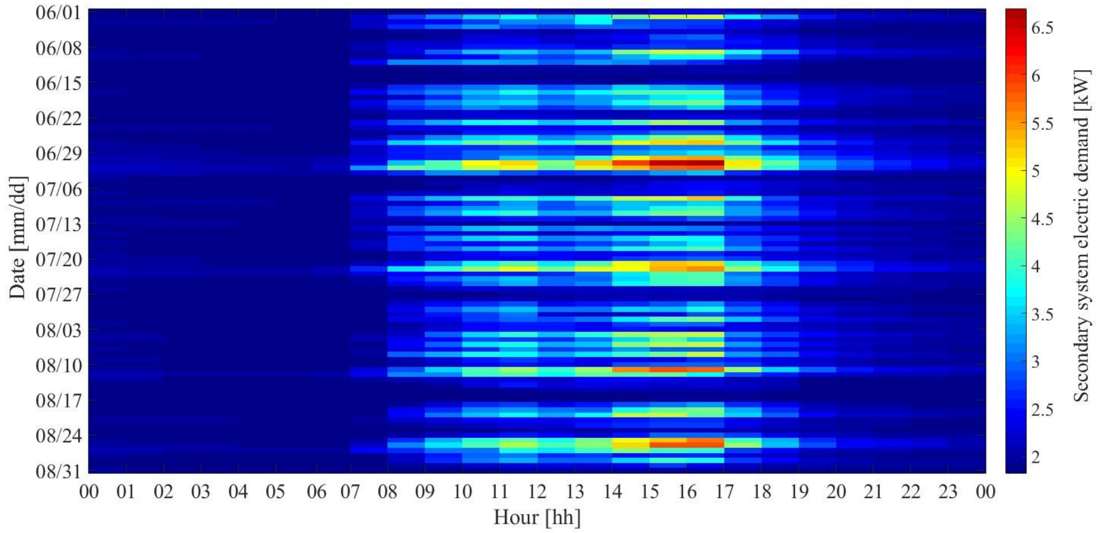

Figure 28: Secondary system electric demand carpet plot for the synthetic case study

Figure 29 presents a day plot for the secondary system electric demand from 21/07/2014 to 27/07/2014 (one week prior to the testing data). Weekdays can be noticed to have larger electric demand loads commencing at approximately 06:00 to 07:00 with peaks occurring at approximately 14:00 to 17:00 followed by a ramp down of operation.

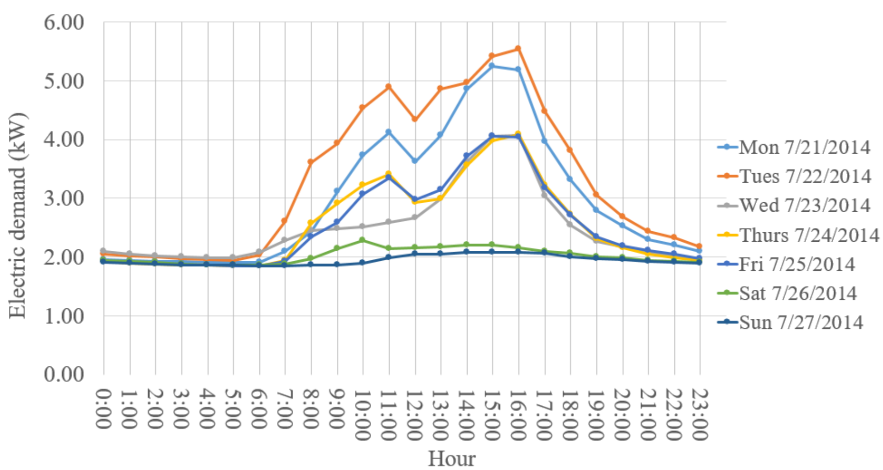

Figure 29: Secondary system electric demand day plot for the synthetic case study

The probability distribution of the electric demand for the secondary system is presented in Figure 30. Within Figure 30 the weekday periods are depicted in black, while the weekend periods are depicted in white. From the figure, it shows that the frequency of occurrences for the electric demand follows an asymmetrical distribution skewed to the right (mean is to the right of the median).

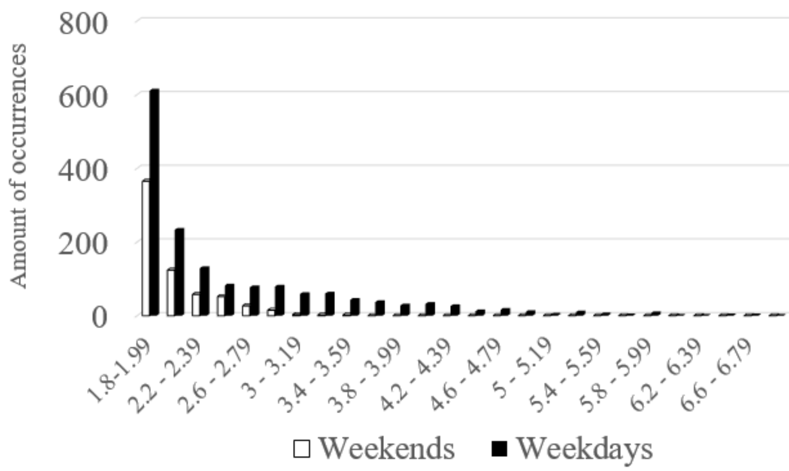

Figure 30: Secondary system electric demand probability distribution for the synthetic case study

Table 14 presents summary statistics for the full, weekday, and weekend datasets. For the weekday and weekend datasets, the data for each day type was found, isolated and extracted from the full dataset and placed into a separate column for analysis. The maximum electric demand for the system during weekday operation is 6.69 kW with a minimum value of 1.83 kW. The weekend shows a maximum of 3.58 kW and a minimum of 1.83 kW.

Table 14: Secondary system electric demand summary statistics for the synthetic case study

Summary Statistics	Full dataset (kW)	Weekends (kW)	Weekdays (kW)
Mean	2.38	2.08	2.51
Quartile 1	1.89	1.87	1.90
Median	2.04	1.95	2.12
Quartile 3	2.59	2.19	2.89
Maximum values	6.69	3.58	6.69
Minimum values	1.83	1.83	1.83
Range	4.86	1.76	4.86
Standard Deviation	0.75	0.30	0.84

5.4.3.2 Preliminary analysis air-side cooling load

The thermal load on the air-side of the HVAC system can be calculated through equation 5-5. The air-side cooling load over the full dataset is presented in Figure 31 through a carpet plot.

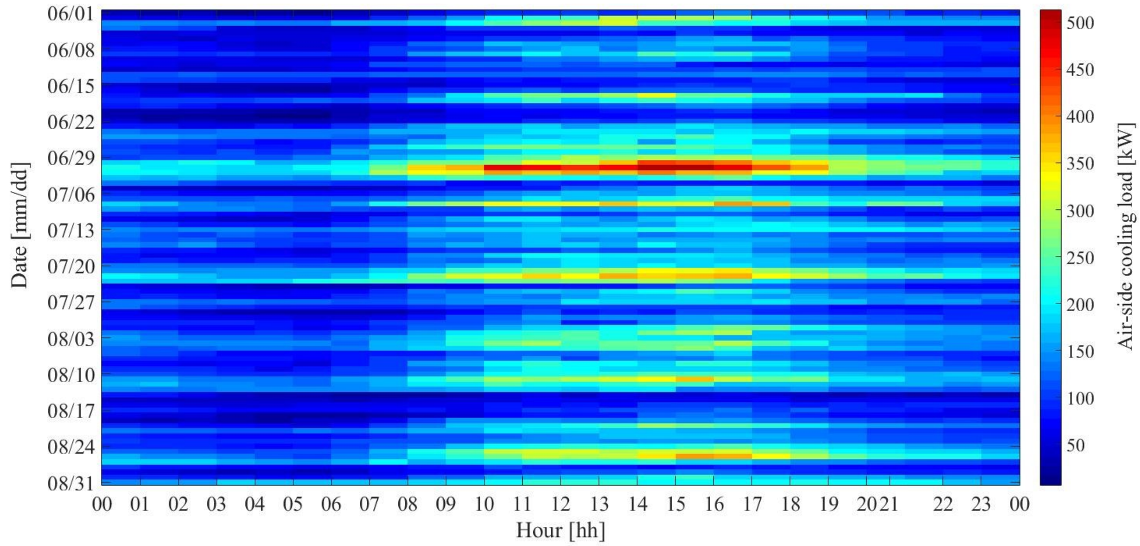

Figure 31: Air-side cooling load carpet plot for the synthetic case study

The blue color squares of Figure 31 correspond to periods when the demand for cooling is low (10 to 100 kW) while the red colors correspond to periods of high demand (400 to 525 kW). The daily operation shows a ramp up occurring from 06:00 to 07:00. The ramp ups begins similarly as the electric demand of the secondary system also begins to increase its demand. Ramp-downs for the air-side cooling load begin decreasing approximately at 16:00 to 17:00.

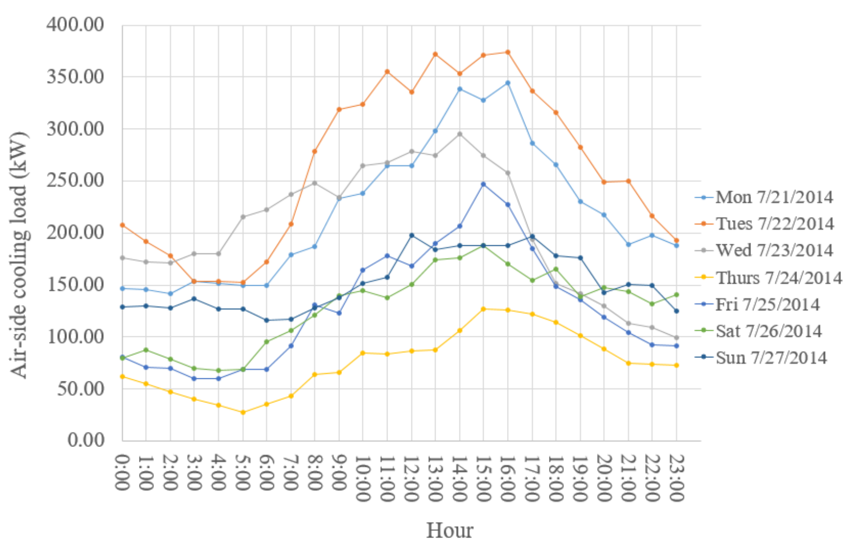

Figure 32: Air-side cooling load day plot for the synthetic case study

The day plot for the air side cooling load is presented in Figure 32 using one week of data prior to the testing dataset (21/07/2014 to 27/07/2014). The figure shows the diverse patterns occurring with Monday through Wednesday containing the largest thermal demands, followed by Thursday with the lowest.

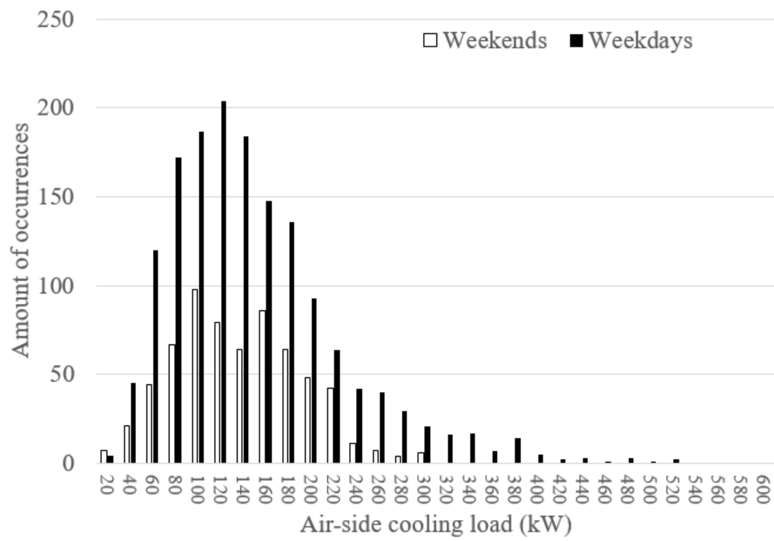

Figure 33: Air-side cooling load probability distribution for the synthetic case study

Figure 33 provides a probability distribution graph for the air-side cooling load. The weekday values are depicted in black, while the weekend values are depicted in white. Bins for the frequency distribution are separated every 20 kW. From Figure 33 the most frequent occurrences are 120 kW for weekdays and 100 kW for weekends. Furthermore, summary statistics for the air-side cooling load are presented in Table 15. The peak demand for weekdays was found to be 512.88 kW and 298.60 kW for weekends.

Table 15: Air-side cooling load summary statistics for the synthetic case study

Summary Statistics	Full dataset (kW)	Weekends (kW)	Weekdays (kW)
Mean	136.21	127.22	139.94
Quartile 1	85.52	86.03	85.12
Median	124.78	124.13	125.15
Quartile 3	174.58	167.48	176.56
Maximum values	512.88	298.60	512.88
Minimum values	6.30	6.30	18.06
Range	506.58	292.30	494.82
Standard Deviation	71.44	55.46	76.80

5.4.3.3 Preliminary analysis water-side cooling load

The water-side cooling load can be calculated based on equation 5-6. The cooling load is outputted from eQuest over the full year and the data is extracted for the summer operation from June 1st until August 31st, 2014. Figure 34 presents the carpet plot for the water-side cooling system. Low demand periods (10 to 100 kW) can be seen in blue, while the high demand periods (400 to 550 kW) can be seen in red. The carpet plot of the water-side cooling load shows an increase beginning from 06:00 to 07:00, as previously shown with the air-side cooling load.

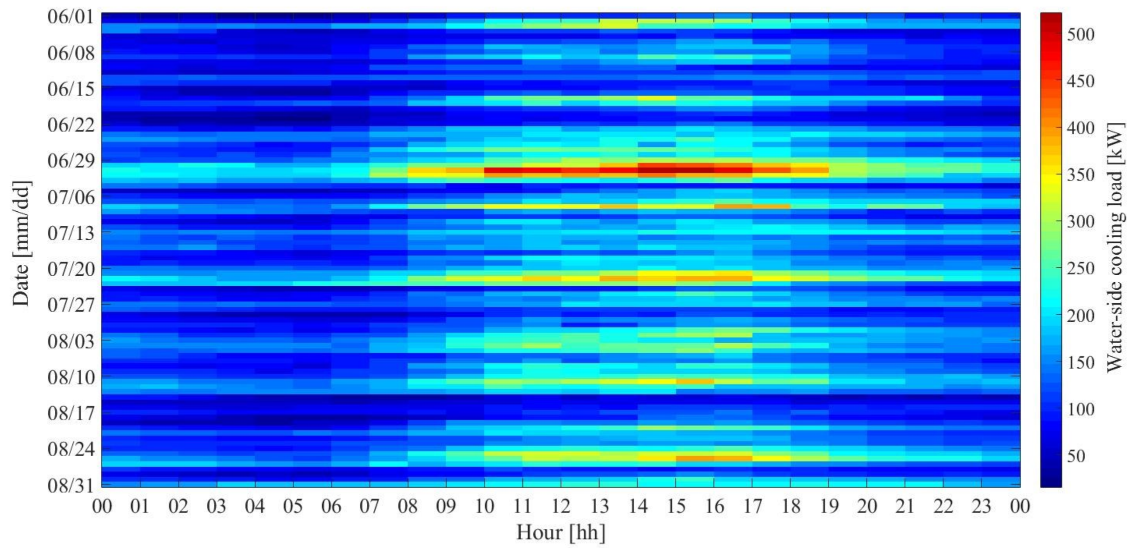

Figure 34: Water-side cooling load carpet plot for the synthetic case study

Figure 35 presents the day plot for the water-side cooling load from 21/07/2014 to 27/07/2014. The daily patterns can be seen to follow closely with that of the air-side cooling load. However, a main difference is that the water-side cooling load is larger than that of the air-side load.

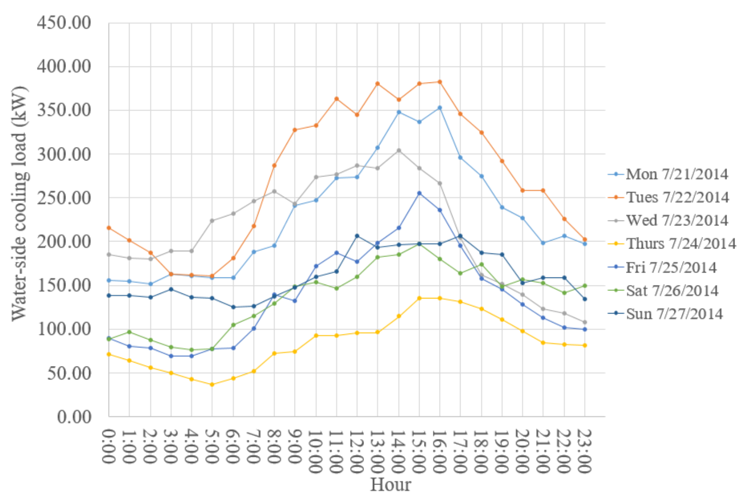

Figure 35: Water-side cooling load day plot for the synthetic case study

The probability distribution for the waters-side cooling load is illustrated in Figure 36 over the full dataset. Days are separated into weekends and weekdays and plotted based on the frequency of occurrences over step sizes of 20 kW. The black bars of Figure 36 present the distribution of weekdays, while the white bars present the distribution during weekends. The figure shows a right skewed distribution with most occurrences of operation at 120 kW for weekdays and 100 kW for weekends.

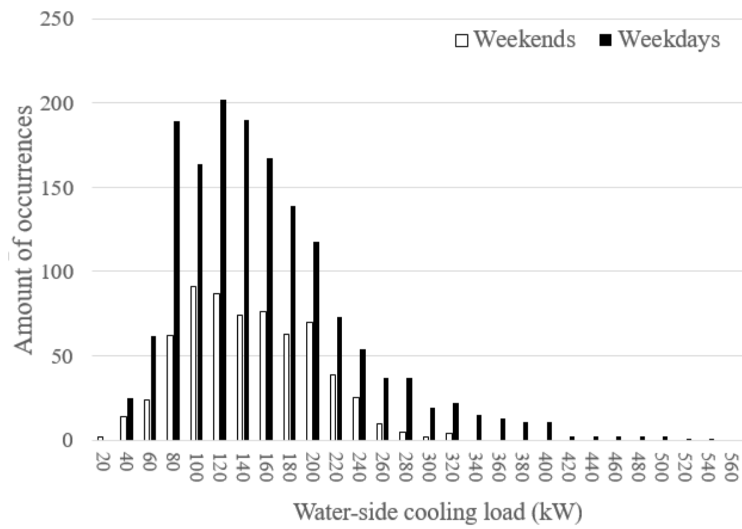

Figure 36: Water-side cooling load probability distribution for the synthetic case study

Table 16 provides the summary statistics for the water-side cooling load over the full, weekday, and weekend data. The maximum demand shows 521.51 kW for weekdays and 307.41 for weekends. Furthermore, weekdays show a minimum demand of 27.40 kW while weekends show a minimum demand of 15.86 kW.

Table 16: Water-side cooling load summary statistics for the synthetic case study

Summary Statistics	Full dataset (kW)	Weekends (kW)	Weekdays (kW)
Mean	145.37	136.37	149.11
Quartile 1	94.63	95.17	94.31
Median	133.84	133.21	134.39
Quartile 3	183.77	176.50	186.21
Maximum values	521.51	307.41	521.51
Minimum values	15.86	15.86	27.40
Range	505.65	291.54	494.11
Standard Deviation	71.40	55.42	76.76

5.4.3.4 Preliminary analysis primary system electric demand

The electric demand for the primary system is shown in equation 5-4. The hourly output electric demand for the chilled water supply pump, chiller, condenser pump, and cooling tower fans are extracted from the eQuest hourly report. These values are then summed (equation 5-4) together in order to create the dataset for the primary systems electric demand over the full year. From this point the data is extracted from June 1st until August 31st, 2014 in order to isolate the data for this case study. Figure 37 provides the carpet plot for the primary systems electric demand over the full dataset to be applied in this study. From the figure, a rise in demand can be seen occurring at approximately 05:00 to 07:00. Furthermore, reductions in demand can be seen occurring at approximately 17:00 depending on the day. The red colored cells of Figure 37 shows the high demand periods with values of (140 to170 kW) and low demand periods with a range of (50 to 80 kW) as shown by the blue colored cells.

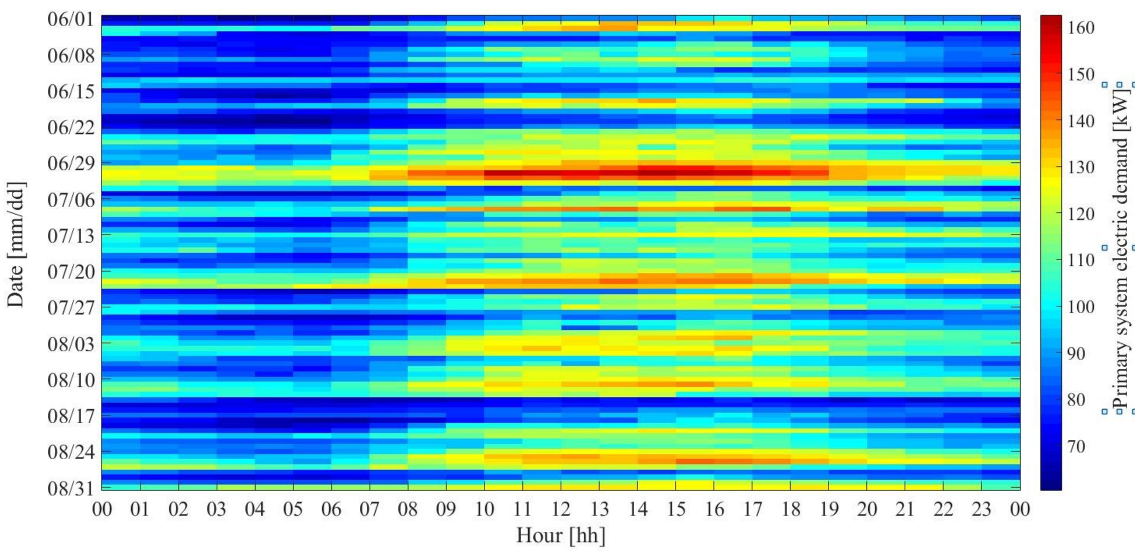

Figure 37: Primary system electric demand carpet plot for the synthetic case study

In order to help visualize the diverse daily patterns, a day plot for the primary systems electric demand is provided in Figure 38 for 21/07/2014 to 27/07/2014. The data for the figure is one week prior to that of the validation and testing data sets. The figure shows a relatively more rounded profile with fewer large variations as that of the previous loads.

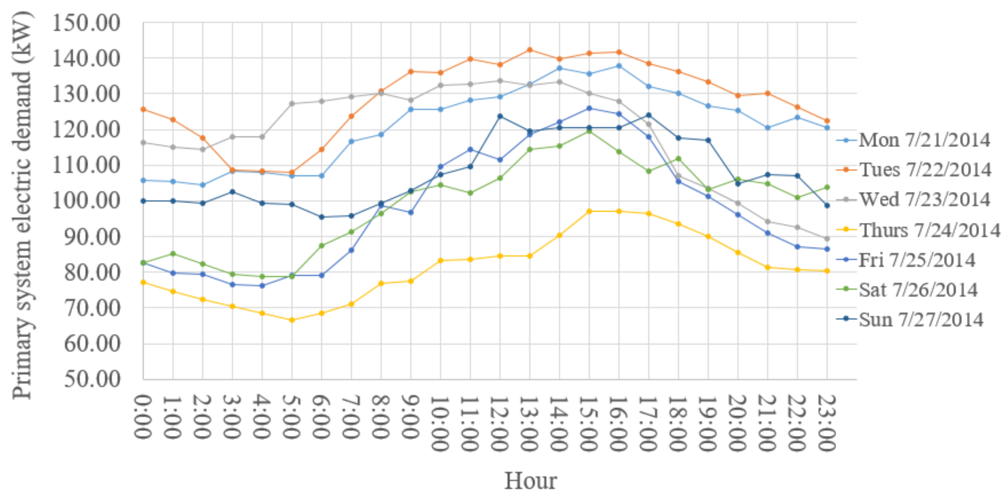

Figure 38: Primary system electric demand day plot for the synthetic case study

The probability distribution for the primary systems electric demand is shown in Figure 39. The full dataset is broken into weekdays, shown in black, and weekends which are depicted in white. The values for the distributions are separated based on bins of 5 kW starting from 60 kW and ending with 170 kW. From the figure, it can be seen that 95 kW have the most occurrences for weekdays, in comparison to weekends with 90 kW.

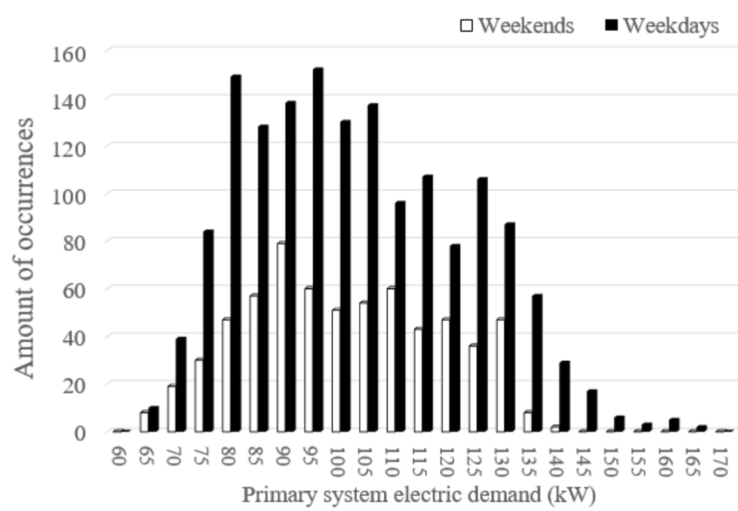

Figure 39: Primary system electric demand probability distribution for the synthetic case study

Table 17 provides the summary statistics for the electric demand of the primary system from June 1st to August 31st, 2014. The maximum peaked value for the weekday is 162.39 kW and 135.47 kW for the weekend. The minimum values for both the weekday and weekend are similar with 63.76 kW and 60.45 kW respectively.

Table 17: Primary system electric demand summary statistics for the synthetic case study

Summary Statistics	Full dataset (kW)	Weekends (kW)	Weekdays (kW)
Mean	99.62	98.33	100.15
Quartile 1	84.51	84.92	84.46
Median	97.57	97.40	97.65
Quartile 3	114.49	112.31	114.88
Maximum values	162.39	135.47	162.39
Minimum values	60.45	60.45	63.76
Range	101.93	75.02	98.62
Standard Deviation	19.01	17.45	19.59

5.4.4 Forecasting model construction: Monolithic approach

5.4.4.1 Selection of regressors: Monolithic model

The selection of the regressors for the monolithic model was selected based on the work completed in reference [188] which selected the regressors based on available data, cross correlation and physics based equations. The work herein diverges from the previous work in that while similar regressors (features) are applied, the time steps for the regressors are different. For example, in the previous work, the target variable of the electric demand of the secondary system was forecasted using the air supply flow rate which varied lags from (t, t-1,…, t-90) at 15 minute time steps. For this work, the selection of time steps for the regressors (air flow) was based on the hyperparameter search conducted. However, the monolithic model consists of one large model exploring the effects of many regressors from the HVAC system applied in a single DL model. For this reason, the regressors applied for the monolithic model consists of all the regressors used in the previous work (selected for each regressor) and incorporated into a single model. The list of regressors applied as inputs for the monolithic model can be seen in Table 18, in addition to the target variables.

Table 18: Selected regressors of the monolithic model for the synthetic case study

Target variable(s)	Regressor (s)
• Secondary system electric demand • Primary system electric demand	• Occupation • Hour • Outdoor dry bulb temperature and enthalpy • Ratio of outside air to total air flow rate • Mixing box dry bulb temperature and specific enthalpy • Cold deck temperature, specific enthalpy, and total air supply flow rate • AHU air-side cooling load • Chilled water supply temperature, return temperature, flow rate and cooling load • Secondary system electric demand • Primary system electric demand

5.4.4.2 Hyperparameter tuning results: Monolithic model

In order to find the optimal hyperparameters, the method described in section 5.3.4.2 was applied. First, a grid search of different AE architectures was accomplished. This grid search varied the number of inputs and set boundary conditions based on the number of inputs. The architectures were trained on a sub-set of the training data (85%) and then tested on the remaining training sub-set (15%). The average performances over the remaining (15%) were recorded. The architecture was then re-randomized with new weights, and the process was repeated multiple times for each given architecture. The architectures which produced the lowest averaged error were selected and are presented in Table 19.

Table 19: Autoencoder results for the monolithic model for the synthetic case study

Number of lags	**H1**	**H2**	**H3**	**H4**	**H5**	**Average CV(RMSE) (%)**
5	95	95	40	95	95	5.28
6	114	95	30	95	114	5.49
7	133	95	45	95	133	6.84
8	152	95	40	95	152	6.85
9	171	95	45	95	171	7.53
10	190	95	45	95	190	8.39
11	209	95	45	95	209	8.42
12	228	90	45	90	228	9.01
18	342	95	45	95	342	10.32
24	456	95	45	95	456	11.19

Next, after identifying the AE architectures which produced the lowest errors for each given number of lags, the encoder architectures (H1 to H3) were extracted. LSTM models were then coupled to the encoder architectures and a grid search for the EN-LSTM model was then conducted. The LSTM model hidden layer units were varied from 75 to 200 in the first hidden layer, while the last most layer of the LSTM contained a fully connected layer. Each given EN-LSTM architecture was trained and applied to the validation data recording the average performance. The average value for each architecture was recorded and is presented in Table 20.

Table 20: EN-LSTM monolithic hyperparameter results for the synthetic case study

Number of lags	Average CV(RMSE) of EN-LSTM single point architectures applied to validation dataset (%)											
	75 LSTM Units		100 LSTM Units		125 LSTM Units		150 LSTM Units		175 LSTM Units		200 LSTM Units	
	SEC	PRI	SEC	PRI	SEC	PRI	SEC	PRI	SEC	PRI	SEC	PRI
5	13.08	10.72	10.44	8.79	12.39	11.22	10.81	10.72	13.51	11.69	10.66	9.67
6	13.63	9.55	9.75	7.90	10.64	9.82	8.98	7.25	10.83	8.60	9.00	8.54
7	7.89	7.60	7.50	6.79	6.86	6.39	8.73	8.19	7.23	8.03	7.55	7.82
8	6.02	5.85	6.24	6.72	6.99	7.79	5.37	7.58	6.04	9.03	5.95	8.56
9	4.75	5.62	3.83	5.25	3.46	6.22	2.85	5.98	3.23	5.52	3.25	5.83
10	4.88	5.61	4.88	5.61	3.50	3.84	3.35	4.14	3.49	5.26	5.01	6.40
11	4.38	7.42	4.08	8.45	3.25	4.62	3.51	4.74	3.02	5.07	2.86	4.13
12	3.58	6.44	3.55	5.40	3.28	4.53	3.41	4.79	4.91	5.17	3.33	5.14
18	15.12	12.26	8.37	6.83	4.56	6.29	12.57	11.80	13.62	13.13	10.55	11.58
24	5.50	6.02	4.36	4.90	4.54	13.17	5.82	10.76	3.16	10.54	5.67	9.84

Next, averages were computed to identify the best performing range of architectures. The range of performances is presented in Table 21. The top performing architectures were identified with 11 input lags (current values and the past ten hours) with an encoder architecture of 209-95-45 applied to LSTM models in a range of 125 to 200 units.

Table 21: EN-LSTM ensemble hyperparameter results for the monolithic model for the synthetic case study

Number of lags	Average CV(RMSE) for monolithic ensemble (%)					
	75 to 200 LSTM units		100 to 200 LSTM units		125 to 200 LSTM units	
	Sec Sys	Prim Sys	Sec Sys	Prim Sys	Sec Sys	Prim Sys
5	11.82	10.47	11.56	10.42	11.84	10.83
6	10.47	8.61	9.84	8.42	9.86	8.55
7	7.63	7.47	7.58	7.44	7.59	7.61
8	6.10	7.59	6.12	7.93	6.09	8.24
9	3.56	5.74	3.32	5.76	3.20	5.89
10	4.18	5.14	4.04	5.05	3.84	4.91
11	3.52	5.74	3.35	5.40	3.16	4.64
12	3.68	5.24	3.69	5.01	3.73	4.91
18	10.80	10.32	9.93	9.93	10.33	10.70
24	4.84	9.21	4.71	9.84	4.79	11.08

In order to find the optimal number of training days for the EN-LSTM models, the top performing architecture was selected from Table 20 and consists of the: 11 lagged, EN 209-95-45, and a 200-12 LSTM. This EN-LSTM architecture was then trained with varied lengths of data from: 20, 30, 40, 50 and 59 days. With each variation of training data length, the models were trained and then applied to the validation dataset and the forecasting performance was recorded. Figure 40 presents the results from the varied lengths of training data and shows that the model obtains the lowest error with the maximum amount of training data.

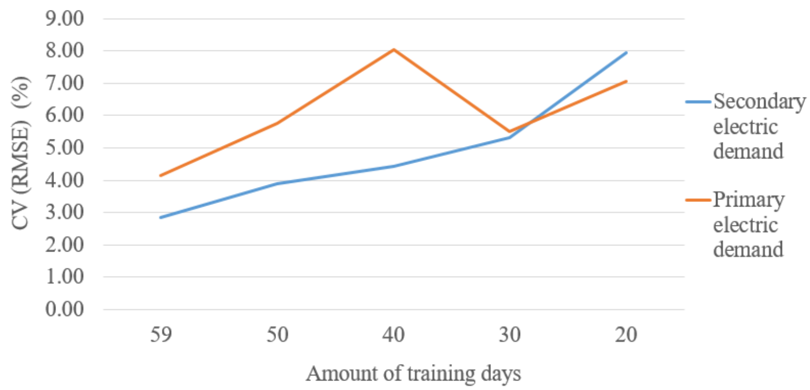

Figure 40: Adjustment of training days for monolithic model for the synthetic case study

5.4.5 Forecasting results: Monolithic model

This section presents the results for the monolithic model forecasting for each target variable applying data at 08:00 July 30th 2014 to forecast 09:00 to 14:00 July 30th, 2014. Thus, this section presents the F+1 forecasts for the monolithic model for both target variables. The architecture of the monolithic model applied was based on the hyperparameter tuning method results and consisted of: (i) the features described in Table 18 (ii) 11 inputs for each feature (t to t-10 hrs.) (iii) an encoder architecture of 209-95-45 hidden neurons (iv) four LSTMs with 125 to 200 hidden layer units, and (v) trained with 59 days of data.

Figure 41 presents the F+1 forecasts (first set of forecasts) and the synthetic data values. The synthetic data is shown by the red lines in both graphs, while the forecasted values are shown by the black lines. The top graph presents the F+1 for the secondary system electric demand while the bottom graph presents the F+1 for the primary system electric demand. The results for the F+1 for the target variable of the secondary system electric demand was 0.28 kW RMSE and 8.83% CV(RMSE) over the forecast horizon. Furthermore, the F+1 performance for the target variable of the primary system electric demand was 7.62 kW RMSE and 7.59% CV(RMSE). Based on the forecasted values of the primary and secondary system, the total HVAC systems electric demand is forecasted by the summation of both systems which yielded an forecasting error of 7.86 kW RMSE and 7.59% CV(RMSE).

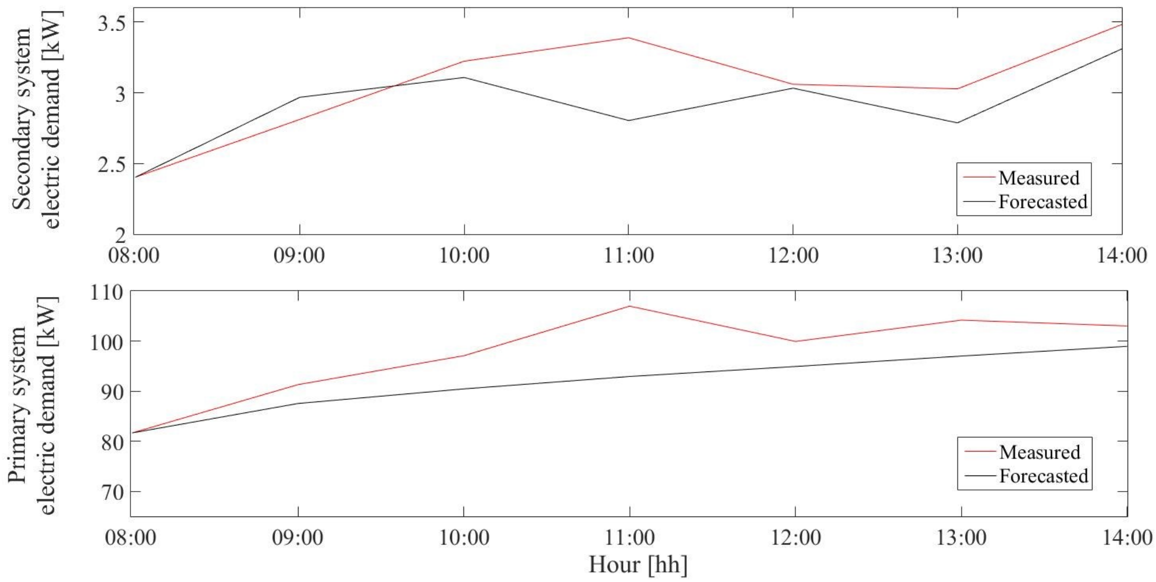

Figure 41: F+1 forecasts for the monolithic model applied to the synthetic case study

5.4.6 Forecasting model construction: Sequential approach

5.4.6.1 Selection of regressors for the target variables of the sequential approach

Similar to the monolithic model, the selection of regressors for the sequential approach is based on the work completed in reference [188]. The sequential model differs from previous work in that forecasts were applied as inputs within this model. Furthermore, forecasting models completed in reference [188] targeted the electric demand of the chillers and an additional model for the electric demand of the cooling towers. For the work herein, such demands are summed along with the electric demand of the pumps into the primary system electric demand.

Table 22 provides an overview for the regressors applied for each forecasting model of the sequential approach. It should be noted, that all the regressors applied current and historical values as inputs into the forecasting models. However, those variables listed in Table 22 marked with an asterisk (*) apply future estimates as inputs in addition to the current and historical values.

Table 22: Selected regressors for the sequential of the synthetic case study

Target variable(s)	Regressor (s)
• Secondary system electric demand	• Occupation • Total air supply flow rate • Secondary system electric demand
• Air-side cooling load	• Outdoor dry bulb temperature and enthalpy • Mixing box dry bulb temperature • AHU air-side cooling load • Total air supply flow rate • Secondary system electric demand *
• Water-side cooling load	• Outdoor dry bulb temperature and enthalpy • Ratio of outside air to total air flow rate • Mixing box dry bulb temperature • Water-side cooling load • Air-side cooling load *
• Primary system electric demand	• Outdoor dry bulb temperature and enthalpy • Chilled water return temperature and flow rate • Chiller and cooling tower electric demand • Primary system electric demand • Water-side cooling load *

5.4.6.2 Secondary system electric demand: Hyperparameter tuning results

The grid search was conducted based on the features, inputs, and boundary conditions for the encoder architectures. The top performing architectures for the AE models of the secondary system were recorded and are presented in Table 23. It should also be noted that for the sequential model, the number of lags was extended up to 36 compared to the monolithic model which stop its search at 24. This was done to explore the EN-LSTM models ability for handling longer sequences of input data.

Table 23: Secondary system electric demand AE model of the synthetic case study

Number of lags	H1	H2	H3	H4	H5	Average CV(RMSE) (%)
5	15	15	5	15	15	155.00
6	18	15	5	15	18	135.36
7	21	20	5	20	21	100.37
8	24	15	5	15	24	91.37
9	27	25	10	25	27	57.57
10	30	30	10	30	30	50.66
11	33	30	10	30	33	49.74
12	36	35	15	35	36	36.41
18	54	50	20	50	54	28.27
24	72	70	30	70	72	24.36
30	90	90	40	90	90	24.68
36	108	95	45	95	108	27.07

Given the results of the AE search, the top performing encoder architectures (H1 to H3) were coupled to LSTM models and a grid search was conducted. Each single point architecture was trained and applied to the validation dataset and the average performance of the model was recorded. The results for the grid search are presented in the left hand side of Table 24. Additionally, the top performing range of ensemble of EN-LSTM architectures was calculated and is shown in the right hand side of Table 24. Based on the grid search approach applied, the EN-LSTM ensemble which obtained the lowest error is identified as nine inputs, with a 27-25-10 encoder architecture and 125 to 200 neurons.

Table 24: EN-LSTM hyperparameter results for the secondary system electric demand model of the synthetic case study

Number of lags	Average CV(RMSE) of single point EN-LSTM architectures (%)						Average CV(RMSE) of ensemble of EN-LSTM architectures (%)		
	75 LSTM Units	**100 LSTM Units**	**125 LSTM Units**	**150 LSTM Units**	**175 LSTM Units**	**200 LSTM Units**	**75 to 200 LSTM units**	**100 to 200 LSTM units**	**125 to 200 LSTM units**
5	8.16	9.24	10.66	10.06	9.65	10.06	9.64	9.93	10.11
6	7.58	7.64	7.76	7.38	7.49	7.63	7.58	7.58	7.57
7	7.86	6.06	5.99	5.39	5.20	5.73	6.04	5.67	5.58
8	3.88	3.78	5.44	3.80	4.12	3.98	4.17	4.22	4.33
9	2.95	2.59	2.27	2.42	2.06	1.76	2.34	2.22	2.13
10	3.46	2.79	2.03	2.68	2.12	2.31	2.56	2.38	2.28
11	3.11	2.58	2.39	2.68	2.45	2.41	2.60	2.50	2.48
12	3.83	3.42	3.04	2.31	3.27	2.93	3.13	2.99	2.89
18	2.85	2.92	3.00	2.15	2.45	1.86	2.54	2.48	2.37
24	2.82	2.79	3.22	3.20	2.91	2.49	2.91	2.92	2.96
30	2.70	2.62	5.22	3.71	2.77	3.32	3.39	3.53	3.76
36	4.69	3.55	3.28	3.63	2.69	4.84	3.78	3.60	3.61

In order to tune the number of training days, a single point EN-LSTM was selected and applied to the validation dataset with varied lengths of training time. The architecture selected consisted of a 9 lagged, 27-25-10 EN, and 200-6 LSTM. The performance of such tuning is presented in Figure 42. Based on the results of Figure 42, the models obtain the lowest error over the validation dataset with the maximum number of training days. Therefore, the architecture to be applied to the testing data set consists of nine lagged, 27-25-10 EN, 125 to 200 hidden unit LSTM, and trained with 59 days of data.

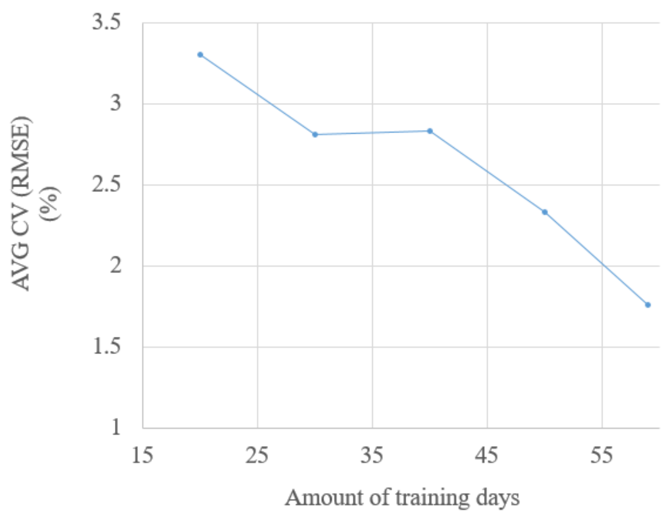

Figure 42: Adjustment of training days for secondary system electric demand model (synthetic case study)

5.4.6.3 Air-side cooling load: Hyperparameter tuning results

The results for the AE grid search for the air-side cooling load forecasting model is presented in Table 25. The regressors used are listed in Table 22 and consists of 6 variables plus six additional time lags (t+1 to t+6) for the secondary system electric demand regressor.

Table 25: Autoencoder results for the air-side cooling load model applied to the synthetic case study

Number of lags	H1	H2	H3	H4	H5	Average CV(RMSE) (%)
5	36	35	15	35	36	7.10
6	42	40	15	40	42	7.20
7	48	45	20	45	48	6.96
8	54	50	20	50	54	7.61
9	60	55	25	55	60	7.36
10	66	55	25	55	66	7.39
11	72	70	30	70	72	7.35
12	78	70	30	70	78	7.48
18	114	90	35	90	114	8.31
24	150	95	45	95	150	8.70
30	186	95	35	95	186	9.56
36	222	95	45	95	222	9.73

Based on the AE grid search, the top performing encoder architectures were selected for each lag (H1 to H3) and coupled with LSTM models. A grid search was accomplished for the EN-LSTM models as they were applied to the validation dataset. The results of the grid search for the single point models is presented in the left side of Table 26 and the calculated averages for the ensemble models is presented in the right side of Table 26. The range which obtained the lowest error is observed to be a five lagged, 36-35-15 EN, and with 100 to 175 LSTM units.

Table 26: EN-LSTM hyperparameter results for the air-side cooling load model for the synthetic case study

Number of lags	Average CV(RMSE) of single point EN-LSTM architectures (%)						Average CV(RMSE) of ensemble of EN-LSTM architectures (%)		
	75 LSTM Units	**100 LSTM Units**	**125 LSTM Units**	**150 LSTM Units**	**175 LSTM Units**	**200 LSTM Units**	**75 to 200 neurons**	**100 to 175 neurons**	**100 to 200 neurons**
5	14.68	15.45	14.44	16.48	14.45	19.37	15.81	15.20	16.04
6	19.92	22.67	18.14	18.65	17.00	19.71	19.35	19.11	19.23
7	19.36	14.97	18.66	16.13	20.89	20.86	18.48	17.66	18.30
8	21.98	20.83	18.91	21.48	20.28	19.66	20.53	20.38	20.23
9	21.92	17.75	19.83	21.74	21.86	26.98	21.68	20.29	21.63
10	22.73	23.56	22.76	24.12	19.29	23.48	22.66	22.43	22.64
11	18.46	20.11	22.60	17.63	19.43	20.97	19.87	19.94	20.15
12	17.68	19.81	16.59	18.27	17.73	16.10	17.70	18.10	17.70
18	22.58	21.40	20.70	19.66	21.19	19.68	20.87	20.74	20.52
24	19.23	22.08	30.44	25.57	22.17	24.80	24.05	25.06	25.01
30	26.16	29.95	31.39	33.54	30.22	32.81	30.68	31.27	31.58
36	30.33	34.77	29.72	28.78	36.95	37.53	33.01	32.56	33.55

The architecture of a five lagged, 36-35-15 EN, and 125-6 LSTM was selected in order to explore the number of training days required for the model. Figure 43 presents the results for the amount of training data and the performance of the selected model when applied to the validation dataset. It can be observed that the model obtains the lowest error with the maximum number of training data. Based on the search conducted for the air-side cooling load, the forecasting model to be applied consists of an EN-LSTM ensemble with five inputs, 36-35-15 encoder architecture, 100 to 175 hidden layer units in the LSTM models, and 59 days of training data.

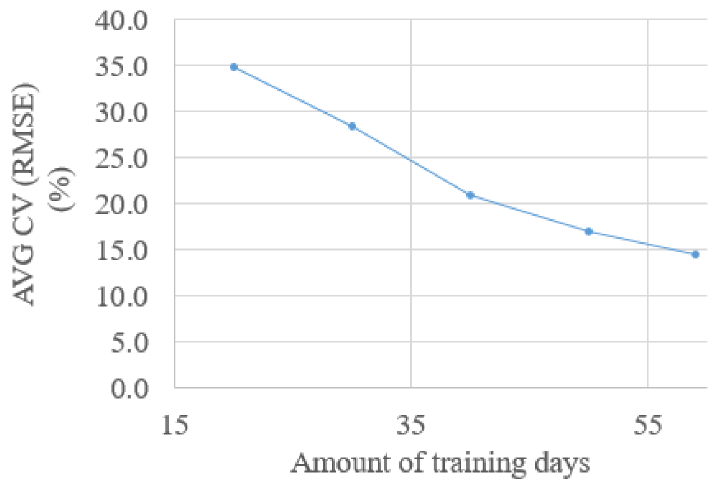

Figure 43: Adjustment of training days for the air-side forecasting model (Synthetic case study)

5.4.6.4 Water-side cooling load: Hyperparameter tuning results

This section presents the results of the hyperparameter tuning of the water-side cooling load model in the sequential forecasting approach. Table 27 provides the results obtained from the grid search of the autoencoder.

Table 27: Autoencoder results for the water-side cooling load model of the synthetic case study

Number of lags	H1	H2	H3	H4	H5	Average CV(RMSE) (%)
5	36	35	15	35	36	3.65
6	42	40	15	40	42	3.61
7	48	45	20	45	48	3.67
8	54	50	15	50	54	3.98
9	60	60	20	60	60	3.87
10	66	65	30	65	66	3.97
11	72	70	30	70	72	3.95
12	78	75	35	75	78	4.03
18	114	90	30	90	114	4.73
24	150	95	45	95	150	5.15
30	186	95	45	95	186	5.57
36	222	95	45	95	222	6.04

After the grid search for the autoencoder, a grid search of the EN-LSTM single point forecasting models targeting the water-side cooling load was then accomplished. The results of the combined models is presented in Table 28. The single point architecture found to have the lowest error based on the grid search conducted consists of a five lagged, 36-35-15 encoder architecture, and with 200-6 LSTM (8.11% AVG CV(RMSE)). The ensemble which obtained the lowest error based on the grid search conducted was found to be five lags, 36-35-15 encoder, and 100 to 200 units in the LSTM models (9.84% CV(RMSE)).

Table 28: EN-LSTM hyperparameter results of the water-side cooling load model for the synthetic case study

Number of lags	Average CV(RMSE) of single point EN-LSTM architectures (%)						Average CV(RMSE) of ensemble of EN-LSTM architectures (%)		
	75 LSTM Units	100 LSTM Units	125 LSTM Units	150 LSTM Units	175 LSTM Units	200 LSTM Units	75 to 200 neurons	100 to 175 neurons	100 to 200 neurons
5	12.07	10.66	10.24	8.79	11.41	8.11	10.21	10.28	9.84
6	9.85	12.65	11.08	14.70	10.48	13.62	12.07	12.23	12.51
7	14.03	11.28	10.61	13.91	9.45	9.49	11.46	11.31	10.95
8	10.40	12.69	11.04	10.92	10.43	9.06	10.76	11.27	10.83
9	17.32	15.68	12.45	13.17	12.19	12.26	13.84	13.37	13.15
10	13.12	13.88	12.35	14.73	11.24	13.74	13.18	13.05	13.19
11	14.01	13.56	10.44	11.19	11.32	11.39	11.99	11.63	11.58
12	11.50	13.21	13.16	11.90	14.30	11.24	12.55	13.14	12.76
18	19.72	15.47	19.12	17.55	13.11	13.99	16.49	16.31	15.85
24	17.20	16.38	15.17	11.86	17.03	16.75	15.73	15.11	15.44
30	20.99	14.99	17.78	13.93	21.85	12.65	17.03	17.14	16.24
36	17.72	16.97	16.65	14.94	17.07	15.30	16.44	16.41	16.19

The single point forecasting model with the lowest error (five inputs, 35-15 encoder, and 200 LSTM units) was selected to explore the effects of adjusting the amount of training days. The results of adjusting the lengths of training data are presented in Figure 44 and show the lowest error with the largest length of training data.

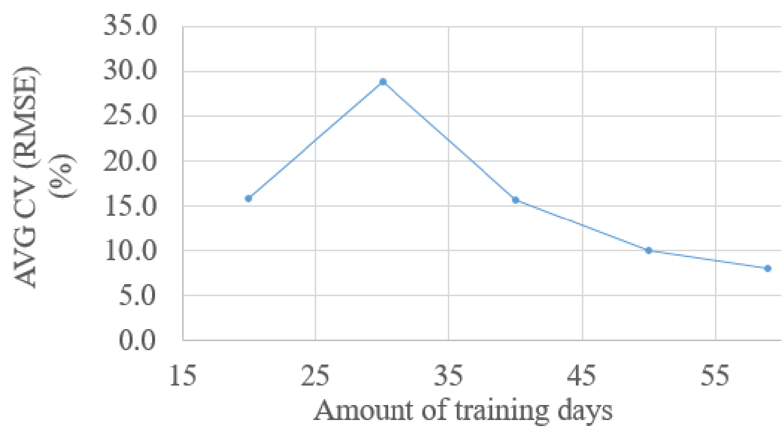

Figure 44: Adjustment of training days for the water-side cooling load forecasting model (synthetic case study)

Based on the search conducted, the architecture of five inputs, 36-35-15 EN, 100 to 200 LSTM units, and 59 days of training data is selected as the water-side forecasting model hyperparameters which will be applied to the testing dataset.

5.4.6.5 Primary system: Hyperparameter tuning results

The primary system electric demand contains the load profiles for the chiller, cooling tower and water pumps of the HVAC system. This section presents the results found from the hyperparameter optimization stage of the primary system model within the sequential forecasting approach. Firstly, a grid search was conducted for the AE model based on regressors presented in Table 22. The results of the grid search are presented in Table 29 which provide the top performing AE architecture for each number of lags.

Table 29: AE results for the primary system electric demand model of the synthetic case study

Number of lags	H1	H2	H3	H4	H5	Average CV(RMSE) (%)
5	46	45	20	45	46	5.46
6	54	50	15	50	54	5.91
7	62	60	20	60	62	5.63
8	70	70	30	70	70	5.36
9	78	75	35	75	78	5.61
10	86	80	35	80	86	5.48
11	94	90	35	90	94	5.72
12	102	95	45	95	102	5.58
18	150	95	35	95	150	7.37
24	198	95	40	95	198	7.94
30	246	95	35	95	246	8.76
36	294	90	35	90	294	8.92

The results of the autoencoder search were then applied to a grid search for the EN-LSTM models. Table 30 provides the performance results for the single point EN-LSTM models (left hand side) and ensemble based models (right hand side). The EN-LSTM single point forecasting model which obtained the lowest error (2.13% CV(RMSE)) consists of the five input, 46-45-14 encoder and 200-6 LSTM model. Furthermore, the EN-LSTM model which obtained the lowest calculated error is the five lagged, 46-45-20 encoder, and with 100 to 200 LSTM units (2.40% CV(RMSE)).

Table 30: EN-LSTM hyperparameter results for the primary system electric demand model of the synthetic case study

Number of lags	**Average CV(RMSE) of single point EN-LSTM architectures (%)**						**Average CV(RMSE) of ensemble of EN-LSTM architectures (%)**		
	75 LSTM Units	**100 LSTM Units**	**125 LSTM Units**	**150 LSTM Units**	**175 LSTM Units**	**200 LSTM Units**	**75 to 200 neurons**	**100 to 175 neurons**	**100 to 200 neurons**
5	2.55	2.48	2.35	2.49	2.54	2.13	2.42	2.46	2.40
6	3.77	4.83	4.06	3.58	3.71	3.91	3.98	4.04	4.02
7	2.94	3.12	4.06	2.98	3.31	3.06	3.24	3.37	3.31
8	3.20	3.32	3.22	3.08	3.38	3.23	3.24	3.25	3.25
9	4.18	4.07	5.13	4.04	3.63	3.48	4.09	4.22	4.07
10	4.03	3.34	3.73	3.63	3.34	3.78	3.64	3.51	3.56
11	4.86	4.18	4.16	4.10	4.18	3.73	4.20	4.15	4.07
12	6.20	6.35	4.22	6.31	5.11	5.20	5.57	5.50	5.44
18	5.39	4.77	5.50	4.67	4.90	4.86	5.01	4.96	4.94
24	6.74	5.07	5.56	4.56	5.31	6.72	5.66	5.13	5.44
30	4.79	5.25	4.96	4.71	5.48	5.74	5.15	5.10	5.23
36	5.51	4.92	4.24	5.36	6.80	6.17	5.50	5.33	5.50

Selecting the five lagged, 46-45-14 encoder and 200-6 LSTM model, the performance of the model was explored with varied lengths of training data and then applied to the validation dataset. The average performance of the model was recorded and the results are presented in Figure 45. The results demonstrate that the model performs with the lowest error (2.13% CV(RMSE)) with the longest length of training data. Based on these results, the hyperparameters which will be applied to the testing data set consists of a five lagged, a 46-45-20 encoder, 100 to 200 LSTM units in the hidden layer and with a length of 59 days of training data.

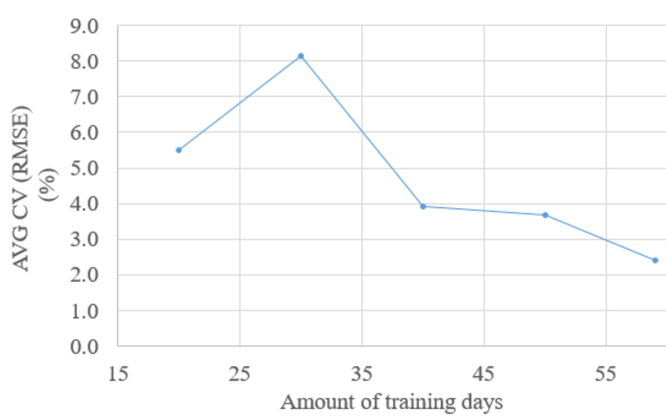

Figure 45: Tuning of the training days for the primary system electric demand model for the synthetic study

5.4.7 Forecasting results: Sequential model

This section presents the results for each target variable of the sequential model as it is applied to generate the first forecasting set (F+1). Inputs were applied from 08:00 July 30th 2014 in order to generate forecasts for 09:00 to 14:00 July 30th 2014. The target variables are forecasted sequentially, one after each other passing the forecasted values to be used as inputs to the subsequent model. The F+1 forecasts for each target variable are shown in Figure 46. With the figure, the synthetic values are shown with the red lines of each graph and the forecasted values are shown with the black lines.

The top most graph of Figure 46 presents the F+1 for the target variable of the secondary systems electric demand. From the figure, it can be seen that the forecasts for the secondary system electric demand follow closely with the profile and show acceptable performance with an error of 0.22 kW RMSE and 6.83% CV(RMSE).

The second target variable forecasted is the air-side cooling load. Based on current and historical data for the regressors and the forecasts of the secondary system, the forecast generated for the air-side cooling load is presented in the second graph (from the top) of Figure 46. The error of this model over the forecast horizon is 23.73 kW and 17.49% CV(RMSE).

Based on the forecasts for the previous model, the third model targets forecasting the water-side cooling load. The output F+1 forecast for the water-side cooling load is shown in Figure 46 (second from the bottom). The error obtained for this forecasting model over the horizon was 27.13 kW RMSE and 18.76% CV(RMSE).

The last DL model forecast within the sequential approach targets forecasting the primary system cooling load. Based on the regressors and the forecasts from the third model, the forecasts for the primary systems electric demand are generated. The first forecasting set (F+1) for the primary system electric demand is shown in the bottom most graph of Figure 46. The error for this model is 14.19 kW RMSE and 14.14% CV(RMSE).

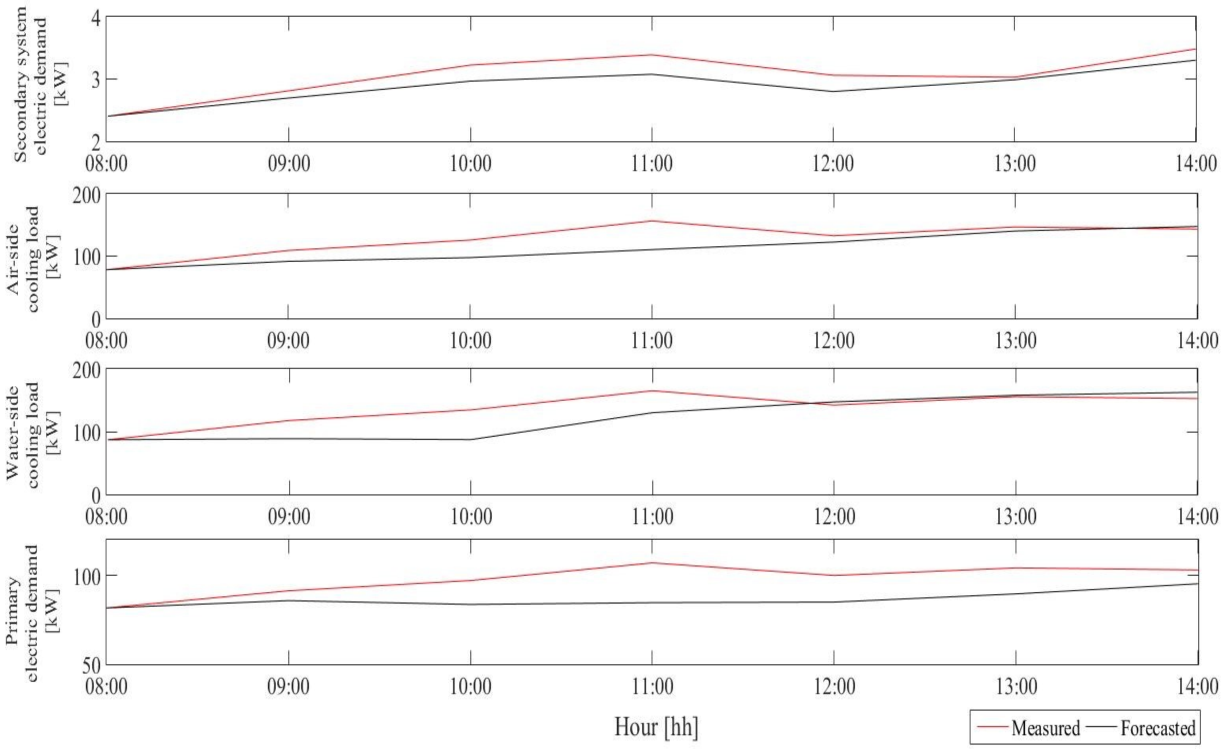

Figure 46: F+1 forecasts for the sequential model of the synthetic case study

The final forecast of the sequential forecasting model is for the total HVAC systems electric demand. This is achieved by the addition of the forecasts for the primary and secondary systems. Based on the F+1 forecasts for both HVAC systems, the forecasts generated by the sequential approach obtained an error of 14.39 kW RMSE and 13.90% CV(RMSE) for the F+1 forecast of the total HVAC systems electric demand.

5.4.8 Comparison of forecasting results approaches

This section compares the results for both system approaches applied to forecast the target variables. For this work, the testing data set consists of one day of consecutive forecasts starting at 08:00 7/30/2014 and continuing until 08:00 7/31/2014. The target variables for both system based forecasting models are calculated each hour over the forecast horizon. For the monolithic approach, the DL model forecasts based on current and past values of the regressors. For the sequential approach, the DL models forecasts based on historical usage, current values, and forecasts generated by the upstream model. The hyperparameters applied are based on search results conducted and are shown in Table 31.

Table 31: Summary of hyperparameters for the synthetic data case study

System approach	Target variables	Lags [h]	AE Architecture	LSTM units
Monolithic	Electric demand of the secondary system Electric demand of the primary system	11	209-95-45	125 to 200
Sequential	Electric demand of the secondary system	9	27-25-10	125 to 200
	Air-side cooling load	5	36-35-15	100 to 175
	Water-side cooling load	5	36-35-15	100 to 200
	Electric demand of the primary system	5	46-45-20	100 to 200

At each hour, the system based forecasting models generate their forecasts for the target variables. The forecasting error is then calculated between the forecasted values and the synthetic data from the case study over the forecast horizon. The performance of both system based models is then recorded hour by hour as the models progress through the testing dataset. Upon completion, the error metrics at each hour are averaged. The error results for each system based model and their target variables over the dataset are presented in Table 32.

Table 32: Forecasting results for the synthetic data case study over the testing data set

Target Variable	**Monolithic approach**		**Sequential approach**	
	RMSE (kW)	**CV(RMSE) (%)**	**RMSE (kW)**	**CV(RMSE) (%)**
Electric demand of the secondary system	0.13	4.94	0.13	5.07
Air-side cooling load	N/A	N/A	15.57	13.79
Water-side cooling load	N/A	N/A	15.34	12.54
Electric demand of the primary system	5.11	5.41	7.26	7.24
Electric demand of the total HVAC system	5.25	5.43	6.09	6.21

The results demonstrate that both system based approaches have minor differences in forecasting the electric demand of the HVAC system. The thermal cooling loads present a larger forecasting error compared to that of the electric loads. It is difficult to gauge the performance of the models as there is no set standard benchmark. Therefore for the work herein, we apply a benchmark provided by ASHRAE of 30% CV(RMSE). However, it should be noted that the ASHRAE benchmark is meant for hourly prediction of whole building energy use [139]. Comparing the models applied in this work to the ASHRAE benchmark, it can be observed that the system based models have adequate performance with an error of less than 7.3% CV(RMSE) for the electric demand and 13.8% CV(RMSE) for the thermal cooling loads.

5.5 Conclusion of the synthetic case study

This work contributes to the accomplishment of the second objective for this thesis. Two system based approaches were successfully applied to forecast the target variables of the electric demand of the HVAC system. The monolithic approach applied a single large DL model to forecast the electric demand of the secondary and primary systems. In contrast, the sequential approach applied multiple DL based models to forecast the secondary system electric demand, air-side cooling load, water-side cooling load, and the electric demand of the primary cooling system. The multi-step forecasting models applied in this work aid in the estimation of the future electric demand for the HVAC system. The short term forecasting models applied in this case study forecast over a horizon of six-hours in advance for an institutional building. The synthetic data was obtained from an eQuest model and over the summer cooling period of 2014 at hourly time steps. The results of the proposed system based approaches showed good forecasting performance; however, both such approaches have their relative merits. For example, the monolithic approach is quicker to tune the hyperparameters for; however, it does not forecast the thermal cooling loads. In contrast, the sequential approach requires a longer time for hyperparameter tuning; however, allows for additional target variables to be forecasted.

Chapter 6: System-level forecasting method using measurement data

6.1 Objectives

This chapter contributes to the main objective of this thesis in presenting a high performance forecasting model targeting the overall electric demand of a HVAC system. This chapter applies the system based deep learning models to a case study using measurement data obtained for a buildings BAS system.

6.2 Methods

6.2.1 Information Gathering

For this case study, data is obtained from the BAS of an existing building. A database is created from the control system, using 15 minute measurements of various equipment. The database for this work contains approximately 170 variables starting in 2006 until the present day. From the original database, measured variables were selected based on domain knowledge and combined with derived variables in order to create an overall database of 49 variables. The database created contains variables to describe the environmental conditions on site, variables measured in the AHUs, chilled water flow rates, and central plant information. Table 33 provides a list of both measured and derived variables.

6.2.2 Preprocessing of data

Measured values from sensors typically contain missing, faulty, and outlier based values. There are a variety of reasons why such effects may occur including: uncalibrated sensors, unexpected events in the system operation, and faulty/loose connections between the sensors and/or the data acquisition system. Due to such reasons, cleaning the data is an important step in the development of accurate forecasting based models as the models are calibrated with such data. Missing values were identified, and if the time period was less than two hours, the data from before and after the data samples were interpolated to fill in the missing data. If however, the missing values of data were longer than two hours, such data was ignored and removed (across all variables). Inconsistencies were observed in the outdoor air temperature and humidity sensors. Values here were then compared with those of Le Cam [188]; whom observed such inconsistencies and corrected them with on-site measurements.

Table 33: Measured and derived variables for the measurement case study

Variables	Point	Description	Measured/derived indicator (M/D)	Units
Target	GE_SSE_load	Electric demand of the secondary system in GE	D	kW
	AHU1&2 CC_Load	Cooling coil load on air side of both AHUs in GE	D	kW
	GE_CC_Load	Building cooling load for GE	D	kW
	GE_PSE_load	GE Electric demand contribution to CP	D	kW
Outdoor and temporal	OUT_T	Outdoor temperature	M	ºC
	OUT_H	Outdoor relative humidity	M	%
	OUT_E	Outdoor enthalpy	D	kJ
	Hour	Hour of the day	D	0…24
	Occupation	Occupied or unoccupied period	D	0,1
AHU1	AHU1_Ret_Flow	Return air flow rate	M	L/s
	AHU1_Ret_Temp	Return air temperature	M	ºC
	AHU1_Ret_RH	Return air relative humidity	M	%
	AHU1_Ret_En	Enthalpy of the return air	D	kJ/kg
	AHU1_Mod_CG_V	Cooling coil valve modulation	M	%
	AHU1_Mix_damp	Mixed air damper modulation	M	%
	AHU1_T_mix	Mixed air temperature	M	ºC
	AHU1_Sup_Flow	Supply air flow rate	M	L/s
	AHU1_Sup_Tem	Supply air temperature	M	ºC
	AHU1_Sup_RH	Supply air relative humidity	M	%
	AHU1_Sup_En	Supply air enthalpy	D	kJ/kg
AHU2	AHU2_Ret_Flow	Return air flow rate	M	L/s
	AHU2_Ret_Temp	Return air temperature	M	ºC
	AHU2_Ret_RH	Return air relative humidity	M	%
	AHU2_Ret_En	Enthalpy of the return air	D	kJ/kg
	AHU2_Mod_CG_V	Cooling coil valve modulation	M	%
	AHU2_Mix_damp	Mixed air damper modulation	M	%
	AHU2_T_mix	Mixed air temperature	M	ºC
	AHU2_Sup_Flow	Supply air flow rate	M	L/s
	AHU2_Sup_Tem	Supply air temperature	M	ºC
	AHU2_Sup_RH	Supply air relative humidity	M	%
	AHU2_Sup_En	Supply air enthalpy	D	kJ/kg
AHU totals	AHU1&2_Sup_Flow	Supply air flow rate for both AHUs	D	L/s
	AHU1&2_Sup_Tem	Average air supply temperature for both AHUs	D	ºC
GE water-side cooling load	GE_CW_flow	Chilled water flow rate entering GE	M	L/s
	GE_CW_temp_sup	Chilled water temperature entering GE	M	ºC
	GE_CW_tem_ret	Chilled water temperature leaving GE	M	ºC
Central plant	CH1_CHW pump	Chiller 1 chilled water supply pump operation	M	ON/OFF
	CH1_E	Chiller 1 electric demand	M	kW
	CH1_CW_pump	Chiller 1 condenser water pump operation	M	ON/OFF
	CT-1	Cooling tower electric demand	M	kW
	CH2_CHW pump	Chiller 2 chilled water supply pump operation	M	ON/OFF
	CH2_E	Chiller 2 electric demand	M	kW
	CH2_CW_pump	Chiller 2 condenser water pump operation	M	ON/OFF
	CT-2	Cooling tower electric demand	M	kW
	CP_PSE_load	Electric demand of the primary system in CP	D	kW
	CP_CHW_flow	Central plant chilled water supply flow rate	M	L/s
	CP_CHW_sup_tem	Central plant chilled water supply temperature	M	ºC
	CP_CHW_ret_temp	Central plant chilled water return temperature	M	ºC
	CP_CC_load	Central plant cooling load	D	kW

6.2.3 Preliminary analysis

6.2.3.1 Exploratory analysis of the control variables

An exploratory analysis is applied to target variables and a few key controllable variables. The methods applied in the exploratory analysis for the target variables will include: (i) carpet plots over the full dataset, (ii) a probability distribution graph, and (iii) summary statistics for the dataset. Due to the larger dataset and variables, day plots are not provided in this analysis.

6.2.3.2 Feature selection

For the measurement dataset, feature selection was primary based on what are the recorded variables from the BAS system. It is desired not to place additional sensors and leverage the available BAS system data in order to generate the developed models. Based on available measured and derived variables, regressors were then selected for each target variable based on reference [188] which selected regressors based on cross correlation and physics based equations. Feature extraction is accomplished for this work through the use of an auto encoder.

6.2.4 Forecasting model construction

6.2.4.1 Feature scaling

After the preprocessing of data and feature selection, the first step in the development of the forecasting models is the normalization of data. For the measurement case study, the min-max method was applied to normalize the regressor and target data. This approach is the same as that applied in the component model and the synthetic data case study. The governing equation for the min-max method is previously presented in equation 4-5.

6.2.4.2 Hyperparameter optimization

The hyperparameter search followed the same procedure as outlined in section 5.3.4.2. However, modifications to the boundary conditions of the search were needed due to the larger number of variables and increased amount of data. The modification for the boundary search of the input regressors were modified to 4, 6, 8, 10, and 12. LSTM boundary conditions were modified to 50, 100, 200, and 300 units. This was a result of a larger number of possible AE architectures and the increased training times for larger models.

6.2.4.3 Ensemble forecasting approach

The ensemble models applied in this work were similar to those describe in section 5.3.4.3 and consisted of a homogenous ensemble of EN-LSTMs. The encoder compresses the data, which is applied to the inputs of four LSTM models. Output forecasts are generated by each model in the ensemble and the overall output forecast is computed by combining the single point forecasts through an equal weights approach.

6.3 Measurement data case study

6.3.1 Case study description

The case study for this work is applied to the Genomics (GE) building located in Loyola campus of Concordia University. This is the same case study applied to the component based model and consequently, a description of the building and HVAC system can be found in section 4.3.1. For this case study, measurement data is extracted from the BAS over the time range of June 1st 2014 to September 1st 2014 at 15 minute time interval steps.

6.3.2 Governing equation

The overall governing equation for the following case study is presented in section 5.4.2 with equation 5.5. The secondary system for this case study consists of the four air supply fans. However, for this case study the primary system is part of a larger system which supplies chilled water to multiple buildings and is located at the central plant (CP) of the campus. In order to forecast the future electric demand of the GE building in relation to the requirements of the central plant primary system, the following equation derived by Le Cam is applied [188]:

$$\dot{E}^{t+15}_{GE,prim,sys} = 0.12 * \dot{E}^{t+15}_{CP.prim,sys} \qquad 6\text{-}1$$

Where $\dot{E}^{t+15}_{GE,prim,sys}$ is the future electric demand requirement for the GE building [kW], and $\dot{E}^{t+15}_{CP.prim,sys}$ is the future electric demand requirements for the central plants [kW]. The equation was derived based on measurements for the summer operation during 2014.

6.3.3 Preliminary analysis of the target variables

6.3.3.1 Preliminary analysis of the secondary system electric demand

The electric demand for the secondary cooling system is calculated based on equations 4-9 to 4-11 and presented in section 4.2.5 of the component model. However, this work differs from the component model in the type of forecasting model applied. The component based model applied a grey-box approach. In contrast, the approach for the system based models apply a black-box based approach. Therefore, the electric demand of the secondary system was calculated as a derived variable and used in training, validation and testing. For preliminary analysis purposes; the power input ratio (PIR) is applied for the secondary system electric demand. The governing equation for the PIR is calculated through equation 6-2 as shown by ASHRAE Fundamentals [9] and as a percentage.

$$\text{Power input ratio (PIR)} = \frac{P_{\text{part load}}}{P_{\text{design}}} * 100 \tag{6-2}$$

Where $P_{\text{part load}}$ is the fans power at part load (kW) and P_{design} is motor at full load/design (kW).

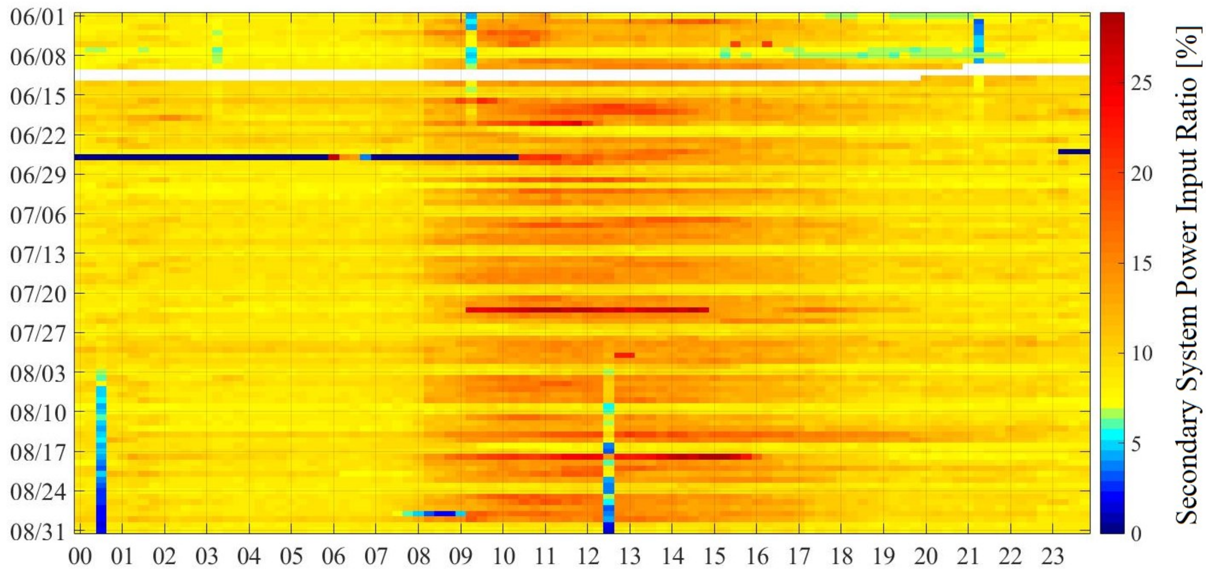

Figure 47: Secondary System Power Input Ratio for the measurement case study

The daily operation of the secondary systems electric demand (four supply fans in two AHUs) is displayed over the acquired dataset in a carpet plot presented in Figure 47. The demand varies from 0% in dark blue to approximately 30% of the design capacity in dark red. The daily operation can be observed and show an approximate 10% PIR during unoccupied times. Occupied times appear starting from 08:00 until 18:00 and show a PIR that operates between approximately 12 to 30%.

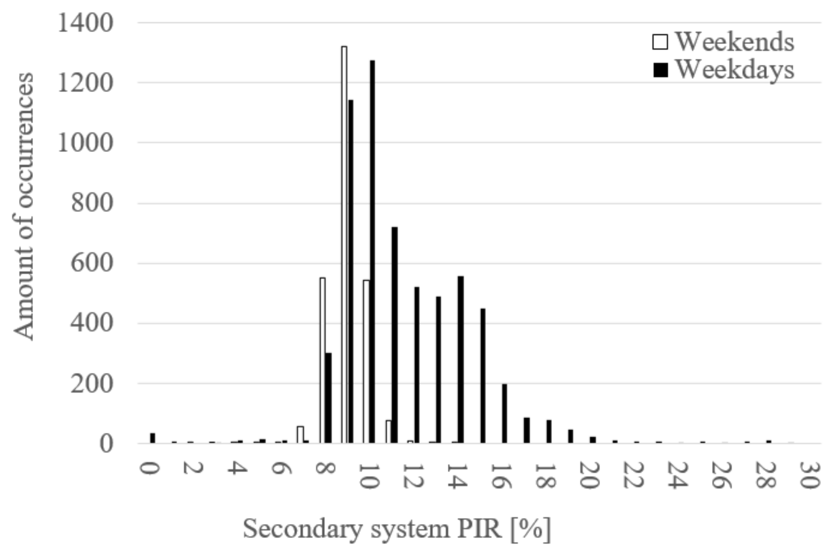

Figure 48: Probability distribution of the secondary system PIR for the measurement case study

The probability distribution for the secondary systems PIR is shown in Figure 48. White bars depicted the distributions of PIR during the weekends, while the black bars shows the distribution during weekdays. The median value for weekday PIR was observed to be 10.23% while for weekend it was observed to be 9.38%. The summary statistics for the PIR of the secondary system are presented in Table 34. Maximum values show a PIR of 13.91% for weekends and 28.87% for weekdays.

Table 34: Summary statistics secondary system PIR for the measurement case study

Summary Statistics	Full dataset [%]	Weekends [%]	Weekdays [%]
Mean	10.28	8.49	11.05
Minimum values	0.00	1.51	0.00
Maximum values	28.87	13.91	28.87
Range	28.87	12.40	28.87
Quartile 1	8.46	8.01	8.99
Median	9.38	8.42	10.23
Quartile 3	11.69	8.99	12.99
Standard Deviation	2.87	0.93	3.07

6.3.3.2 Preliminary analysis of the air-side cooling load

The cooling load on the air side of the AHU is calculated using equation 5.8 in section 5.2.4. It is assumed that the air density is constant at 1.2 [kg/m^3] at 101 [kPa] and 20 [ºC]. For this case study, the cold deck temperature is calculated by taking the measurements of the air supply available from the BAS and subtracting the temperature rise of the supply fans. From reference [188], it was observed that the fans exhibited a temperature increase of approximately 1.8 [ºC] from measurements of the supply air and mixed air when cooling coil values are closed. Furthermore, it was assumed that the absolute humidity remained constant across the fans.

Mixed air enthalpy is calculated from temperature measurements from the BAS system and derived relative humidity. The governing equations applied in order to derive the relative humidity are shown in equations 6-3 to 6-5.

$$\mathrm{RM_{MA}} = \alpha * \mathrm{RH_{OA}} + (1 - \alpha) * \mathrm{RH_{RA}} \tag{6-3}$$

$$\alpha = \frac{\dot{\mathrm{V}}_{\mathrm{sup,a}} - \dot{\mathrm{V}}_{\mathrm{ret,a}}}{\dot{\mathrm{V}}_{\mathrm{sup,a}}} \tag{6-4}$$

$$\alpha = \frac{\mathrm{T_{mix,a}} - \mathrm{T_{ret,a}}}{\mathrm{T_{out,a}} - \mathrm{T_{mix,a}}} \tag{6-5}$$

Where $\mathrm{RM_{MA}}$ is the mixed air relative humidity, $\mathrm{RH_{OA}}$ is the outdoor air relative humidity, $\mathrm{RH_{RA}}$ is the relative humidity of the return air, $\dot{\mathrm{V}}_{\mathrm{sup,a}}$ is the supply air volumetric flow rate [L/s], $\dot{\mathrm{V}}_{\mathrm{ret,a}}$ is the volumetric flow rate of the return air [L/s], $\mathrm{T_{mix,a}}$ is the temperature of the mixing box, $\mathrm{T_{ret,a}}$ is the temperature of the return air, $\mathrm{T_{out,a}}$ is the outdoor air temperature, and α is the outdoor air flow ratio. When the mixing dampers are completely closed α is equal to 1. However, when the mixing dampers are completely open, α is derived from equation 6-4. When the dampers are partially open, α is derived using equation 6-5. It was observed that over the dataset, there are only few days in which the dampers are opened (less than three). Figure 49 presents the outdoor air ratio calculated over the dataset and as a percentage. Blue values indicate 100% while red values correspond to 60%.

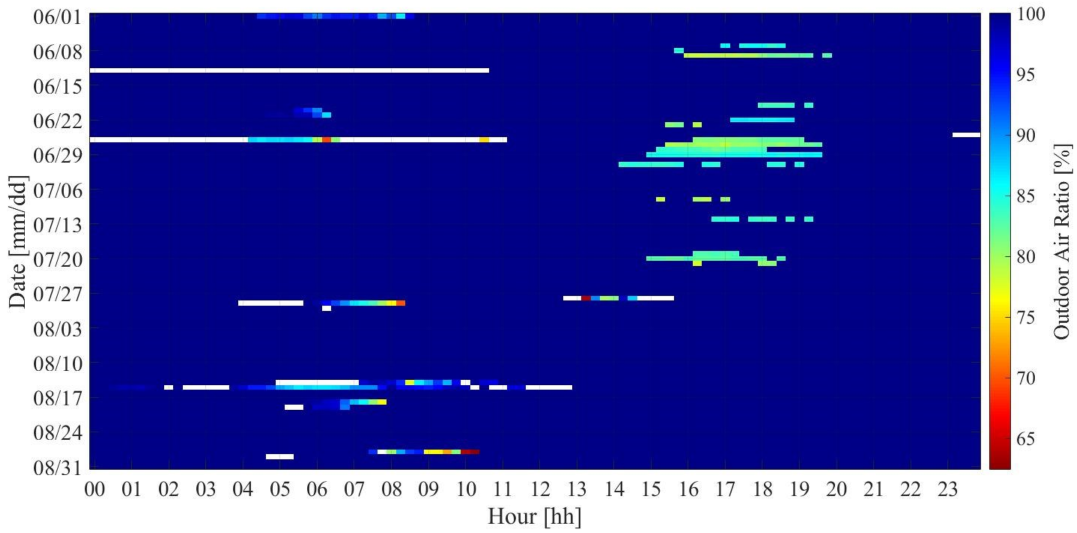

Figure 49: Carpet plot for the outdoor air ratio for the measurement case study

Based on the governing equations, the air-side cooling load is calculated over the dataset. Figure 50 presents a carpet plot for the air-side cooling load. The blue to dark blue values correspond to values of 0 to 900 kW. Furthermore, yellow to red colors indicate periods of "free cooling mode" in which the cooling coil, heating coil, and humidifier are not operated. Such times are omitted from the training and application of the system based models. From Figure 50, it can be observed that the cooling load begins to increase in demand at approximately 08:00 following closely with the PIR which also begins to ramp up at this hour.

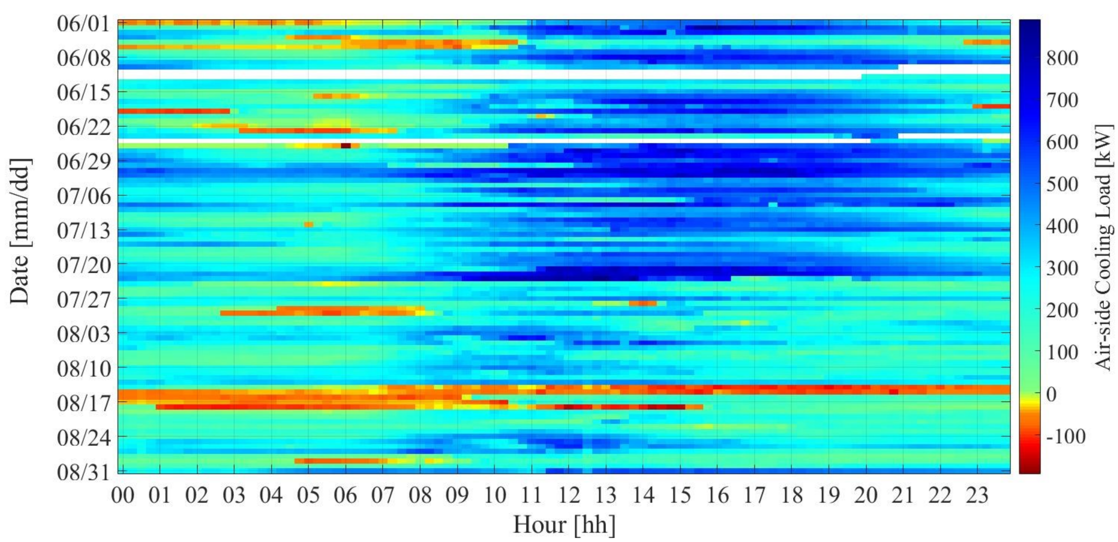

Figure 50: Carpet plot of the air-side cooling load for the measurement case study

The probability distribution for the air-side cooling load of GE is presented in Figure 51 for weekdays and weekends. The black values correspond to weekday occurrences and the white values correspond to weekends. Higher loads are seen occurring for weekdays in comparison to weekends in which building occupancy is low.

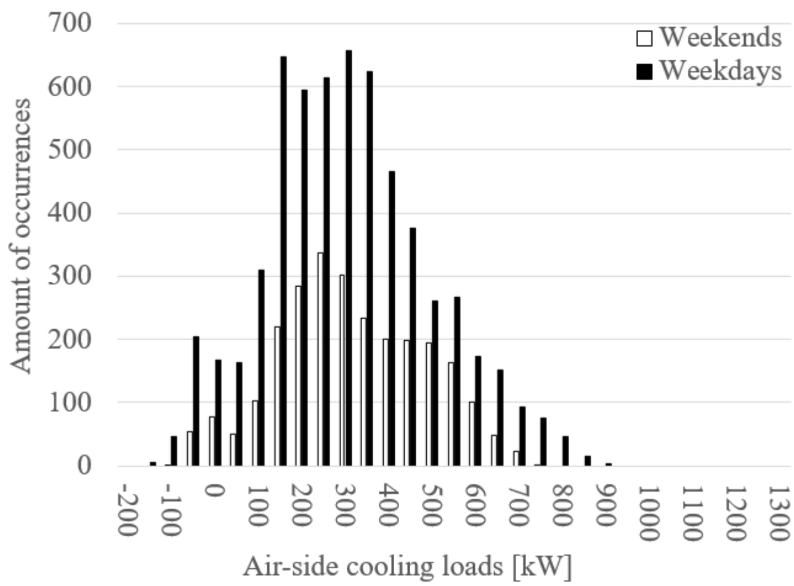

Figure 51: Probability distribution of the air-side cooling load for the measurement case study

Summary statistics for the air-side cooling loads are provided in Table 35 for the full, week day and weekend values. Maximum demand values were observed for weekdays to be 888 kW and 716 kW for weekends. Median values were observed to be 267 kW for weekdays and 277 kW for weekends.

Table 35: Summary Statistics air-side cooling load for the measurement case study

Summary Statistics	Full dataset (kW)	Weekends (kW)	Weekdays (kW)
Mean	283.71	292.30	279.97
Minimum values	-191.40	-107.72	-191.40
Maximum values	887.68	716.12	887.68
Range	1079.08	823.84	1079.08
Quartile 1	155.99	173.51	145.20
Median	271.06	276.82	267.20
Quartile 3	404.69	421.02	396.35
Standard Deviation	183.92	166.88	190.74

6.3.3.3 Preliminary analysis of the water-side cooling load

The GE water-side cooling load is calculated through equation 5-9 in section 5.4.2. This case study assumes that the heat capacity of the water remains constant at 4.2 [kJ/kg°C]; moreover, the density of the water is also assumed constant at 1,000 [kg/m^3].

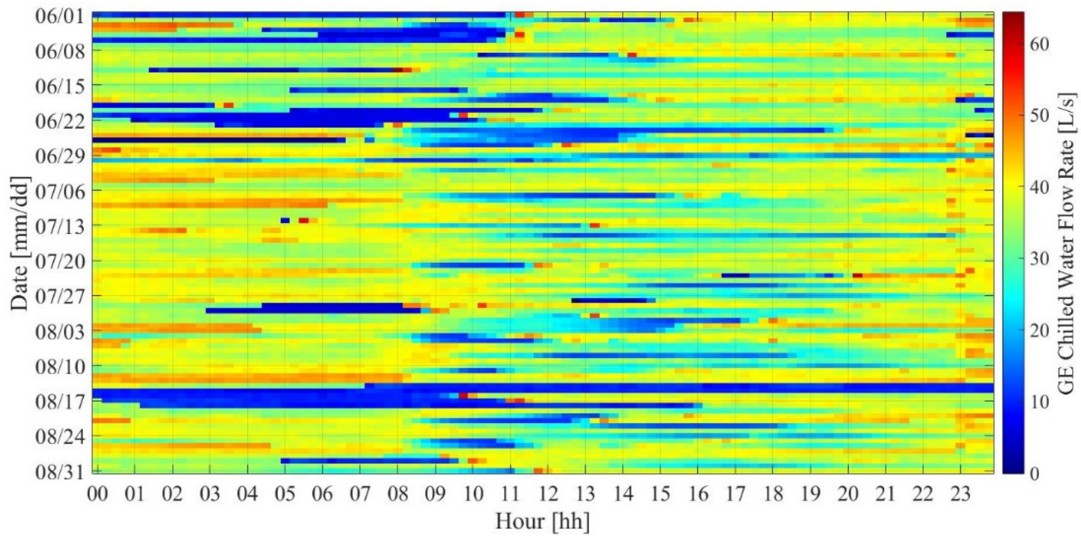

Figure 52: Carpet plot for the GE chilled water flow rate for the measurement case study

An exploratory analysis is conducted on the chilled water flow rate to the GE building over the applied dataset and shown in Figure 52. The figure demonstrates a higher flow rate (40 L/s) during periods of 00:00 to 08:00 and approximately 18:00 to 23:00. This is shown in the yellow and orange colors within Figure 52. Furthermore, lower demand (25 L/s) periods are observed during 08:00 to 18:00 as shown in the light blue colors in Figure 52.

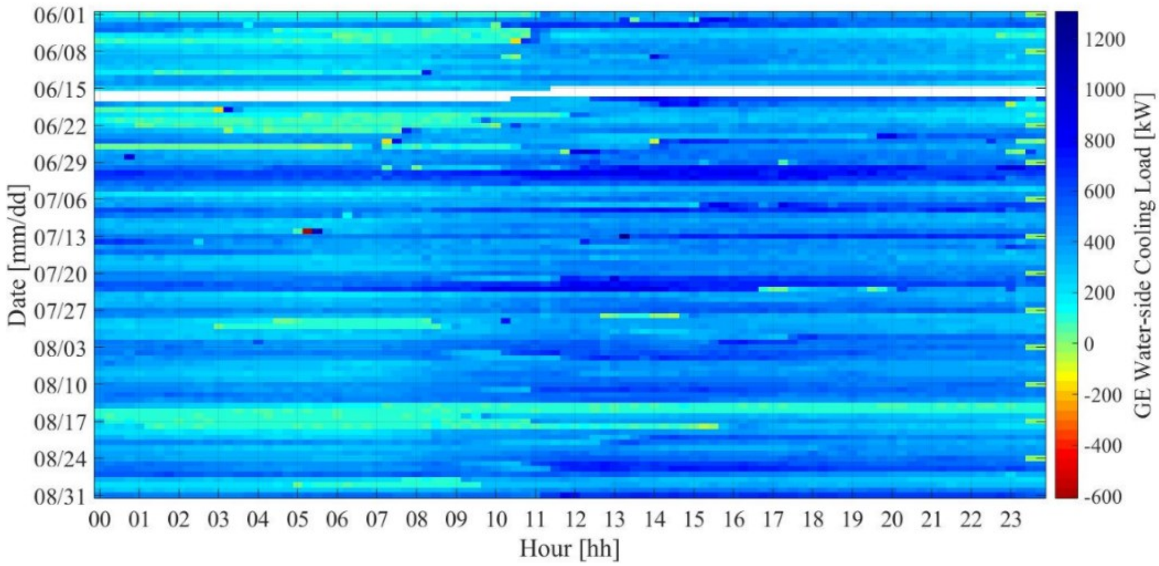

Figure 53: Carpet plot for GE water-side cooling load for the measurement case study

The exploratory analysis of the GE water-side cooling load begins with a carpet plot and is presented in Figure 53. The figure demonstrates similarities with the air-side cooling load and the PIR showing an increasing cooling demand beginning at approximately 08:00. Higher demand periods show cooling loads with values greater than 600 kW while low demand periods show loads at approximately 200 kW. It should be noted, that certain days can be observed which have no cooling load. For instance, 8/17/214, such instances are when the outdoor temperature is low.

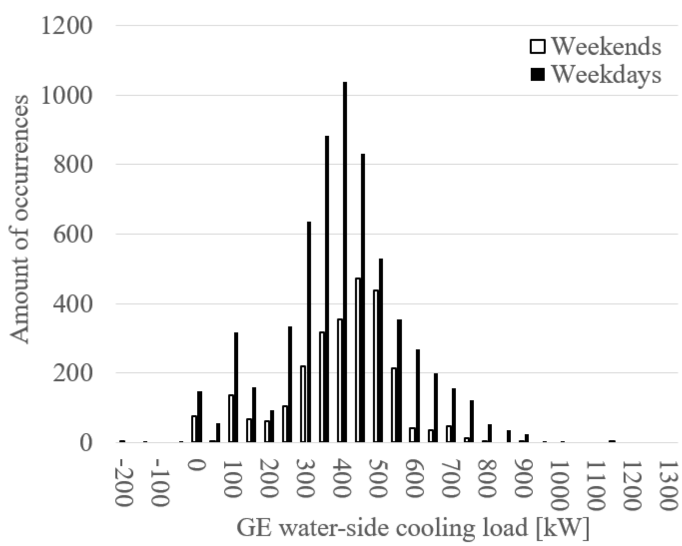

Figure 54: Probability distribution of the water-side cooling load for the measurement case study

A probability distribution graph is presented in Figure 54 for the GE water-side cooling load with datasets separated in to weekdays (black) and weekends(white). Weekdays demonstrate a larger number of high demand loads compared to that of weekends. Summary statistics are provided for the GE water-side cooling load and are presented in Table 36. The median values were observed to be 379 kW for the full dataset, 372 kW for weekdays, and 395.43 for weekends. Furthermore, the mean values were observed to be 376 kW for weekdays, 370 kW for weekends, and 374 kW for the overall dataset.

Table 36: Summary statistics for the water-side cooling load for the measurement case study

Summary Statistics	Full dataset (kW)	Weekends (kW)	Weekdays (kW)
Mean	374.25	370.33	375.88
Minimum values	-607.49	-607.49	-164.54
Maximum values	1306.25	1306.25	989.75
Range	1913.74	1913.74	1154.29
Quartile 1	290.23	297.38	287.45
Median	379.22	395.34	372.41
Quartile 3	464.08	464.45	463.87
Standard Deviation	162.43	148.17	167.97

6.3.3.4 Preliminary analysis of the primary system electric demand

The primary systems electric demand is not a directly recorded value by the BAS and requires computation in order to be extracted. Among the components which contribute to the overall electric demand of the central plant primary system, some are directly monitored values while others required modifications. Beginning with the direct measurements, the electric demand of both 900 ton chillers is monitored and recorded by the BAS. Therefore, no modifications are required in order to capture the electric demand of the chiller components.

However, the system pumps require modification in order to be able to compute the primary system electric demand. The central plant system contains two constant speed pumps assigned to the chilled water loop with a design capacity of 74.6 kW for each pump. Furthermore, there are two constant speed pumps for the condenser water loop with a design capacity of 55.9 kW for each pump. Within the BAS, all four pumps are monitored and recorded as [ON/OFF]. As a consequence, modifications are required to apply the monitored values. The approach applied in this work consisted of assigning the manufacturers rated capacity for times when pumps were 'ON' and zero for when pumps were 'OFF'. It was assumed that the pumps are still operating at their rated specifications and no degradation has occurred.

Focusing on the cooling tower fans, the BAS system measures the fan modulation for each fan. In order to derive the electric demand; the manufactures equation (6-6) is applied.

$$\dot{E}_{CT} = P_{CT}\left(\frac{mod_t}{100}\right)^3 \tag{6-6}$$

Where P_{CT} is the designed power demand for each fan (29.8 kW) and mod_t is the fan modulation at time t.

Based on the system and the available BAS measurements, equation 6-7 was applied in order to calculate the electric demand of the central plant primary cooling system.

$$\dot{E}_{p,s} = \dot{E}_{CHW,p1} + \dot{E}_{CH1} + \dot{E}_{CW,p1} + \dot{E}_{CT1} + \dot{E}_{CHW,p2} + \dot{E}_{CH2} + \dot{E}_{CW,p2} + \dot{E}_{CT2} \tag{6-7}$$

Where $\dot{E}_{p,s}$ is the electric demand for the central plant primary system, $\dot{E}_{CHW,p}$ is the electric demand for the chilled water pumps 1 and 2, $\dot{E}_{CH}$ is the electric demand for chillers 1 and 2, $\dot{E}_{CW,p}$ is the electric demand condenser water pumps 1 and 2, and $\dot{E}_{CT}$ is the electric demand for the cooling tower fans 1 and 2.

After the modifications were applied for the components within the overall HVAC system, a dataset of the central plant's primary system was completed. Figure 55 provides a carpet plot for the primary system electric demand. Values in red depict times with no electrical consumption and such values are omitted from the training and testing of the models. Higher demand periods (800 to 1,200 kW) can be seen in cyan to blue colors. Lower demand periods (200 to 500 kW) can be seen in orange and yellow colors.

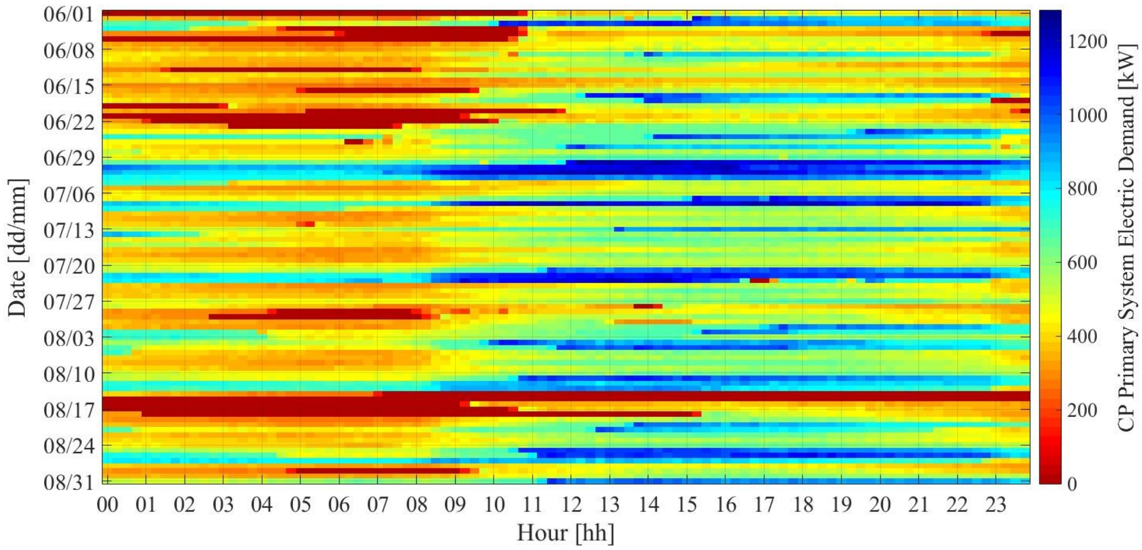

Figure 55: Carpet plot for the central plant primary cooling system for the measurement case study

A probability distribution graph for the primary systems electric demand is shown in Figure 56. The weekday values are shown in black, while the weekend values are shown in white. Higher demand periods appear in weekdays, along with a larger number of such periods. The times of zero demand are shown in the left most side of the figure. Such zero times are omitted from the application of this study along with the values of other regressors at the same time steps.

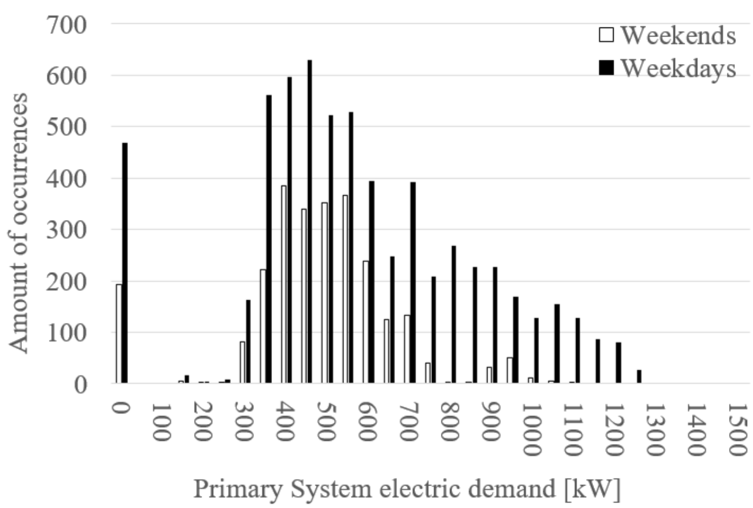

Figure 56: Probability distribution of central plant primary cooling system for the measurement case study

Table 37 provides the summary statistics for the central plant primary system electric demand. The maximum electric demands were observed to be 1,281 kW for weekdays and 1,063 kW for weekends. Mean values were observed as 552 kW for weekdays, 454 kW for weekends, and 523 kW for the full dataset.

Table 37: Summary statistics for the central plant primary cooling system for the measurement case study

Summary Statistics	Full dataset (kW)	Weekends (kW)	Weekdays (kW)
Mean	523.03	454.22	551.61
Minimum values	0.00	0.00	0.00
Maximum values	1281.43	1062.80	1281.43
Range	1281.43	1062.80	1281.43
Quartile 1	376.93	373.57	379.38
Median	490.77	458.56	515.92
Quartile 3	661.41	549.35	736.04
Standard Deviation	256.78	186.48	275.85

6.3.4 Forecasting model construction: Monolithic approach

6.3.4.1 Selection of regressors

The monolithic approach applies all the regressors into one overall model. The selection of regressors for the monolithic model is based on BAS data (measured and derived), reference [188] and physics based equations. In reference [188], regressors were selected for each target variable applied to the same system. The target variables in reference [188] differed as models were applied to forecast the electric demand of the chillers and cooling towers; in contrast, this work directly forecasts the electric demand of the primary system. However, the regressors found for the chillers and cooling tower fans are applied as regressors for the work herein. Table 38 provides the list of regressors to be applied within the monolithic forecasting model.

Table 38: Selected regressors and the target variables of the monolithic approach of the measurement case study

Target variable(s)	Regressor(s)	
• Secondary system electric demand • Primary system electric demand	• Hour • Occupation • Outdoor air temperature • Outdoor enthalpy • AHU 1&2 cooling coil valve modulation • AHU 1&2 mixed air temperature • AHU average supply air temperature • AHU total air supply • Chilled water flow rate entering GE • Secondary System electric demand • GE air-side cooling load • GE water-side cooling load	• Chiller operation • Chiller 1 and 2 electric load • Cooling tower 1 and 2 electric load • Chilled water flow rate central plant • Chilled water return temperature • Central plant cooling load • Primary system electric load

6.3.4.2 Hyperparameter tuning

Due to the large amount of regressors, the computational time and resources were significant for models which contained a larger number of lagged values (greater than 12 lags). Such values were observed to have a large computational time, non-convergence issues, and did not result in a significant improvement in performance. As the increased computational time went against the objective of this thesis of creating a fast and accurate forecasting model; the boundary conditions for the monolithic model was reduced to the lags of 4, 6, 8, 10, and 12. Table 39 presents the average results for the autoencoder hyperparameter search applied over the training dataset. H1 to H5 present the best architectures observed for each given lag. The architectures represent the full undercomplete autoencoder.

Table 39: Autoencoder hyperparameter search results of the monolithic model applied to the measurement case study

Number of lag	H1	H2	H3	H4	H5	Average CV(RMSE) (%)
4	84	75	40	75	84	6.10
6	126	115	70	115	126	6.04
8	168	160	85	160	168	5.96
10	210	190	110	190	210	5.95
12	252	240	140	240	252	6.21

Selecting the top performing encoder architectures presented in Table 39 (H1 to H3), LSTM models were coupled to each and the forecasting performance was explored over a combined dataset of the training and validation data. The forecasting performance of the combined EN-

LSTM single point forecasting models was recorded and averaged; the results of the different architectures are presented in Table 40. The results demonstrate an increasing error for a greater number of lagged values. The term N/A refers to situations of non-convergence occurring during training.

Table 40: EN-LSTM hyperparameter search results for the measurement case study

Number of lags	Average CV(RMSE) Secondary system electric demand (%)				Average CV(RMSE) Primary system electric demand (%)			
	50 LSTM Units	100 LSTM Units	200 LSTM Units	300 LSTM Units	50 LSTM Units	100 LSTM Units	200 LSTM Units	300 LSTM Units
4	11.84	13.48	12.83	14.16	6.34	5.88	6.30	8.01
6	13.58	10.53	12.45	12.82	14.54	8.99	6.65	7.54
8	14.92	11.68	12.60	14.66	25.55	12.73	14.70	9.93
10	8.38	10.87	12.43	12.71	38.86	21.02	8.59	11.79
12	12.95	13.97	16.24	N/A	23.60	11.11	10.57	N/A

Based on the forecasting results for the single point forecasting models presented in Table 40, ensemble models are estimated in order to find the boundary conditions for the monolithic forecasting model which will be applied to the testing dataset. The results demonstrate that a lag of 4 variables with an 84-75-40 encoder and 50 to 200 LSTM units contained the lowest forecasting error over the validation dataset.

Table 41: EN-LSTM ensemble hyperparameter research results for the measurement case study

Number of lags	Average CV(RMSE) for monolithic ensemble (%)							
	50 to 200 LSTM Units		50 to 300 LSTM Units		100 to 200 LSTM Units		100 to 300 LSTM Units	
	Sec Sys	Prim Sys	Sec Sys	Prim Sys	Sec Sys	Prim Sys	Sec Sys	Prim Sys
4	12.71	6.17	13.08	6.63	13.15	6.09	13.49	6.73
6	12.19	10.06	12.35	9.43	11.49	7.82	11.93	7.72
8	13.07	17.66	13.47	15.73	12.14	13.72	12.98	12.45
10	10.56	22.83	11.10	20.07	11.65	14.81	12.00	13.80
12	14.39	15.09	N/A	N/A	15.11	10.84	N/A	N/A

In order to tune the number of training days for the model, the single point forecasting model which produced the lowest error was selected. This was the 4 lagged, 84-75-40 EN, and 50 unit LSTM model. The length of training days were varied at 59, 50, 40, 30, and 20 days in length. For each number of training days, the single point forecasting model was re-trained and then applied to the validation dataset. The forecasting performance over the validation data was recorded and the performance was then averaged. Figure 57 presents the results for the adjustment of the length of training data with the EN-LSTM model. The results demonstrate that as the length of training days (horizontal axis) increases, the forecasting error (vertical axis) begins to decrease. Based on the result of the hyperparameter optimization phase, the forecasting model to be applied will consist of a 4 lagged, 84-75-40 encoder, 50 to 200 LSTM units, and 59 days of training data.

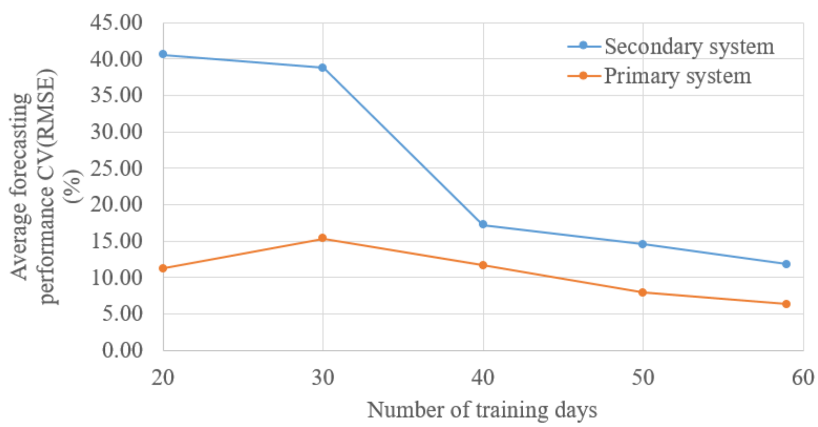

Figure 57: Adjustment of training days for a single point forecasting monolithic model applied to the measurement case study

6.3.5 Forecasting model results: Monolithic approach

This section presents the results for the monolithic model forecasting the target variables at 08:45 July 30th, 2014 over the horizon of 09:00 to 14:45. The applied forecasting model consisted of four EN-LSTM models, using 84-75-40 encoder architecture, and 50 to 200 LSTM units, and trained with 59 days of data.

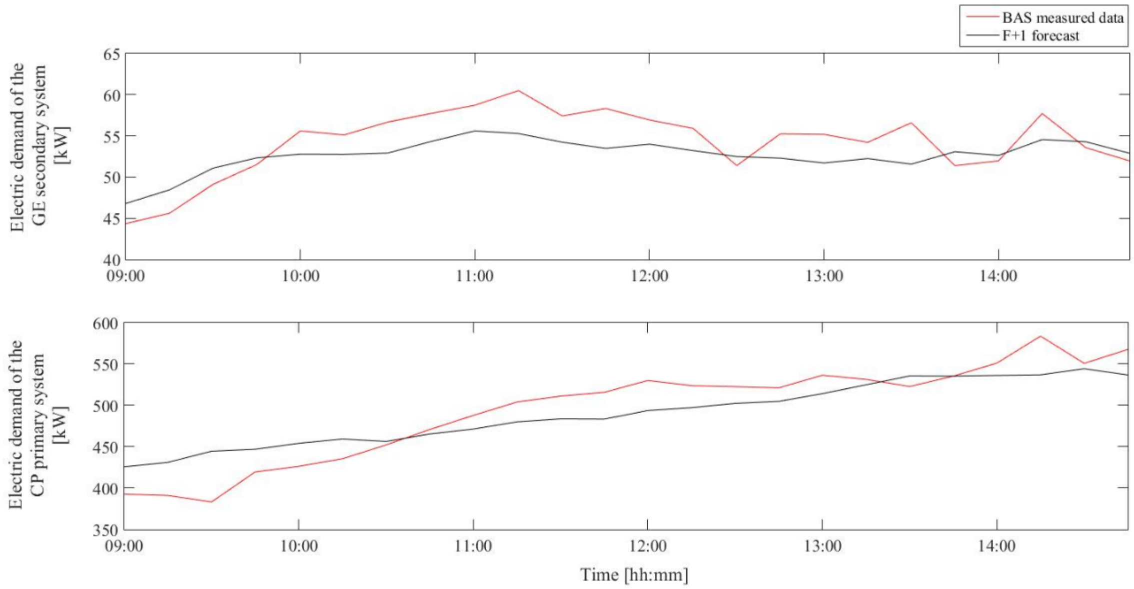

Figure 58: F+1 forecasts for the monolithic model applied to the measurement case study

Figure 58 presents the F+1 for the measurement case study. The top graph presents the output forecasts for the GE secondary system electric demand while the bottom graph presents the output forecasts for the CP primary system electric demand. The red lines in both graphs indicate the measured values for the target variables and the black lines indicate the forecasted values. The monolithic model performance results of the F+1 forecasts are presented in Table 42.

Table 42: F+1 forecasts results for the monolithic model applied to the measurement case study

Target variables	**Performance indices**	
	CV(RSME) (%)	RMSE (kW)
Electric demand of the GE secondary system	5.45	2.96
Electric demand of the CP primary system	5.55	27.43
Electric demand of the GE HVAC System	4.75	5.40

6.3.6 Forecasting model construction: Sequential approach

6.3.6.1 Selection of regressors for the target variables of the sequential model

The selection of regressors for the sequential model is based on reference [188] with some small variations. Firstly, this work applied forecasts from the upstream DL models as inputs to the subsequent model. Secondly, the work in reference [188] targeted forecasting the electric demand of the chillers and cooling towers as two separate models each with its own set of regressors. In contrast, the approach in this work forecasts the overall electric load of the primary system, and combines both sets of regressors into a single set. The historical data of the primary system is additionally incorporated as well. An overview for each forecasting model in the sequential approach is listed in Table 43. The table provides a list of the target variables and the regressors applied as inputs to each forecasting model. All values apply current and historical measurement data as inputs, however, those variables marked with an asterisks (*) additionally provide future estimates as inputs.

Table 43: Sequential model regressors applied to the measurement data case study

Target variable (s)	Regressor (s)	
• Secondary system electric demand	• Hour • Occupation	• GE AHU total air supply • Secondary System electric demand
• GE Air-side cooling load	• Hour • Outdoor air temperature • Outdoor enthalpy • GE AHU total air supply	• AHU 1&2 mixed air temperature • GE air-side cooling load • Secondary System electric demand*
• GE Water-side cooling load	• Hour • Outdoor air temperature • Outdoor enthalpy • AHU 1&2 mixed air temperature • Chilled water flow rate entering GE	• AHU 1&2 cooling coil valve modulation • AHU average supply air temperature • GE water-side cooling load • GE air-side cooling load*
• Primary system electric demand	• Hour • Outdoor air temperature • Outdoor enthalpy • Chiller operation • Chiller 1 and 2 electric load • Cooling tower 1 and 2 electric load	• Chilled water flow rate central plant • CP chilled water return temperature • Central plant cooling load • Primary system electric load • GE water-side cooling load*

6.3.6.2 Hyperparameter tuning results: Secondary system electric demand

This section presents the results found from the hyperparameter search for the secondary system electric demand model within the sequential approach. First a grid search was conducted to find the AE architectures which could best reconstruct the inputs given a certain number of lagged values and regressors. Based on the regressors listed in Table 43 the results of the grid search conducted are presented in Table 44. Within Table 44 H refers to the hidden layer.

Table 44: Autoencoder hyperparameter search results for the secondary system model applied to the measurement data case study

Number of lags	H1	H2	H3	H4	H5	Average CV(RMSE) (%)
4	16	10	5	10	16	7.47
6	24	20	15	20	24	6.27
8	32	20	15	20	32	7.29
10	40	35	25	35	40	6.24
12	48	45	40	45	48	5.87

The second step in the model development, involves a grid search for coupled EN-LSTM models. The encoder values of Table 44 are combined with LSTM forecasting models, trained and then applied to the validation dataset. The performance of each architecture is recorded over the validation data set and then averaged. The results demonstrate that a four lagged, 16-10-5 EN with 100 to 200 LSTM units obtained the lowest forecasting error. Therefore, this model is to be selected and applied to the testing dataset.

Table 45: EN-LSTM single point and ensemble hyperparameter research of the secondary system model applied to the measurement case study

Number of lag	Average CV(RMSE) of single point EN-LSTM architectures (%)				Average CV(RMSE) of ensemble of EN-LSTM architectures (%)			
	50 LSTM Units	100 LSTM Units	200 LSTM Units	300 LSTM Units	50 to 200 LSTM Units	50 to 300 LSTM Units	100 to 200 LSTM Units	100 to 300 LSTM Units
4	12.28	11.80	10.90	12.19	11.66	11.79	11.35	11.63
6	13.09	10.63	12.59	13.06	12.10	12.34	11.61	12.09
8	14.02	12.83	12.47	11.84	13.11	12.79	12.65	12.38
10	14.42	12.17	12.67	12.38	13.09	12.91	12.42	12.41
12	13.89	14.13	13.96	13.02	13.99	13.75	14.04	13.70

Selecting the top performing architecture (4 lagged, 16-10-5 EN, 200 unit LSTM) the length of training days is varied and the model is tested. The training data length is varied over 59, 50, 40, 30, and 20 days. For each given length of training data, the selected architecture is trained, applied to the validation data, and the performance is averaged. Figure 59 presents the results from the adjustment of days for the architecture applied. The results demonstrate that the model performs with the lowest CV(RMSE) when the training length is at a maximum.

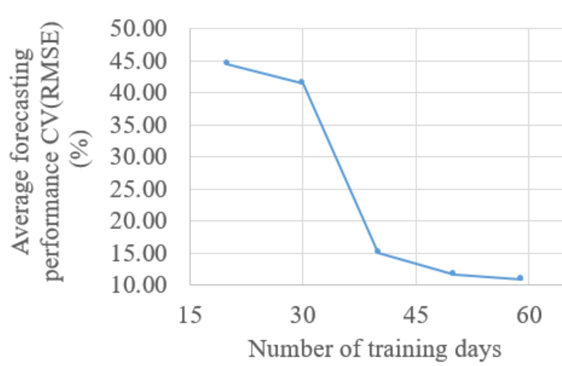

Figure 59: Adjustment of training days for a single point secondary system forecasting model applied to the measurement case study

6.3.6.3 Hyperparameter tuning results: Air-side cooling demand

This section outlines the hyperparameter search results of the GE air-side cooling load for the sequential forecasting model. Table 46 presents the results of the AE hyperparameter search.

Table 46: Autoencoder hyperparameter search results for the air-side cooling model for the measurement case study

Number of lags	H1	H2	H3	H4	H5	Average CV(RMSE) (%)
4	56	50	45	50	56	3.38
6	72	55	40	55	72	3.87
8	88	85	45	85	88	3.97
10	104	100	80	100	104	4.45
12	120	115	70	115	120	4.52

Selecting the encoder architectures (H1 to H3) from Table 46, LSTM models are coupled to each encoder in order to explore the forecasting performance of each EN-LSTM model. The models are trained, applied to the validation dataset, and the average forecasting performance is recorded. The results for the single point forecasting models of the air-side cooling load are presented in Table 47. Furthermore, the ensemble boundaries were calculated and are similarly presented. It was observed that the 4 lagged 56-50-45 encoder, and 100 to 200 LSTM model obtained the lowest average forecasting performance over the validation dataset; therefore, such a model will be applied to the testing data.

Table 47: EN-LSTM single point and ensemble hyperparameter research of the air-side cooling model applied to the measurement case study

Number of lag	Average CV(RMSE) of EN-LSTM architecture applied to validation dataset (%)				Average CV(RMSE) of ensemble of EN-LSTM architectures applied to validation dataset (%)			
	50 LSTM Units	100 LSTM Units	200 LSTM Units	300 LSTM Units	50 to 200 LSTM Units	50 to 300 LSTM Units	100 to 200 LSTM Units	100 to 300 LSTM Units
4	40.39	30.72	28.78	33.51	33.30	33.35	29.75	31.00
6	50.84	49.56	45.46	40.84	48.62	46.68	47.51	45.29
8	45.01	52.12	37.66	34.51	44.93	42.32	44.89	41.43
10	44.64	34.73	46.41	39.32	41.93	41.28	40.57	40.15
12	46.03	43.74	35.18	45.62	41.65	42.64	39.46	41.52

The 4 lagged, 56-50-45 EN, and 200 unit LSTM model is selected in order to explore the effects of tuning the training days for a single point forecasting model. Figure 60 presents the results from the adjustment of training days for the air-side cooling load forecasting model when applied to the validation dataset. It is observed that the model performs with the lowest average error with the maximum number of training days.

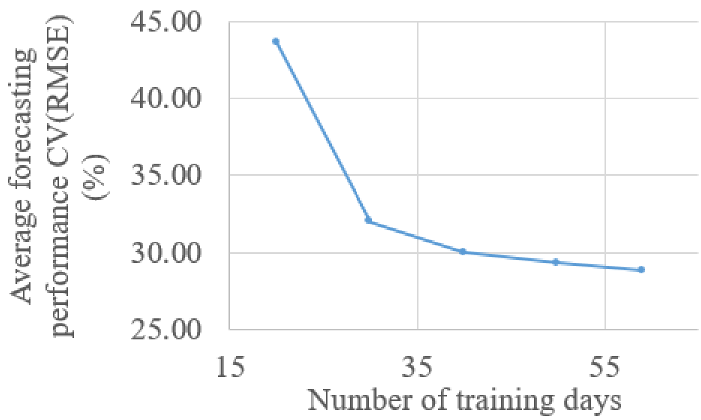

Figure 60: Adjustment of training days for a single point air-side cooling forecasting model applied to the measurement case study

6.3.6.4 Hyperparameter tuning results: Water-side cooling demand

This section outlines the hyperparameter search results for the GE water-side cooling load of the sequential forecasting model. Table 48 presents the architecture search results for the AE. It can be observed that the four lag model obtained the lowest average reconstruction error.

Table 48: Autoencoder hyperparameter search results for the water-side cooling model applied to the measurement case study

Number of lags	**H1**	**H2**	**H3**	**H4**	**H5**	**Average CV(RMSE) (%)**
4	68	65	50	65	68	4.67
6	90	85	40	85	90	5.21
8	112	105	95	105	112	5.30
10	134	130	105	130	134	5.33
12	156	150	100	150	156	5.47

The encoder architectures of Table 48 were coupled to various LSTM models and a grid search was conducted to explore the various architectures of combined EN-LSTM models. Table 49 presents the average forecasting results for the single point EN-LSTM models applied to forecast the GE water-side cooling load over the validation dataset; furthermore, the average results of the ensemble models are calculated. It was observed that the eight lagged, 112-105-95 EN and 100 to 200 LSTM unit model obtained the lowest CV(RMSE) of 8.47%. Therefore, this architecture will be applied to the testing dataset.

Table 49: EN-LSTM single point and ensemble hyperparameter research of the water-side model applied to the measurement case study

Number of lag	Average CV(RMSE) of EN-LSTM architecture applied to validation dataset (%)				Average CV(RMSE) of ensemble of EN-LSTM architectures applied to validation dataset (%)			
	50 LSTM Units	100 LSTM Units	200 LSTM Units	300 LSTM Units	50 to 200 LSTM Units	50 to 300 LSTM Units	100 to 200 LSTM Units	100 to 300 LSTM Units
4	10.22	11.34	11.22	12.41	10.93	11.30	11.28	11.66
6	12.04	10.92	8.42	8.47	10.46	9.96	9.67	9.27
8	12.71	8.47	8.47	11.62	9.88	10.31	8.47	9.52
10	10.75	9.19	9.82	14.70	9.92	11.11	9.51	11.24
12	12.21	8.83	18.87	13.98	13.31	13.47	13.85	13.90

The single point forecasting model which obtained the lowest average CV(RMSE) was selected in order to tune the number of training days. This consisted of a six lagged, 90-85-40 EN with 200 LSTM units. Figure 61 provides the results observed which demonstrate that the forecasting models obtain the lowest error with the maximum number of training days. Therefore, such a length will be used for the application of the forecasting models to the testing dataset.

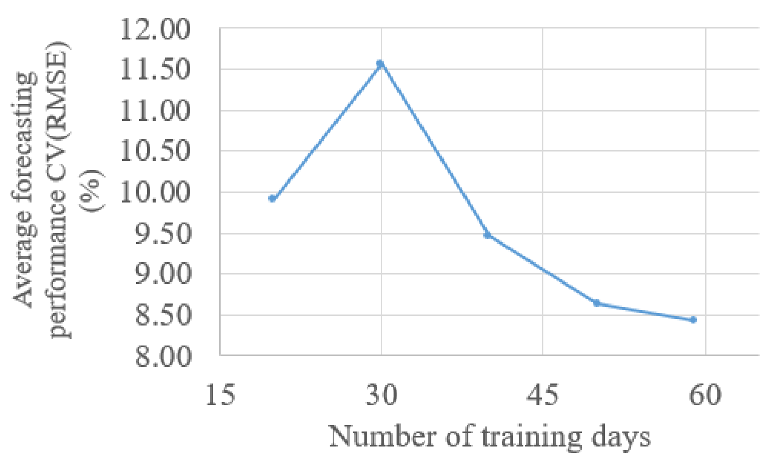

Figure 61: Adjustment of training days for a single point water-side cooling forecasting model applied to the measurement case study

6.3.6.5 Hyperparameter tuning results: Primary system electric demand

This section describes the results from the hyperparameter search for the primary system electric demand model. Table 50 presents the result from the AE architecture search; a different of less than 2% CV(RMSE) was observed from the architectures with four to twelve lags.

Table 50: Autoencoder hyperparameter search results of the primary system model applied to the measurement case study

Number of lags	H1	H2	H3	H4	H5	Average CV(RMSE) (%)
4	68	65	50	65	68	4.05
6	90	80	65	80	90	4.58
8	112	105	80	105	112	5.09
10	134	120	105	120	134	5.32
12	156	140	110	140	156	5.43

The EN architectures of Table 50 were then coupled to LSTM models. The EN-LSTM single point forecasting models were trained and then applied to the validation dataset recording the average performance. The results are presented in Table 51 on the left hand side for the single point forecasting models and on the right hand side for the ensemble models. The architecture of four lags, 68-65-50 EN, and 100 to 300 LSTM units is observed to obtain the lowest forecasting error and will be applied to the testing dataset.

Table 51: EN-LSTM single point and ensemble hyperparameter research of the primary system model applied to the measurement case study

Number of lag	Average CV(RMSE) of EN-LSTM architecture applied to validation dataset (%)				Average CV(RMSE) of ensemble of EN-LSTM architectures applied to validation dataset (%)			
	50 LSTM Units	100 LSTM Units	200 LSTM Units	300 LSTM Units	50 to 200 LSTM Units	50 to 300 LSTM Units	100 to 200 LSTM Units	100 to 300 LSTM Units
4	6.39	6.91	6.62	5.25	6.64	6.29	6.77	6.26
6	7.68	7.02	7.18	8.98	7.29	7.71	7.10	7.73
8	8.03	10.27	13.10	7.70	10.47	9.78	11.69	10.36
10	19.08	10.48	11.50	9.29	13.69	12.59	10.99	10.42
12	27.59	7.89	11.61	31.01	15.70	19.52	9.75	16.84

An architecture was selected in order to tune the number of training days required; this consisted of a 4 lagged, 68-65-50 EN, and a 300 LSTM unit model. The lengths varied similar to the other experiments. The average CV(RMSE) results were recorded for each length of training data and the results are presented in Figure 62. It can be observed from the figure that the minimal performance error occurs with the maximum amount of training days for this case study.

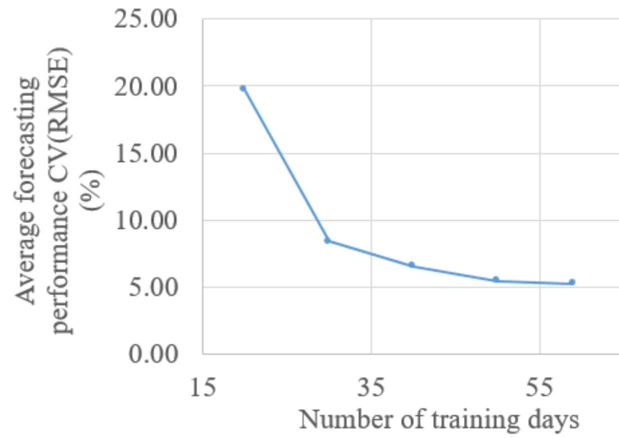

Figure 62: Adjustment of training days for a single point primary system forecasting model applied to the measurement case study

6.3.7 Forecasting model results: Sequential approach

The results of the F+1 forecasts for the sequential approach are presented in this section. The sequential approach forecasts the target variables at 08:45 over a six hour horizon of 09:00 to 14:45 on July 30th, 2014. The output forecasts are presented in Figure 63 which show the output forecasts in black lines and measurement data in red.

The model begins by forecasting the GE secondary system electric demand shown in the top most graph of Figure 63, which then passes the output forecasts to the sub-sequent model forecasting the air-side cooling load (seen in the graph second from the top of Figure 63). The output forecasts are then passed to the next model which forecasts the GE water-side cooling load (second from the bottom) and then the electric demand of the primary system (bottom most graph of Figure 63). Next, the total electric demand of the GE HVAC system is forecasted by the summation of the forecasts for the primary and secondary system.

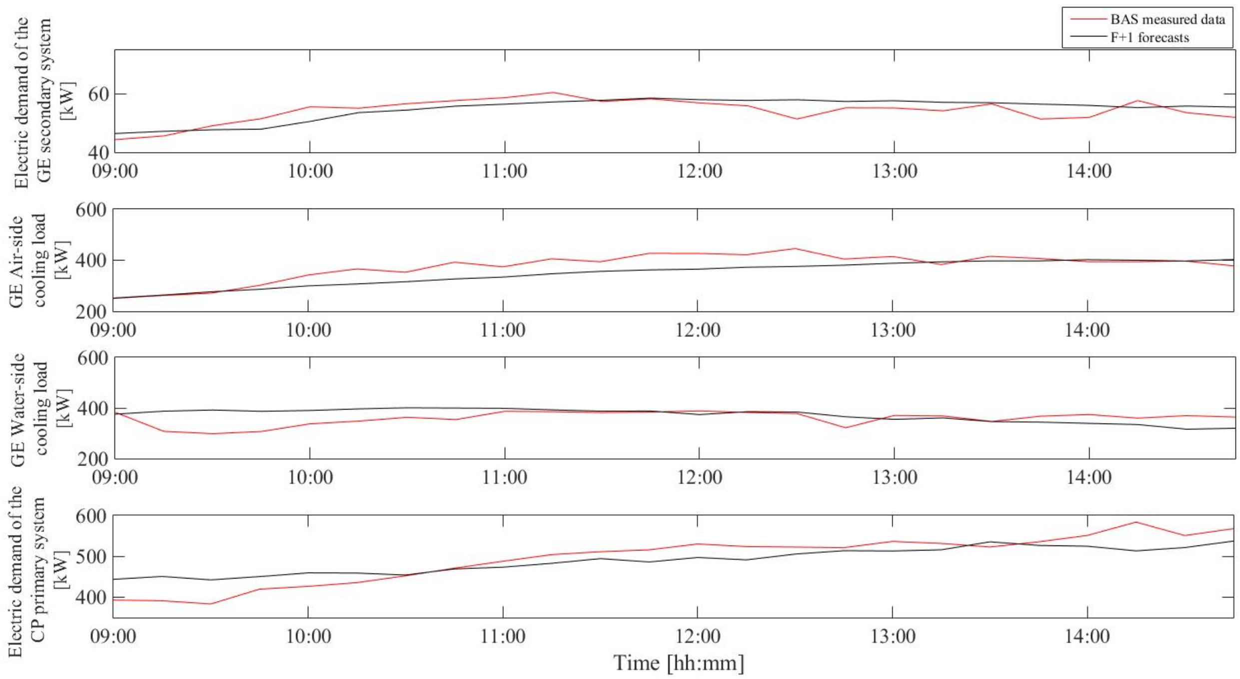

Figure 63: F+1 forecasts for the sequential model applied to the measurement case study

The results of the F+1 forecasts for the sequential approach are presented in Table 52. The forecasts fit well with measurements showing a 5.42% CV(RMSE) for the electric demand of the secondary system and 6.53% CV(RMSE) for the electric demand of the CP primary system. The electric demand of the GE HVAC system shows an error of 3.71% CV(RMSE).

Table 52: F+1 forecasting results for the sequential model applied to the measurement case study

Target variables	**Performance indices**	
	CV(RSME) (%)	RMSE (kW)
Electric demand of the GE secondary system	5.42	2.94
GE Air-side cooling load	10.23	38.45
GE Water-side cooling load	11.37	40.85
Electric demand of the CP primary system	6.53	32.29
Electric demand of the GE HVAC System	3.71	4.21

6.4 Comparison of forecasting model results

Both system based forecasting approaches (monolithic and sequential) were applied to a testing data set of one day of consecutive forecasts starting from 08:45 7/30/2014 and continuing until 08:45 7/31/2014. The hyperparameters used for both approaches remain consistent with those found during the search and are summarized in Table 53.

Table 53: Summary of the hyperparameters for the measurement data case study

System approach	Target variables	Lags [15 min]	AE Architecture	LSTM units
Monolithic	Electric demand of the secondary system Electric demand of the primary system	4	84-75-40	50 to 200
Sequential	Electric demand of the secondary system	4	16-10-5	100 to 200
	Air-side cooling load	4	56-50-45	100 to 200
	Water-side cooling load	6	90-85-40	100 to 200
	Electric demand of the primary system	4	68-65-50	100 to 300

At each time step, 15 minutes, forecasts are generated over the forecast horizon and the performance is calculated comparing estimated values with the measurement values. Retraining in both forecasting approaches occurs every six-hours of the testing set. Upon completion of the models over the testing dataset, the performance is then averaged for both approaches. The results of the system based approaches are presented in Table 54.

The results of this case study demonstrate that the monolithic approach obtains a slightly smaller (less than 1% CV(RMSE)) forecasting error than the sequential approach. The results of this study are consistent with those found in the synthetic data case study which additionally observed a higher error in the sequential approach, though at a larger difference. Overall, both approaches show good performances in forecasting the future electric demands with a performance range of 5.59% to 8.33% CV(RMSE). The performance range observed for the thermal cooling loads was 10.65% to 33.41% CV(RMSE).

Table 54: Forecasting results for the measurement data case study over the testing data set

Target Variable	Monolithic approach		Sequential approach	
	RMSE (kW)	**CV(RMSE) (%)**	**RMSE (kW)**	**CV(RMSE) (%)**
Electric demand of the GE secondary system	2.71	6.24	2.80	6.52
GE Air-side cooling load	N/A	N/A	70.46	33.41
GE Water-side cooling load	N/A	N/A	35.03	10.65
Electric demand of the CP primary system	34.98	7.78	36.22	8.33
Electric demand of the GE HVAC System	5.51	5.59	5.58	5.78

Furthermore, if we use ASHRAE's error of 30% for building energy prediction models as a benchmark for performance; then the models developed in this work demonstrate adequate forecasting performance. This is particularly valid as the model development within this case study, over a forecast horizon, should expect a larger margin of forecasting error. This is due to the fact that more calculations may be required compared to that of a prediction model and the uncertainty of the future becomes greater the further ahead of time a forecast is generated.

6.5 Conclusions of the measurement case study

The work completed in this chapter contributes to the completion of the second objective of this thesis. In this chapter, two system based approaches were applied to target various energy demands of the HVAC system. The overarching goal of both approaches is to target the overall electric demand for the HVAC system. The monolithic approach applied one large forecasting model. The inputs for this model used current and past values for the regressors. Furthermore, multiple target variables are outputted by the model. In contrast, the sequential approach applied multiple comparatively smaller forecasting models coupled together. Each model within the overall approach targeted a specific energy load within the HVAC system. Inputs for the model consisted of current, past, and forecasted values from the upstream model. Both approaches were applied to a case study of an institutional building over the summer cooling period and utilized measurement data obtained from the buildings BAS. The forecast horizon for all models in this work was six-hours ahead at a granularity of 15-minute time steps. The results of both system based approaches demonstrate adequate performance. Furthermore, it was observed that both models obtained similar performances in forecasting the electric demand of the HVAC system and larger errors were observed in forecasting the results of the thermal cooling loads. The results are consistent with those of the synthetic case study.

Chapter 7: Forecasting with off-site weather data

This section explores the application of using off-site weather data applied to a forecasting model calibrated with on-site data. For this work, the off-site weather data was obtained from publically available data provided by Environment Canada and is from the closest available source, Pierre-Trudeau airport, approximately 8 km away from Loyola Campus. The off-site weather data is applied to the monolithic approach developed with on-site measurement data in order to explore the performance effects.

7.1 Objectives

The objectives of this section are to explore the effects of: (i) when weather data from the closest airport is applied as an input in the event of a failure of the on-site weather station, and (ii) when publically available forecasted airport weather data is applied as an input.

7.2 Methods

7.2.1 Forecasting model

This work will applied the monolithic forecasting model tuned in Chapter 6 to forecast the future demand of the HVAC system. The tuned hyperparameters of the model developed in Chapter 6 included: a 4 lagged, 84-75-40 encoder, and 50 to 200 LSTM units. This work will apply the same architecture without a new hyperparameter search. Therefore, even with the application of future forecasted weather information, the architecture will be kept constant.

7.2.2 Data conversion method of hourly to sub-hourly data

The airport data obtained consists of hourly time step data. In contrast, the on-site measurement data of the Genome building is recorded at 15-minute time steps. Therefore, in order to substitute and apply the airport data to the monolithic model, a conversion is required. In order to convert the hourly data into 15-minute time steps; linear interpolation is applied. This is a result of time constraints; however, future work may applied different approaches such as a ML based model.

7.2.3 Correlation between airport weather data to on-site measurement data

Two approaches for applying airport weather data to the GE forecasting model are explored in this research. In the first approach, airport weather data is converted to 15 minute time step data and then applied to the forecasting model. In the second approach, airport weather data is converted to

15 minute time step data, then used to estimate the local GE weather, and then applied to the forecasting model.

In order to estimate local GE weather data based on airport weather data, the conversions developed by Le Cam et al. over the summer period of June to August 2015 [213] are applied. These conversions were derived to convert weather data from Pierre-Trudeau airport to Loyola campus and specifically focus on: the outdoor air temperature, relative humidity and enthalpy. The equations are presented in 7-1 to 7-3 [213]:

$$T_{GE} = 0.95 * T_{airport} + 2.0 \tag{7-1}$$

$$RH_{GE} = 1.09 * RH_{airport} - 18.5 \tag{7-2}$$

$$h_{GE} = 0.73 * h_{airport} + 9.8 \tag{7-3}$$

Within their work, the models developed by Le Cam et al. demonstrated a performance of: (i) 0.91 R^2 and a RMSE of 1.3ºC for the air temperature conversion, (ii) 0.89 R^2 and a RMSE of 5.7% for the relative humidity, and (iii) 0.82 R^2 and a RMSE of 3.5 kJ/kg for the outdoor air enthalpy [213]. This work assumes that the models developed by Le Cam et al. fit well for 2014 and therefore, no modifications to the equations are required.

7.2.4 Forecasting future weather approach

This work explores the application of weather forecasts as an input regressor. However, the case study for this work is 2014. To the best of this author's knowledge; there are no databases which contain historical records sets of weather forecasts. Rather, only databases for historical weather. As a consequence, this work assumes that the future forecasted weather fits well with the future measurement data; therefore, this work will apply future weather measurement values as a proxy to weather forecasts.

7.2.5 Summary of scenarios for the inclusion of off-site weather data

Based on the objectives and the approaches for applying the off-site weather data, multiple studies (scenarios) are generated. The list below provides a summary of the various scenarios applied in this chapter: Furthermore, it was desired to explore the performance of each scenario against a benchmark model. In order to achieve this, the benchmark performances of the monolithic model shown in chapter 6 is selected and is termed benchmark within the list.

Benchmark: Monolithic forecasting model developed in chapter 6

Scenario 1: Use of airport weather data as inputs in place of historical local GE weather data

Scenario 2: Use of estimated local GE weather data, which is obtained via the airport weather data with equations (7.1 to 7.3)

Scenario 3: Use of historical lags of GE measurement data, with the addition of outdoor air temperature forecasts from the on-site GE weather data, as an input

Scenario 4: Use of historical lags of GE measurement data, with the addition of outdoor air temperature forecasts from the airport weather data, as an input

Scenario 5: Use of historical lags of GE measurement data, with the addition of GE weather forecasts obtained via the airport weather data forecasts and equations (7.1 to 7.3)

Scenarios 1 and 2 explore the performance effects from the substitution of off-site weather data applied as an input to the forecasting model calibrated with on-site weather data. Scenarios 3, 4, and 5 explore the effects from the application of forecasted weather data applied as an input to the calibrated forecasting model. For such scenarios, the length of the forecasted weather data applied as an input to the monolithic model is of equal length to the forecast horizon (six hours in advance). It should be noted that both scenarios 3 and 5 are similar; however, they differ in an important aspect. Scenario 3 bases the weather forecasts from the GE weather station, whereas scenario 5 obtains weather forecasts from the airport and is then used to estimate the forecasted weather data for GE using equations (7.1 to 7.3).

7.3 Results

This section presents the results from the application of airport and future forecasted weather data to the monolithic forecasting model. First, a comparison between weather datasets is presented. Figure 64 compares outdoor air temperature from GE, Trudeau airport, and Trudeau airport converted to GE over the week of July 21 to July 27th, 2014.

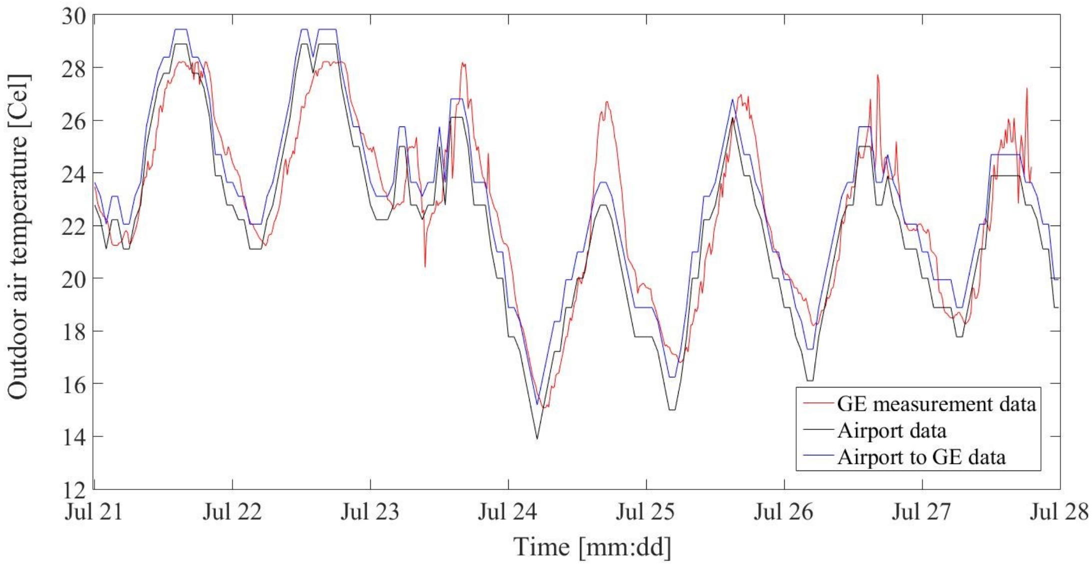

Figure 64: Outdoor air temperature measurements

Comparing airport weather data to GE on-site measurement data from June 1 to August 31st demonstrates an error of 1.88ºC RMSE for the outdoor air temperature and 7.53 kJ/kg RMSE for the outdoor enthalpy. Next, comparing GE on-site weather data with airport weather data converted to GE from June 1st to August 31st demonstrates an RMSE of 1.66ºC for outdoor air temperature and 8.12 kJ/kg RMSE from outdoor air enthalpy. The conversion of relative humidity (equation 7.2) was not applied as it is not an input regressor to the forecasting model.

7.3.1 Substitution of historical weather airport data as inputs to the monolithic model

This section explores the performance effects from the substitution of off-site weather data in the event of a failure of on-site weather sensors. Scenarios 1 and 2 are trained based on GE on-site measurement data from 00:00 June 1st to 8:45 July 30th, 2014 and then applied to forecast 09:00 to 14:45 July 30th, 2014. It was assumed that the weather sensors obtained a failure after 08:30, therefore, the substitution of off-site data is applied at the 08:45 data sample (t). Figure 65 presents the forecasts for both target variables at 09:00 July 30th (F+1). The top graph presents the forecasts for the GE secondary system electric demand, while the bottom graph presents the forecasts for the CP primary system electric demand. The red lines in both graphs indicate BAS measured values while the black lines represent the scenario 0 (baseline forecasting model developed in section 6.3.5). The blue lines show the results from the implementation of scenario 1 (airport data). Furthermore, the green lines presents the forecast with the application of GE data estimated based off of airport weather data (scenario 2).

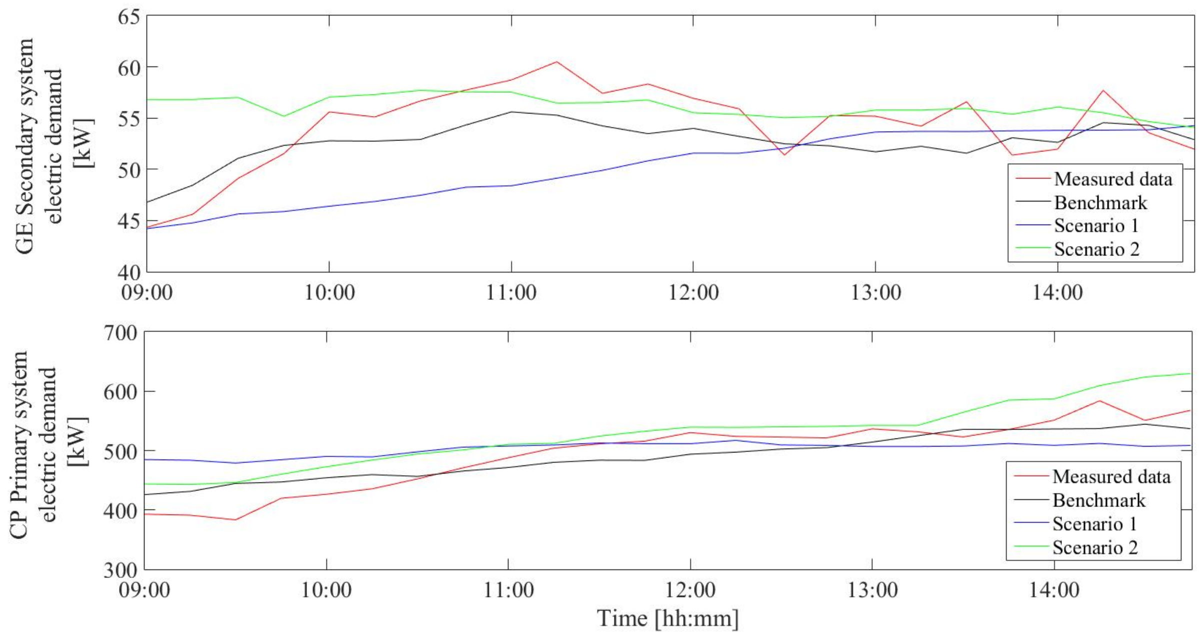

Figure 65: F+1 with substituted weather data

The performances of the F+1 forecasts depicted in Figure 65 are presented in Table 55. The observations demonstrate that the benchmark model obtained the smallest CV(RMSE) with an error of 5.45% for the secondary system and 5.55% for the primary system. Scenario 1, which applied converted airport weather data obtained the largest forecasting errors with 10.73% and 9.81% CV(RMSE) for the electric demand of the secondary and primary systems respectively. Furthermore, scenario 2 which estimated GE weather data based on airport weather data obtained the error of 7.96% and 7.77% CV(RMSE) for the electric demand of the secondary and primary systems.

Table 55: F+1 performance results for scenarios 1 and 2 from 09:00 to 14:45

Scenario	**Forecasting Error CV(RMSE) (%)**		
	GE Secondary System electric demand	**CP Primary System electric demand**	**Electric demand of the GE HVAC System**
Benchmark	5.45	5.55	4.75
1	10.73	9.81	5.72
2	7.96	7.77	7.09

7.3.2 Inclusion of forecasted weather data as inputs for the monolithic model

This section explores the application of the forecasted air temperature data applied as an input to the monolithic model. For this work, it is assumed that the future weather forecasts follow closely with the future measured values. Therefore, this work applies the future weather data as a proxy to forecasted weather data. At each time step, the monolithic forecasting model continues to apply current and historical values based on GE measurement data (t, t-1,…t-3); however, additional forecasted outdoor air temperatures are applied as inputs. The overall hyperparameters of the monolithic forecasting model remain unchanged in this work. Scenarios 3, 4 and 5 are trained based on GE on-site measurement data from 00:00 June 1st to 8:45 July 30th, 2014 and then applied to forecast 09:00 to 14:45 July 30th, 2014.

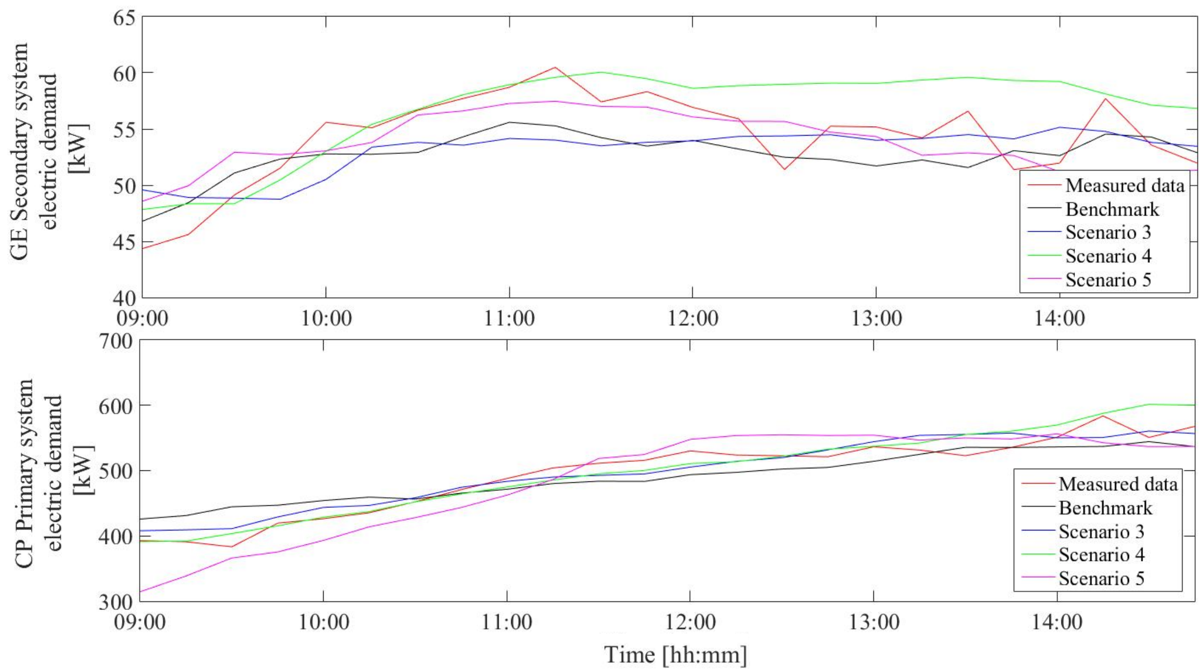

Figure 66: F+1 forecasts with forecasted weather data included as inputs

Figure 66 presents the forecast generated for July 30th at 09:00 to 14:45. The top most graph presents the BAS measured values along with forecasts generated by each scenario for the GE secondary system electric demand. The bottom graph of Figure 66 presents the BAS measured data for the primary system electric demand and the forecasts generated by each scenario. Within both graphs the red lines indicated BAS measured values, the black lines indicate the benchmark forecasts, the blue lines indicate scenario 3 forecasts, the green lines indicate scenario 4 forecasts, and the magenta lines indicate scenario 5 forecasts. The performance results for the F+1 forecasts depicted in Figure 66 are presented in Table 56. The results demonstrate adequate forecasting performance.

Table 56: F+1 performance results for scenarios 3, 4, and 5 from July 30th at 09:00 to 14:45

Scenario	Forecasting Error CV(RMSE) (%)		
	GE Secondary System electric demand	**CP Primary System electric demand**	**Electric demand of the GE HVAC System**
Benchmark	5.45	5.55	4.75
3	5.99	3.49	3.80
4	6.77	3.66	4.45
5	4.91	6.22	3.88

7.3.3 Comparison of scenarios

All scenarios were applied to a testing data set consisting of one day of consecutive forecasts. The data set ranged from 08:45 7/30/2014 and continuing until 08:45 7/31/2014. At each 15-minute time step, forecasts are generated over the forecast horizon and the performance is calculated in comparison to measurement data; the results are presented in Table 57.

Table 57: Forecasting results for each scenario over the testing data set

Scenario	Secondary System electric demand		Primary System electric demand		Electric demand of the GE HVAC System	
	RMSE (kW)	**CV(RMSE) (%)**	**RMSE (kW)**	**CV(RMSE) (%)**	**RMSE (kW)**	**CV(RMSE) (%)**
Benchmark	2.71	6.24	34.98	7.78	5.51	5.59
1	3.58	8.22	39.76	9.06	7.02	7.32
2	3.40	7.91	40.78	9.32	6.68	7.01
3	3.43	7.88	35.10	7.76	6.36	6.46
4	3.10	7.02	32.30	7.24	5.31	5.35
5	3.24	7.54	33.92	7.56	5.99	6.21

Two main points may be noticed from the results. Firstly, in the event of an on-site weather sensor failure; the airport weather data can be substituted into the forecasting model with a minor (less than 2%) decrease in model accuracy. This can be observed in comparing the results of model accuracy for scenarios 1 and 2 to the benchmark accuracy. Secondly, the inclusion of forecasted weather as an input shows a slight reduction in the forecasting of the electric demand. However, this may be attributed to the need to re-tune the hyperparameters of the monolithic model. Despite this, the results from the inclusion of off-site weather data demonstrate adequate forecasting performance over the testing data set.

Chapter 8: Main contributions and future work

8.1 Thesis Contributions

This thesis contributes to the field of demand response research by investigating issues related to short term forecasting. Two different methods were proposed within this thesis for forecasting the electric demand of a heating, ventilation and air conditioning system over a forecast horizon of six-hours: the system model and the component model.

The system based model is a global model which targets forecasting the overall electric demand of an HVAC system. In addition, one system based model provided further forecasts for the thermal energy loads for the HVAC in addition to the electric demand. The system based models provide a tool for the building operators to estimate the various sub-systems within the HVAC which influence the overall electric demand.

The component based model focuses on the future estimation for a component within the overall HVAC system and is based on a grey-box approach. An ANN is applied to forecast a controlled/measured variable within the HVAC system (e.g. supply air flow rate), which is then coupled with a physical model that forecasts the future electric demand of the supply fans.

Based on the objectives of this thesis, the main contributions of this thesis include:

1. A literature review focusing on how ANNs have been applied to forecasting building energy use, and a second literature review focusing on how DL models have been applied for forecasting building energy use.
2. Multi-step ahead forecasting for the electric demand of an HVAC system operating in the cooling/summer period with a forecast horizon of up to six-hours in advance. The data-driven models applied are trained on measurement and synthetic data for a building and consist of state of the art based ML models. The work herein shows the adaptability and performance of such models leveraging existing data from the BAS. The models applied in this work contribute to the overall research field of demand response by the investigation of different approaches for forecasting the future electric demand of large systems within CI buildings.
   I. The system based model forecasts various loads within the HVAC system providing insight for the future electric demand of the sub-systems. Thus, the

system based models facilitate the testing of different demand response based strategies for the overall HVAC system.

II. The component based model estimates the future electric demands for a component within the HVAC system and thus facilitates the testing of different demand response based strategies for a component.

3. Validation of the proposed forecasting models on a case study for an existing building. The systems based model was applied on a case study with two different data sources. Synthetic data for the building obtained from an eQuest simulation for the building and measurement data obtained from the buildings BAS system. The building for this case study consists of the Genomic research center located on Loyola campus at the Concordia University. The component based model has been validated on a component (air handling unit) for the same building with measurement data.
4. A comparison was accomplished exploring the performance effects of applying off-site weather data with a system based forecasting model. Firstly, weather data from Trudeau airport was applied exploring the performance effects of substituting off-site weather data into a calibrated forecasting model. Secondly, this work explored the performance effects of incorporating future weather forecasts as an input.

8.2 Thesis Limitations

The limitations and difficulties of this work is discussed in the following paragraphs. Firstly, the performance of data-driven models is impacted by the quantity and quality of the data obtained. Clearly incorrect sensor measurement values were omitted from the development of the forecasting models within. In addition, sensor values were double checked for accuracy whenever possible (e.g. outdoor air temperature from a local sensor was compared with historical weather data or two sensor measurements at similar locations were compared to each other). However, after such screening and preprocessing steps were applied, it was assumed measurements were correct and that the sensors were reasonably calibrated. Therefore, a limitation of this work was to the amount and quality of the data available.

Furthermore, an additional limitation of this work is with the extraction of data from the synthetic case study. For this work, it was assumed that the eQuest model developed in reference [187], was accurate and did not require significant modifications. Therefore, a limitation of this work included only minor modifications to the eQuest model for the GE building. An example of one such modifications would be the substitution of weather data from 2011 to 2014.

The performance of ML models are significantly impacted by the hyperparameter optimization of the architectures. Hyperparameter selection of the models is crucial to the successful application of the selected models over the task applied. It was desired to explore all such combinations of all different hyperparameters available, and repeat each architecture multiple times to ensure satisfactory observation of its performance. However, in order to accomplish such a task, it would significantly increase the computational load and development time of the forecasting models. As such, time and computational constraints would not allow for such a search. Therefore, heuristics were applied in order to help reduce the search range of the hyperparameters. However, the exploration of hyperparameters beyond the boundary limitations imposed in this work may lead to interesting results and could be worth exploring in future work.

Another limitation to be discussed is with regards to the data conversion methods applied. There are numerous such methods or approaches in order to convert hourly data to sub-hourly data. The exploration of all such methods, their effectiveness of conversion and there effects on performance while important, are beyond the scope of this work. Future work may wish to explore such different techniques and there overall performance effects.

8.3 Future work

Firstly, it should be noted that the forecasting energy models applied within this work may extend beyond the scope of demand response to other energy saving based approaches. For example, other approaches such as: demand side management, fault detection and diagnosis, model predictive control, smart grids, etc. may benefit from the implementation of such models. Nevertheless, the scope of this work was limited to the application of the forecasting models. As such, future work may focus on the implementation of the developed forecasting models outside the field of demand response.

Secondly, the scope of this thesis was limited to the development of the forecasting models. While, the uncertainty of measurement values, sensitivity of the developed forecasting models, and uncertainty of the future estimations is valuable, it was beyond the scope of this work due to time constraints. Future work may benefit from the exploration of such aspects.

Finally, the literature reviews conducted in this work identified many research gaps for the applications of ANN and DL based models and provided lists of such gaps as summaries at the end of sections 2.2 and 2.3. A few of the most prominent research gaps include: lack of grey box models, sub-system and lighting applications, sub-hourly forecast horizons, residential case studies, and long term forecasting. Therefore, future work may wish to explore the effectiveness of ANNs, both shallow and deep models, on the identified research gaps.

References

[1] Independent Electricity System Operator, "Q&A: Demand Response," Independent Electricity System Operator, 20 January 2018. [Online]. Available: http://www.ieso.ca/sector-participants/engagement-initiatives/working-groups/-/media/files/ieso/document-library/working-group/demand-response/drwg-DR_QA.pdf. [Accessed 20 January 2018].

[2] U.S. Energy Information Agency, "Commercial Building Energy Consumption Survery," 2012. [Online]. Available: https://www.eia.gov/consumption/commercial/reports/2012/energyusage/.

[3] Merriam-Webster, "Definition of Forecast," Merriam-Webster, 30 January 2018. [Online]. Available: https://www.merriam-webster.com/dictionary/forecast. [Accessed 30 January 2018].

[4] Merriam-Webster, "Definition of Predict," Merriam-Webster, 10 August 2018. [Online]. Available: https://www.merriam-webster.com/dictionary/predicting. [Accessed 10 August 2018].

[5] Oxford dictionary, "estimate," Oxford dictionary, 30 August 2018. [Online]. Available: https://en.oxforddictionaries.com/definition/estimate. [Accessed 30 August 2018].

[6] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," International Journal of Foreasting, vol. 32, no. 3, pp. 914-938, 2016.

[7] T. Hong and M. Shahidehpour, "Load Forecasting: Case Study," United States Department of Energy, Washington , 2015.

[8] K. Song, Y. Baek, D. Hong and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," IEEE Transaction on Power Systems, vol. 20, no. 1, pp. 96-101, 2005.

[9] American Society of Heating, Refrigerating and Air-Conditioning Engineers, "ASHRAE hanbook: Fundamentals," American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, 2017.

[10] H.-X. Zhao and F. Magoules, "A review on the prediction of building energy consumption," Renewable and Sustainable Energy Reviews, vol. 16, no. 6, pp. 3586-3592, 2012.

[11] L. Breiman, "Statistical Modeling: The Two Cultures," Statistical Science, vol. 16, no. 3, pp. 1999-231, 2001.

[12] N. Fumo and M. Biswas, "Regression analysis for prediction of residential energy consumption," Renewable and Sustainable Energy Reviews, vol. 47, pp. 332-343, 2015.

[13] K. Amber, M. Aslam, A. Mahmood, A. Kousar, M. Younis, B. Akbar, G. Chaudhary and S. Hussain, "Energy Consumption Forecasting for University Sector Buildings," Energies, vol. 10, pp. 2-18, 2017.

[14] X. Li and J. Wen, "Review of building energy modeling for control and operation," Renewable and Sustainable Energy Reviews, vol. 37, pp. 517-537, 2014.

[15] B. Arregi and R. Garay, "Regression analysis of energy consumption of teritary buildings," Energy Procedia, vol. 122, pp. 9-14, 2017.

[16] R. Edwards, J. New and L. Parker, " Predicting future hourly residential electrical consumption: A machine learning case study," Energy and Buildings, vol. 49, pp. 591-603, 2012.

[17] J. Massana, C. Pous, L. Burgas, J. Melendez and J. Colomer, "Short-term load forecasting in a non-residential building contrasting models and attributes," Energy and Buildings, vol. 92, pp. 322-330, 2015.

[18] R. Walger da Fonseca, E. Didone and R. Pereira, "Using artificial neural networks to predict the impact of daylighting on building final electric energy requirements," Energy and Buildings, vol. 61, pp. 31-38, 2013.

[19] J. Gooijer and R. Hyndman, "25 years of time series forecasting," International journal of forecasting, vol. 22, pp. 443-473, 2006.

[20] C. Deb, F. Zhang, J. Yang, S. Lee and K. Shah, "A review on time series forecasting techniques for building energy consumption," Renewable and Sustainable Energy Reviews, vol. 74, pp. 902-924, 2017.

[21] J. Grant, M. Eltoukhy and S. Asfour, "Short-term electrical peak demand forecasting in a large government building using aritifical neural networks," energies, vol. 7, pp. 1935-1953, 2014.

[22] G. Newsham and B. Birt, "Building-level occupancy data to improve ARIMA-based electricity use forecasts," in ACM SenSys 2010, Switzerland, 2010.

[23] R. Kalman, "A new approach to linear filtering and prediction problems," Journal of Basic Engineering, vol. 82, no. 1, pp. 35-45, 1960.

[24] Z. Wang and R. Srinivasan, "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models," Renewable and Sustainable Energy Reviews, vol. 75, pp. 796-808, 2017.

[25] A. Ahmad, M. Hassan, M. Abdullah, H. Rahman, F. Hussin, H. Abdullah and R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," Renewable and Sustainable Energy Reviews, vol. 33, pp. 102-109, 2014.

[26] M. Daut, M. Hassan, H. Abdullas, H. Rahman, M. Abdullah and F. Hussin, "Building electrical energy consumption forecasting analysis using conventional and artificial

intelligence methods: A review," Renewable and Sustainable Energy Reviews, vol. 70, pp. 1108-1118, 2017.

[27] A. Dounis, "Artificial intelligence for energy conservation in buildings," Advances in Building Energy Research, vol. 4, pp. 267-299, 2010.

[28] T. Mitchell, Machine Learning, New York City: McGraw, 1997.

[29] V. Vapnik, The nature of statistical learning theory, New York: Springer, 1995.

[30] Q. Li, Q. Meng, J. Cai, H. Yoshino and A. Mochida, "Applying support vector machine to predict hourly cooling load in the building," Applied Energy, vol. 86, pp. 2249-2256, 2009.

[31] Q. Li, Q. Meng, J. Cai, H. Yoshino and A. Mochida, "Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks," Energy Conversion and Management , vol. 50, pp. 90-96, 2009.

[32] M. Le Cam, R. Zmeureanu and A. Daoud, "Cascade-based short-term forecasting method of the electric demand of HVAC system," Energy, vol. 119, no. 1, pp. 1098-1107, 2016.

[33] Y. Chae, R. Horesh, Y. Hwang and Y. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," Energy and Buildings, vol. 111, pp. 184-194, 2016.

[34] Z. Wang, Y. Wang, R. Zeng, R. Srinivasan and S. Ahrentzen, "Random Forest based hourly building energy prediction," Energy and Buildings, vol. 171, pp. 11-25, 2018.

[35] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," The Bulletin of mathematical biophysics, vol. 5, pp. 115-133, 1943.

[36] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," Psychological Review, vol. 65, no. 6, pp. 386-408, 1958.

[37] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," nature, vol. 521, pp. 436-444, 2015.

[38] Y. Bengio, "Learning Deep Architectures for AI," Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1-127, 2009.

[39] Y. Wei, X. Zhang, Y. Shia, L. Xia, S. Pan, J. Wu, M. Han and X. Zhao, "A review of data-driven approaches for prediction and classification of building energy consumption," Renewable and Sustainable Energy Reviews, vol. 82, p. 1027–1047, 2018.

[40] R. Kumar, R. Aggarwal and J. Sharma, "Energy analysis of a building using artificial neural networks: A review," Energy and Buildingss, vol. 65, pp. 352-358, 2013.

[41] K. Amasyali and N. El-Gohary, "A review of data driven building energy consumption prediction studies," Renewable and Sustainable Energy Reviews, pp. 1192-1205, 2018.

[42] T. Ahmad, H. Chen, Y. Guo and J. Wang, "A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review," Energy and Buildings , vol. 165, pp. 301-320, 2018.

[43] M. Bourdeau, X. Zhai, E. Nefzaoui, X. Guo and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data driven techniques," Sustainable Cities and Society, vol. 48, 2019.

[44] S. Mohandes, X. Zhang and A. Mahdiyar, "A comprehensive review on the application of artificial neural networks in building energy analysis," Neurocomputing, vol. 340, pp. 55-75, 2019.

[45] Y. Sun, F. Haghighat and B. Fung, "A review of the-state-of-the-art in data-driven approaches for building energy prediction," Energy and Buildings, vol. 221, no. 110022, 2020.

[46] J. Runge and R. Zmeureanu, "Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review," Energies, vol. 12, no. 17, 2019.

[47] K. Powell, A. Sriprasad, W. Cole and T. Edgar, "Heating, cooling and electrical load forecasting for a large scale district energy systems," Energy, vol. 74, pp. 877-885, 2014.

[48] C. Deb, L. Eang, J. Yang and M. Santamouris, "Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks," Energy and Buildings, vol. 121, pp. 287-297, 2016.

[49] Y. Penya, C. Borges and I. Fernandez, "Short-term load forecasting in non-residential buildings," in IEEE Africon, Livingstone, 2011.

[50] R. Mena, F. Rodriguez, M. Castilla and M. Arahal, "A prediction model based on neural networks for the energy consumption of a bioclimatic building," Energy and Buildings, vol. 82, pp. 142-155, 2014.

[51] M. Sakawa, K. Kato and S. Ushiro, "Cooling load prediction in a district heating and cooling system through simplified robust filter and multilayered neural network," Applied Artificial Intelligence, vol. 15, pp. 633-643, 2010.

[52] A. Ben-Nakhi and M. Mahoud, "Cooling load prediction for buildings using general regression neural networks," Energy converion and management, vol. 45, pp. 2127-2141, 2004.

[53] S. Karatasou, M. Santamouris and V. Geros, "Modeling and predicting building's energy use with artificial neural networks: Methods and results," Energy and Buildings, vol. 38, pp. 949-958, 2006.

[54] P. Gonzalez and J. Zamarreno, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," Energy and Buildings, vol. 37, pp. 595-601, 2005.

[55] I. Fernandez, C. Borges and Y. Penya, "Efficient building load forecasting," in Emerging Technologies & Factory Automation, Toulouse, 2011.

[56] G. Escriva-Escriva, C. Alvarez-Bel, C. Roldan-Blay and M. Ortega, "New artificial neural network prediction method for electrical consumption forecasting based on building end-uses," Energy and Buildings, vol. 43, no. 11, pp. 3112-3119, 2011.

[57] V. Kamaev, M. Shcherbakov, D. Panchenko, N. Shcherbakova and A. Brebels, "Using connectionist systems for electric energy consumption forecasting in shopping centers," Automation and remote control, vol. 75, no. 6, pp. 1075-1084, 2012.

[58] A. Bagnasco, F. Fresi, M. Saviozzi, F. Silvestro and A. Vinci, "Electrical consumption forecasting in hospital facilities: An application case," Energy and Buildings, vol. 103, pp. 261-270, 2015.

[59] C. Fan, F. Xiao and S. Wang, "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques," Applied Energy, vol. 127, pp. 1-10, 2014.

[60] Y. Guo, J. Wang, H. Chen, G. Li, J. Liu, C. Xu, R. Huang and Y. Huang, "Machine learning-based thermal response time ahead energy demand prediction for building heating systems," Applied Energy, vol. 221, no. 1, pp. 16-27, 2018.

[61] S. Farzana, M. Liu, A. Baldwin and U. Hossain, "Multi-model prediction and simulation of residential building energy in urban areas of Chongqing, South West China," Energy and Buildings, vol. 81, pp. 161-169, 2014.

[62] K. Yun, R. Luck, P. Mago and H. Cho, "Building hourly thermal load prediction using an indexed ARX model," Energy and Buildings, vol. 54, pp. 225-233, 2012.

[63] C. Roldan-Blay, G. Escriva-Escriva, C. Alvarez-Bel, C. Roldan-Porta and J. Rodriguez-Garcia, "Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model," Energy and Buildings, vol. 60, pp. 38-46, 2013.

[64] J. G. Jetcheva, M. Majidpour and W.-P. Chen, "Neural network model ensembles for building-level electricity load forecasts," Energy and Buildings, vol. 84, pp. 214-223, 1 12 2014.

[65] Y. K. Penya, C. E. Borges, D. Agote and I. Fernández, "Short-term load forecasting in air-conditioned non-residential Buildings," in IEEE International Symposium on Industrial Electronics, Gdańsk, 2011.

[66] S. Paudel, M. Elmtiri, W. L. Kling, O. L. Corre and B. Lacarrière, "Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network," Energy and Buildings, vol. 70, pp. 81-93, 1 2 2014.

[67] M. Le Cam, A. Daoud and R. Zmeureanu, "Forecasting electric demand of supply fan using data mining techniques," Energy, vol. 101, pp. 541-557, 2016.

[68] S. Jurado, À. Nebot, F. Mugica and N. Avellana, "Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques," Energy, vol. 86, pp. 276-291, 15 6 2015.

[69] L. G. Ruiz, R. Rueda, M. P. Cuéllar and M. C. Pegalajar, "Energy consumption forecasting based on Elman neural networks with evolutive optimization," Expert Systems with Applications, vol. 92, pp. 380-389, 2018.

[70] L. Ruiz, M. Cuéllar, M. Calvo-Flores and M. Jiménez, "An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings," Energies, vol. 9, no. 684, 2016.

[71] A. Tascikaraoglu and B. M. Sanandaji, "Short-term residential electric load forecasting: A compressive spatio-temporal approach," Energy and Buildings, vol. 111, pp. 380-392, 1 1 2016.

[72] A. Garnier, J. Eynard, M. Caussanel and S. Grieu, "Predictive control of multizone heating, ventilation and air-conditioning systems in non-residential buildings," Applied Soft Computing, vol. 37, pp. 847-862, 1 12 2015.

[73] Y. Liu, W. Wang and N. Ghadimi, "Electricity load forecasting by an improved forecast engine for building level consumers," Energy, vol. 139, pp. 18-30, 2017.

[74] B. Dong, Z. Li, S. M. Rahman and R. Vega, "A hybrid model approach for forecasting future residential electricity consumption," Energy and Buildings, vol. 117, pp. 341-351, 2016.

[75] A. Kusiak and G. Xu, "Modeling and optimization of HVAC systems using a dynamic neural network," Energy, vol. 42, no. 1, pp. 241-250, 1 6 2012.

[76] D. Srinivasan, "Energy demand prediction using GMDH networks," Neurocomputing, vol. 72, no. 1-3, pp. 625-629, 2008.

[77] M. Le Cam, R. Zmeureanu and A. Daoud, "Comparison of inverse models used for the forecast of the electric demand of chillers," in Conference of International Building Performance Simulation Association, Chambéry, 2013.

[78] B. Yildiz, J. I. Bilbao and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," Renewable and Sustainable Energy Reviews, vol. 73, pp. 1104-1122, 2017.

[79] A. S. A. Rahman, "Predicting fuel consumption for commercial buildings with machinelearning algorithms," Energy and buildings, vol. 152, pp. 341-358, 2017.

[80] R. Yokoyama, T. Wakui and R. Satake, "Prediction of energy demands using neural network with model identification by global optimization," Energy Conversion and Management, vol. 50, no. 2, pp. 319-327, 1 2 2009.

[81] C. Deb, L. Eang, J. Yang and M. Santamouris, "Forecasting Energy Consumption of Institutional Buildings in Singapore," Procedia Engineering, vol. 121, pp. 1734-1740, 2015.

[82] H. Son and C. Kim, "Forecasting Short-term Electricity Demand in Residential Sector Based on Support Vector Regression and Fuzzy-rough Feature Selection with Particle Swarm Optimization," Procedia Engineering, vol. 118, pp. 1162-1168, 2015.

[83] Z. Hou, Z. Lian, Y. Yao and X. Yuan, "Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique," Applied Energy, vol. 83, no. 9, pp. 1033-1046, 2006.

[84] R. Platon and V. R. M. J. Dehkordi, "Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis," Energy and Buildings, vol. 92, pp. 10-18, 2015.

[85] V. Arabzadeh, B. Alimohammadisagvand, J. Jokisalo and K. Siren, "A novel cost-optimizing demand response control for a heat pump heated residential building," BUILD SIMUL, vol. 11, pp. 533-547, 2018.

[86] S. Ferlito, M. Atrigna, G. Graditi, S. De Vito, M. Salvato, A. Buonanno and G. Di Francia, "Predictive models for building's energy consumption: an Artificial Neural Network (ANN) approach," in XVIII AISEM Annual Conference, Trento, 2015.

[87] T. Ahmad and H. Chen, "Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches," Energy and Buildings, vol. 166, pp. 460-476, 1 5 2018.

[88] G.-C. Liao, "Hybrid Improved Differential Evolution and Wavelet Neural Network with load forecasting problem of air conditioning," International Journal of Electrical Power & Energy Systems, vol. 61, pp. 673-682, 1 10 2014.

[89] K. Kato, M. Sakawa, K. Ishimaru, S. Ushiro and T. Shibano, "Heat Load Prediction through Recurrent Neural Network in District Heating and Cooling Systems," in IEEE International Conference on Systems, Man and Cybernetics, Suntec, 2008.

[90] H. Chitsaz, H. Shaker, H. Zareipour, D. Wood and N. Amjady, "Short-term electricity load forecasting of buildings in microgrids," Energy and Buildings, vol. 99, pp. 50-60, 15 7 2015.

[91] L. Wang, E. W. Lee and R. K. Yuen, "Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach," Applied Energy, vol. 228, pp. 1740-1753, 15 10 2018.

[92] N. K. Thokala, A. Bapna and M. G. Chandra, "A deployable electrical load forecasting solution for commercial buildings," in Proceedings of the IEEE International Conference on Industrial Technology, Lyon, 2018.

[93] D. Chakraborty and H. Elzarka, "Advanced machine learning techniques for building performance simulation: a comparative analysis," Journal of Building Performance Simulation, 2018.

[94] A. Kialashaki and J. R. Reisel, "Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks," Applied Energy, vol. 108, pp. 271-280, 1 8 2013.

[95] A. Kusiak, G. Xu and F. Tang, "Optimization of an HVAC system with a strength multi-objective particle-swarm algorithm," Energy, vol. 36, no. 10, pp. 5935-5943, 1 10 2011.

[96] Z. Li, B. Dong and R. Vega, "A Hybrid Model for Electrical Load Forecasting-a New Approach Integrating Data-Mining with Physics-Based Models," ASHRAE Transactions, vol. 121, pp. 1-9, 2015.

[97] M. Rahman, B. Dong and R. Vega, "Machine Learning Approach Applied in Electricity Load Forecasting: Within Residential Houses Context," ASHRAE Transactions, vol. 121, no. 1-8, 2015.

[98] Y. Yao, Z. Lian, S. Liu and Z. Hou, "Hourly cooling load prediction by a combined forecasting model based on Analytic Heirarchy Process," International Journal of Thermal Sciences, vol. 43, pp. 1107-1118, 2004.

[99] A. Kusiak, G. Xu and Z. Zhang, "Minimization of energy consumption in HVAC systems with data-driven models and an interior-point method," Energy Conversion and Management, vol. 85, pp. 146-153, 2014.

[100] A. Kusiak, Y. Z. Zeng and G. Xu, "Minimizing energy consumption of an air handling unit with a computational intelligence approach," Energy and Buildings, vol. 60, pp. 355-363, 2013.

[101] Y. Zeng, Z. Zhang, A. Kusiak, F. Tang and X. Wei, "Optimizing wastewater pumping system with data-driven models and a greedy electromagnetism like algorithm," Stochastic Environmental Research and Risk Assessment, vol. 30, pp. 1263-1275, 2016.

[102] X. He, Z. Zhang and A. Kusiak, "Performance optimization of HVAC systems with computational intelligence algorithm," Energy and Buildings, vol. 81, pp. 371-380, 2014.

[103] Y. Zeng, Z. Zhang and A. Kusiak, "Predictive modeling and optimization of a multi-zone HVAC system with data mining and firefly algorithms," Energy, vol. 86, pp. 393-402, 2015.

[104] A. Kusiak and M. Li, "Reheat optimization of the variable air volume box," Energy, vol. 35, pp. 1997-2005, 2010.

[105] B. Yan, A. Malkawi and Y. K. Yi, "Case study of applying different energy use modeling methods to an existing building," in Conference of International Building Performance Simulation Association, Sydney, 2011.

[106] M. Macas, F. Moretti, A. Fonti, A. Giantomassi, G. Comodi, M. Annunziato, S. Pizzuti and A. Capra, "The role of data sample size and dimensionality in neural network based forecasting of building heating related variables," Energy and Buildings, vol. 111, pp. 299-310, 2016.

[107] D.-S. Kapetanakis, D. Christantoni, E. Mangina and D. Finn, "Evaluation of machine learning algorithms for demand response potential forecasting," in International Conference on Building Simulation , San Francisco , 2017.

[108] E. Khalil, A. Medhat, S. Morkos and M. Salem, "Neural networks approach for energy consumption in air-conditioned administrative building," in ASHRAE Transactions , 2012.

[109] T. Ahmad, H. Chen, J. Shair and C. Xu, "Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management," International Journal of Refrigeration, vol. 98, pp. 399-409, 2019.

[110] L. Xu, S. Wang and R. Tang, "Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load," Applied Energy, vol. 237, pp. 180-195, 2019.

[111] J. Chou and D. Tran, "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders," Energy, vol. 165, pp. 709-726, 2018.

[112] C. Tian, C. Li, G. Zhang and Y. Lv, "Data driven parallel prediction of building energy consumption using generative adversarial nets," Energy and Buildings, vol. 186, pp. 230-243, 2019.

[113] M. Katsatos and K. Mourtis, "Application of Artificial Neuron Networks as energy consumption forecasting tool in the building of Regulatory Authority of Energy, Athens, Greece," Energy Procedia, vol. 157, pp. 851-861, 2019.

[114] Z. Xuan, Z. Xuehui, L. Liequan, F. Zubing, Y. Junwei and P. Dongmei, "Forecasting performance comparison of two hybrid machine learning models for cooling load of a large-scale commercial building," Journal of Building Engineering, vol. 21, pp. 64-73, 2019.

[115] M. De Felice and X. Yao, "Neural networks ensembles for short-term load forecasting," in IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG), Paris, 2011.

[116] D. Lui and Q. Chen, "Prediction of building lighting energy consumption based on support vector regression," in 9th Asian Control Conference (ASCC), Istanbul, 2013.

[117] M. Gulin, M. Vasak, G. Banjac and T. Tomisa, "Load forecast of a university building for application in microgrid power flow optimization," in IEEE International Energy Conference (ENERGYCON), Cavtat, 2014.

[118] A. Ebrahim and O. Mohammed, "Household Load Forecasting Based on a Pre-Processing Non-Intrusive Load Monitoring Techniques," in IEEE Green Technologies Conference (GreenTech) , Lafayette, 2018.

[119] B. Yuce, M. Mourshed and Y. Rezgui, "A smart forecasting approach to district energy management," Energies, vol. 10, no. 8, 2017.

[120] A. Azadeh, S. Ghaderi and S. Sohrabkhani, "Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors," Energy Conversion and Management, vol. 49, pp. 2272-2278, 2008.

[121] X. Li, J. Wen and E. W. Bai, "Developing a whole building cooling energy forecasting model for on-line operation optimization using proactive system identification," Applied Energy, vol. 164, pp. 69-88, 2016.

[122] B. Gunay, W. Shen and G. Newsham, "Inverse blackbox modeling of the heating and cooling load in office buildings," Energy and Buildings, vol. 142, pp. 200-210, 1 5 2017.

[123] D. Geysen, O. De Somer, C. Johansson, J. Brage and D. Vanhoudt, "Operational thermal load forecasting in district heating networks using machine learning and expert advice," Energy and Buildings, vol. 162, pp. 144-153, 1 3 2018.

[124] J. Schachter and P. Mancarella, "A short-term load forecasting model for demand response applications," in 11th International Conference on the European Energy Market (EEM14), Krakow, 2014.

[125] K. Amarasinghe, D. L. Marino and M. Manic, "Deep neural networks for energy load forecasting," in 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, 2017.

[126] C. Fan, F. Xiao and Y. Zhao, "A short-term building cooling load prediction method using deep learning algorithms," Applied Energy, pp. 222-233, 2017.

[127] G. Fu, "Deep belief network based ensemble approach for cooling load forecasting of air-conditioning system," Energy, vol. 148, pp. 269-282, 2018.

[128] E. Mocanu, P. Nguyen, M. Gibescu and W. Kling, "Deep learning for estimating building energy consumption," Sustainable Energy, Grids and Network, vol. 6, pp. 91-99, 2016.

[129] A. Rahman, V. Srikumar and A. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," Applied Energy, vol. 212, pp. 372-385, 2018.

[130] C. Li, Z. Ding, D. Zhao, J. Y. Yi and G. Zhang, "Building energy consumption prediction: an extreme deep learning approach," energies, vol. 1525, 2017.

[131] D. Koschwitz, J. Frisch and C. v. Treeck, "Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale," Energy, vol. 165, pp. 134-142, 2018.

[132] A. Rahman and A. Smith, "Predicting heating demand and sizing a stratified thermal storage tank using deep learning algorithms," Applied Energy, vol. 228, pp. 108-121, 2018.

[133] J. Reynolds, M. W. Ahmad, Y. Rezgui and J.-L. Hippolyte, "Operational supply and demand optimisation of a multi-vector district energy system using artificial neural networks and a genetic algorithm," Applied Energy, vol. 235, pp. 699-713, 2019.

[134] R. Hribar, P. Potocnik, J. Silc and G. Papa, "A comparison of models for forecasting the residential natural gas," Energy, vol. 167, pp. 511-522, 2019.

[135] M. Cai, M. Pipattanasompom and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techinques," Applied Energy, pp. 1078-1088, 2019.

[136] C. Fan, Y. Sun, Y. Zhao, M. Song and J. Wang, "Deep learning-based feature engineering methods for improved building energy prediction," Applied Energy, vol. 240, pp. 35-45, 2019.

[137] F. Tang, A. Kusiak and X. Wei, "Modeling and short-term prediction of HVAC system with a clustering algorithm," Energy and Buildings, vol. 82, pp. 310-321, 2014.

[138] Hydro Quebec, "Rates for business customers," 19 April 2019. [Online]. Available: http://www.hydroquebec.com/business/customer-space/rates/. [Accessed 19 APril 2019].

[139] ASHRAE, "Guildeline 14-2002, Measurement of Energy and Demand Savings," American Society of Heating, Ventilation, and Air Conditioning Engineers, Atlanta, 2002.

[140] J. Runge and R. Zmeureanu, "A Review of Deep Learning Techniques for Forecasting Energy Use in Buildings," Energies, vol. 14(3), no. 608, 2021.

[141] I. Goodfellow, Y. Bengio and A. Courville, Deep learning, Cambridge: MIT Press, 2016.

[142] O. Laib, M. Khadir and L. Mihaylova, "Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks," Energy, vol. 177, pp. 530-542, 2019.

[143] H. Su, E. Zio, J. Zhang, M. Xu, X. Li and Z. Zhang, "A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model," Energy, vol. 178, pp. 585-597, 2019.

[144] G. Suryanarayana, J. Lago, D. Geysen, P. Aleksiejuk and C. Johansson, "Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods," Energy, vol. 157, pp. 141-149, 2018.

[145] P. Xue, Y. Jiang, Z. Zhou, X. Chen, X. Fang and J. Liu, "Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms," Energy, vol. 188, 2019.

[146] M. Gong, H. Zhou, Q. Wang, S. Wang and P. Yang, "District heating systems load forecasting: a deep neural networks model based on similar day approach," Advances in Building Energy Research, 2019.

[147] G. Xue, Y. Pan, T. Lin, J. Song, C. Qi and Z. Wang, "District Heating Load Prediction Algorithm Based on Feature Fusion LSTM Model," Energies, vol. 12, no. 2122, 2019.

[148] G. Xue, C. Qi, H. Li, X. Kong and J. Song, "Heating load prediction based on attention long short term memory: A case study of Xingtai," Energy, vol. 203, no. 117846, 2020.

[149] N. Wei, C. Li, X. Peng, Y. Li and F. Zeng, "Daily natural gas consumption forecasting via the application of a novel hybrid model," Applied Energy, vol. 250, pp. 358-368, 2019.

[150] M. Son, J. Moon, S. Jung and E. Hwang, "A Short-Term Load Forecasting Scheme Based on Auto-Encoder and Random Forest," in Applied Physics, System Science and Computers III: Proceedings of the 3rd International Conference on Applied Physics, System Science and Computers , Dubrovnik, Springer, 2018, pp. 138-144.

[151] J. Kim, J. Moon, E. Hwang and P. Kang, "Recurrent inception convolution neural network for multi short-term load forecasting," Energy and Buildings, vol. 194, pp. 328-341, 2019.

[152] S. Wang, S. Wang, H. Chen and Q. Gu, "Multi-energy load forecasting for regional integrated energy systems considering temporal dynamic and coupling characteristics," Energy, vol. 195, no. 116964, 2020.

[153] H. Dagdougui, F. Bagheri, H. Le and L. Dessaint, "Neural network model for short-term and very-short-term load forecasting in district buildings," Energy and Buildings, vol. 203, 2019.

[154] K. Zor, O. Celik, O. Timur and A. Teke, "Short-Term Building Electrical Energy Consumption Forecasting by Employing Gene Expression Programming and GMDH Networks," Energies, vol. 13, no. 5, 2020.

[155] S. Idowu, S. Saguna, C. Ahlund and O. Schelen, "Applied machine learning: Forecasting heat load in district heating system," Energy and Buildings, vol. 133, pp. 478-488, 2016.

[156] A. Vesa, N. Ghitescu, C. Pop, M. Antal, T. Cioara, I. Anghel and I. Salomie, "A Stacking Multi-Learning Ensemble Model for Predicting Near Real Time Energy Consumption Demand of Residential Buildings," in IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, 2019.

[157] W. Kong, Y. Dong, Y. Jia, D. Hill, Y. Xu and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841-851, 2019.

[158] C. Fan, J. Wang, W. Gang and S. Li, "Assessment of deep recurrent neural network-based strategies for short-term building energy predictions," Applied Energy, vol. 236, pp. 700-710, 2019.

[159] Z. Wang, T. Hong and M. Piette, "Building thermal load prediction through shallow machine learning and deep learning," Applied Energy, vol. 263, no. 114683, 2020.

[160] J. Yang, K. Tan, M. Santamouris and S. Lee, "Building Energy Consumption Raw Data Forecasting Using Data Cleaning and Deep Recurrent Neural Networks," Buildings, vol. 9, 2019.

[161] N. Somu, G. Raman and K. Ramamritham, "A hybrid model for building energy consumption forecasting using long short term memory networks," Applied Energy, vol. 261, 2020.

[162] C. Nichiforov, G. Stamatescu, I. Stamatescu, V. Calofir, I. Fagarasan and S. Iliescu, "Deep Learning Techniques for Load Forecasting in Large Commercial Buildings," in 22nd International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, 2018.

[163] F. Bunning, P. Heer, R. Smith and J. Lygeros, "Improved day ahead heating demand forecasting by online correction methods," Energy and Buildings, vol. 211, no. 109821, 2020.

[164] C. Zhang, J. Li, Y. Zhao, T. Li, Q. Chen and X. Zhang, "A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process," Energy and Buildings, vol. 225, no. 110301, 2020.

[165] Z. Yong, Y. Xiu, F. Chen, C. Pengfei, C. Binchao and L. Taijie, "Short-term building load forecasting based on similar day selection and LSTM network," in IEEE Conference on Energy Internet and Energy System Integration, BeiJing, 2018.

[166] L. Sehovac, C. Nesen and K. Grolinger, "Forecasting Building Energy Consumption with Deep Learning: A Sequence to Sequence Approach," in IEEE International Congress on Internet of Things (ICIOT), Milan, 2019.

[167] J. Moon, S. Jung, J. Rew, S. Rho and E. Hwang, "Combination of short-term load forecasting models based on a stacking ensemble approach," Energy & Buildings, vol. 216, no. 109921, 2020.

[168] E. Skomski, J. Lee, W. Kim, V. Chandan, S. Katipamula and B. Hutchinson, "Sequence-to-sequence neural networks for short-term electrical load forecasting in commercial office buildings," Energy and Buildings, vol. 226, no. 110350, 2020.

[169] Y. Gao, Y. Ruan, C. Fang and S. Yin, "Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data," Energy and Buildings, vol. 223, no. 110156, 2020.

[170] G. Zhang, C. Tian, C. Li, J. Zhang and W. Zuo, "Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature," Energy, no. 117531, 2020.

[171] X. Luo, L. Oyedele, A. Ajayi, O. Akinade, H. Owolabi and A. Ahmed, "Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings," Renewable and Sustainable Energy Reviews, vol. 131, no. 109980, 2020.

[172] T. Kim and S. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," Energy, vol. 182, pp. 72-81, 2019.

[173] A. Estebsari and R. Rajabi, "Single Residential Load Forecasting Using Deep Learning and Image Encoding Techniques," Electronics, vol. 9, no. 68, 2020.

[174] D. Marino, K. Amarasinghe and M. Manic, "Building Energy Load Forecasting using Deep Neural Networks," in IEEE Industrial Electronics Society , Florence, 2016 .

[175] M. A. Shah, I. A. Sajjad, M. F. N. Khan, M. M. Iqbal, R. Liaqat and M. Shah, "Short-term Meter Level Load Forecasting of Residential Customers Based on Smart Meter's Data," in International Conference on Engineering and Emerging Technologies (ICEET), Lahore, 2020.

[176] T. Hossen, A. Nair, R. Chinnathambi and R. P., "Residential Load Forecasting Using Deep Neural Networks (DNN)," in 2018 North American Power Symposium (NAPS), Fargo, 2018.

[177] E. Choi, S. Cho and D. Kim, "Power Demand Forecasting using Long Short-Term Memory (LSTM) Deep-Learning Model for Monitoring Energy Sustainability," Sustainability, vol. 12, no. 3, 2020.

[178] S. Ryu, J. Noh and H. Kim, "Deep Neural Network Based Demand Side Short Term Load Forecasting," Energies, vol. 10, no. 3, 2017.

[179] R. Jiao, T. Zhang, Y. Jiang and H. He, "Short-Term Non-Residential Load Forecasting Based on Multiple Sequences LSTM Recurrent Neural Network," IEEE Access, vol. 6, pp. 59438-59448, 2018.

[180] C. Li, Z. Ding, J. Yi, Y. Lv and G. Zhang, "Deep Belief Network Based Hybrid Model for Building Energy Consumption Prediction," Energies, vol. 11, no. 242, 2018.

[181] S. Shan, B. Cao and Z. Wu, "Forecasting the Short-Term Electricity Consumption of Building Using a Novel Ensemble Model," IEEE Access, vol. 7, pp. 88093-88106, 2019.

[182] T. Liu, C. Xu, Y. Guo and H. Chen, "A novel deep reinforcement learning based methodology for short-term HVAC system energy consumption prediction," International Journal of Refrigeration, vol. 107, pp. 39-51, 2019.

[183] M. Imani and H. Ghassemian, "Residential load forecasting using wavelet and collaborative representation transforms," Applied Energy, vol. 253, 2019.

[184] T. Liu, Z. Tan, C. Xu, H. Chen and Z. Li, "Study on deep reinforcement learning techniques for building energy consumption forecasting," Energy and Buildings, vol. 208, 2020.

[185] J. Wang, Y. Du and J. Wang, "LSTM based long-term energy consumption prediction with periodicity," Energy, vol. 197, 2020.

[186] S. Makridakis, S. Wheelwright and R. Hyndman, Forecasting Methods and Applications, New Jersey: Wiley & Sons, 2003.

[187] A. Mihai, "Calibration of a Building Energy Model Using Measured Data for a Research," Concordia University, Montreal, QC, 2014.

[188] M. Le Cam, Short-term forecasting of the electric demand of HVAC systems, Ph.D. thesis, Montreal: Concordia University, 2016.

[189] E. Alpaydin, Introduction to Machine Learning, Third Edition, PHI Learning Pvt. Ltd.: Delhi, 2014.

[190] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[191] J. Runge, R. Zmeureanu and M. Le Cam, "Hybrid short-term forecasting of the electric demand of supply fans using machine learning," Journal of Building Engineering, vol. 29, 2020.

[192] S. Taieb, A. Sorjamaa and G. Bontempi, "Multiple-output modeling for multi-step-ahead time series forecasting," Neurocomputing, vol. 73, no. 10-12, 2010.

[193] ASHRAE, "Standard 90.1-2016 Energy Standard for Buildings Except Low-Rise Residential Buildings," American Society of Heating, Refrigerating and Air-Conditioning Engineers Inc., Atlanta, 2016.

[194] L. Fausett, Fundamentals of Neural Networks, London: Pearson, 1993.

[195] X. Yao, "Evolving Artificial Neural Networks," Proceedings of the IEEE, vol. 87, no. 9, pp. 1423-1477, 1999.

[196] A. Kusiak and M. Li, "Cooling output optimization of an air handling unit," Applied Energy, vol. 87, pp. 901-909, 2010.

[197] M. Ning and M. Zaheeruddin, "Neuro-optimal operation of a variable air volume HVAC&R system," Applied Thermal Engineering, vol. 30, no. 5, pp. 385-399, 2010.

[198] H. Chenglei, L. Kangji, L. Guohai and P. Lei, "Forecasting Building Energy Consumption Based on Hybrid PSO-ANN Prediction Model," in Chinese Control Conference, Hangzhou, 2015.

[199] A. Afram, F. Janabi-Sharifi, A. Fung and K. Raahemifar, "Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and casestudy of a residential HVAC system," Energy and Buildings, vol. 141, no. 1, pp. 96-113, 2017.

[200] W. Sarle, "Neural-Nets FAQ," 17 May 2002. [Online]. Available: http://www.faqs.org/faqs/ai-faq/neural-nets/part1/preamble.html. [Accessed 20 April 2017].

[201] J. Zhao, K. Lam, E. Ydstie and O. Karaguzel, "EnergyPlus model-based predictive control within design-build-operate energy information modelling infrastructure," Building Performance Simulation, vol. 8, no. 3, pp. 121-134, 2015.

[202] R. Jovanović, A. A. Sretenović and B. D. Živković, "Ensemble of various neural networks for prediction of heating energy consumption," Energy and Buildings, vol. 94, pp. 189-199, 2015.

[203] J. Yang, H. Rivard and R. Zmeureanu, "On-line building energy prediction using adaptive artificial neural networks," Energy and Buildings, vol. 37, pp. 1250-1259, 2005.

[204] D. Monfet and R. Zmeureanu, "Ongoing commissioning of water-cooled electric chillers using benchmarking models," Applied Energy, vol. 92, pp. 99-108, 2012.

[205] Efficiency Valuation Organization, "International Performance Measurement and Verification Protocol," Efficiency Valuation Organization, Washington, 207.

[206] GE - Genomics building (exterior). [Art]. Concordia university, 2019.

[207] D. Monfet, "New ongoing commissioning approach of central plants: methodology and case," Concordia University, Montreal, QC, 2011.

[208] A. Inanlouganji, T. Reddy and S. Katipamula, "Evaluation fo regression and neural network models for solar forecasting over different short-term horizons," Science and Technology for the Built Environment, vol. 0, pp. 1-10, 2018.

[209] G. Peter and M. Matskevichus, "Hyperparameters Tuning for Machine Learning Models for Time Series Forecasting," in International Conference on Social Networks Analysis, Management and Security, Granada, 2019.

[210] N. Reimers and Gurevych, "Optimal Hyperparameters for Deep LSTM-Networks for Sequence," 2017. [Online]. Available: https://arxiv.org/abs/1707.06799.

[211] B. Feuston and J. Thurtell, "Generalized nonlinear regression with ensemble of Neural Nets: The Great Energy Predictor Shootout," ASHRAE Transactions, vol. 100, no. 2, 1994.

[212] S. Olivieri, G. Henze, C. Corbin and M. Brandemuehl, "Evaluation of commercial building demand response potential using optimal short-term curtailment of heating, ventilation, and air conditioning loads," Journal of Building Performance Simulation, vol. 7, no. 2, pp. 100-18, 2014.

[213] M. Le Cam, R. Zmeureanu, A. Daoud and A. Athienitis, "Forecasting of Cooling Coil Load using On-Site Forecasts of Outdoor Air Conditions versus Forecasts from Closest Airport," in 15th IBPSA Conference, San Francisco, 2017.

Appendix A

Table 58: District level deep learning papers

District Level	Year	Application	DNN type applied	Target Variable	Reference Number
Sector	2018	Industrial sector	DFFNN/OA	Electricity consumption	[120]
	2019	Residential and Industrial sector	LSTM/DFFNN OA	Natural gas consumption	[142]
	2019	Mackey-Glass Sector	LSTM	Natural gas consumption	[143]
City	2018	Rottne district system Karlshamn district	DFNN/OA	Heating load	[144]
	2019	District system	DFFNN/OA	Heating load	[145]
	2018	Non-residential district	DNARX/OA	Heating load Cooling load	[131]
	2019	District system	DFNN/LSTM RNN//OA	Heating load	[146]
	2019	District system	LSTM/OA	Heating load	[147]
	2020	District system	LSTM/OA	Heating load	[148]
	2019	London, Karditsa, Hong Kong, Melbourne systems	LSTM/OA	Natural gas	[149]
	2019	Ljubljana	DRNN/OA	Natural gas	[134]
Complex	2018	University	AE-RF	Electric	[150]
	2019	Industrial complex	CNN/RNN/OA	Electric	[151]
	2020	University	LSTM-FFNN	Electric Heating Cooling	[152]
	2019	University	DFFNN	Electric	[153]
	2020	Hospital complex	GMDH	Electric	[154]
	2019	District mixed buildings	DFFNN	Heating	[133]
Commercial-Residential	2016	District heating system for residential and commercial buildings	DFNN/OA	Heating	[155]
	2018	District heating system for residential and commercial buildings	DFNNN/OA	Heating	[123]
	2017	District residential	DFFNN	Electric	[119]
	2019	District residential (agg.) Residential	LSTM/OA	Electric	[156]
	2019	District Residential (agg) Residential	LSTM/OA	Electric	[157]

Appendix B

Table 59: Building level deep learning based papers

Building Level	Year	Application	DNN type applied	Target Variable	Reference Number
Institutional	2017	Education	AE-DNN/OA	Cooling	[126]
	2019	Education	CNN/AE-DNN/OA	Cooling	[136]
	2019	Education	RNN/GRU/LSTM	Cooling	[158]
	2020	University	LSTM/DFFNN/OA	Cooling	[159]
	2018	University	DRNN/OA	Heating	[132]
	2019	University	GRU	Electric Cooling	[160]
	2020	University	LSTM/OA	Electric	[161]
	2018	Education	LSTM	Electric	[162]
Commercial	2016	Office	DFFNN/OA	Cooling	[121]
	2018	Office	DBN	Cooling	[127]
	2017	Office	DFNN	Hearing Cooling	[122]
	2020	Complex	DFFNN/OA	Heating	[163]
	2020	N/s Public	LSTM/CIFG/GRU/LSTM-ANN/CIFG-ANN/GRU-ANN/OA	Cooling	[164]
	2017	Retail	Extreme-SAE/OA	Overall	[130]
	2018	Hotel	LSTM	Electric	[165]
	2019	Commercial-N/s	GRU/LSTM/RNN/DFFNN	Electric	[166]
	2020	Commercial	AE-RF/DFFNN/OA	Electric	[167]
	2020	Office	RNN-seq2seq/OA	Electric	[168]
	2020	Office	LSTM/CNN/LSTM-AE/LSTM-dense	Electric	[169]
	2018	N/s Public 1 N/s Public 2	DBN/DBEN/OA	Energy	[170]
	2020	Office	DFFNN	Energy	[171]
Residential	2019	Residential	CNN/RNN/RNN-CNN	Electric	[172]
	2020	Residential	CNN/OA	Electric	[173]
	2016	Residential customer	LSTM	Electric	[174]
	2020	Residential	LSTM/OA	Electric	[175]
	2017	Residential customer	CNN/LSTM/FRBM	Electric	[125]
	2018	Residential	LSTM/GRU/RNN/OA	Electric	[176]
Multiple Case studies	2019	Residential/City Hall/Factory/Hospital	LSTM/MIDAS-LSTM	Electric	[177]
	2016	Public Administration/ Retail/R&D/Business/ Healthcare/ Car part/Electronic/ other manufactures	RBM/OA	Electric	[178]
	2018	Industrial Commercial	LSTM/OA	Electric	[179]
	2018	Public health Residential Aggregated residential	LSTMAE-ML/OA	Electric	[129]
	2018	Public-N/s	DBM/DEBM/OA	Energy	[170]
	2018	Retail Office	DBM/OA	Overall	[180]
	2019	Hotel Office	LSTM/GRU/OA	Electric	[181]
	2019	Education Commercial	CNN/GRU/OA	Electric	[135]

Appendix C

Table 60: Sub-meter and component level deep learning papers

Year	Application	DNN type applied	Target Variable	Reference Number
2019	GSHP- Office	AE-DDPG/OA	Electric	[182]
2014	GSHP HVAC	DFFNN/OA	Electric	[124]
2016	Whole building Sub-meters	CRBM/FCRBM/OA	Electric	[128]
2019	Whole building Appliances	LSTM/OA	Electric	[183]
2020	HVAC total	DDPG/OA	HVAC Electric	[184]
2020	Refrigeration system	LSTM/OA	Compressor Electric	[185]

Appendix D

```
% Component Model in Matlab for F+1 with 3 hour validation
%% Step 1: Call in training data
        train_data = xlsread('Direct_search_results.xlsx',1, 'B3074:B4802');
        train_data = train_data';
        train_data = num2cell(train_data);
%% Step 2: Calling in Validation data (previous time step)
        data_preload_1 = xlsread('Direct_search_results.xlsx',1, 'B3086:B4814');
        data_preload_1 =  data_preload_1';
        data_preload_1 = num2cell(data_preload_1);
        val_data = xlsread('Direct_search_results.xlsx',1, 'B4815:B4826');
        val_data = val_data';
        val_data = (val_data.*(27661.44-14838.67))+14838.67;
%% Step 3: Training NARNET until error reached on RMSE validation
        RMSE_val = 1000;
        while RMSE_val > 600
                % Step 3a: Training NARNET
                 net = narnet(1:33,4);
                 [X,Xi,Ai,T] = preparets(net,{},{}, train_data);
                net = adapt(net,X,T,Xi,Ai);
                 nntraintool('close');
                 clear X T Xi Ai t
                % Step 3b: Forcasting over validation period
                [X,Xi,Ai,T] = preparets(net,{},{},data_preload_1);
                [Y1,Xfo,Afo] = net(X,Xi,Ai);
                [netc,Xic,Aic] = closeloop(net,Xfo,Afo);
                [Y2,Xfc,Afc] = netc(cell(0,12),Xic,Aic);
                % Measuring Error
                Forecast_val=cell2mat(Y2);
                Forecast_val =  (Forecast_val.*(27661.44-14838.67))+14838.67;
                RMSE_val  = (immse( val_data, Forecast_val))^0.5;
                CV_val = (RMSE_val/ mean(val_data))*100;
                clear Afc Afo Ai Aic T X Xic Xfc Xfo Xi Y1 Y2
                clc
         end
 %% Step 4: Calling in testing data for F+1 forecast
         test_data = xlsread('Direct_search_results.xlsx',1, 'B4827:B4850');
         test_data = test_data';
         test_data = (test_data.*(27661.44-14838.67))+14838.67;
         data_preload_2 = xlsread('Direct_search_results.xlsx',1, 'B3098:B4826');
         data_preload_2 =  data_preload_2';
         data_preload_2 = num2cell(data_preload_2);
  %%  Step 5: Forecasting F+1 and calculating error
          % Creating forecast
          [X,Xi,Ai,T] = preparets(net,{},{},data_preload_2);
         [Y1,Xfo,Afo] = net(X,Xi,Ai);
          [netc,Xic,Aic] = closeloop(net,Xfo,Afo);
          [Y2,Xfc,Afc] = netc(cell(0,24),Xic,Aic);
        % Calculating Error
        Forecast_testing_1 =cell2mat(Y2);
        Forecast_testing_1 =  (Forecast_testing_1.*(27661.44-14838.67))+14838.67;
        RMSE_testing_1  = (immse( test_data, Forecast_testing_1))^0.5;
         CV_testing_1 = (RMSE_testing_1/ mean(test_data))*100;
```

Appendix E

```
#################### PART 1: Import libraries
        import numpy as np
        import pandas as pd
        import keras
        from sklearn import preprocessing
        from numpy import array
        scaler = preprocessing.MinMaxScaler()
        from keras.models import Sequential
        from keras.layers import Dense
        from keras.layers import LSTM
        from random import randint
## split a multivariate sequence into samples previous
        def split_behind (series, n_steps):
                X  = list()
                for i in range(len(series)):
                        # finding the end
                        end_rr = i + n_steps
                        # seeing if beyond
                        if end_rr > len(series):
                                break
                        # combining
                        seq_b = series [i:end_rr, :]
                        X.append(seq_b)
                return array(X)
## split a multivariate sequence into a head samples (horizon)
        def split_ahead(series, n_steps):
                X  = list()
                for i in range(len(series)):
                        # finding the end
                        end_rr = i + n_steps
                        end_hor = end_rr+24
                        # seeing if beyond
                        if end_hor > len(series):
                                break
                        # combining
                        seq_a = series [end_rr:end_hor, :]
                        X.append(seq_a)
                return array(X)

#################### PART 2: Calling in data (ALL DATA)
## Select the number of time lags for the regressors and hidden layer neurons
        n_steps = 4
        h1 = 75
        h2 = 40

## TRAINING DATA IMPORT AND PREPARATION
## Import regressors training
        xls = pd.ExcelFile(r'C:\Users\NAME\Desktop\1d. Monolithic Forecasting results.xlsx')
        data_reg_full = pd.read_excel(xls, 'R-1')
        del data_reg_full['Date and time']
        n_var = data_reg_full.shape[(1)]
        data_back_pass = data_reg_full.values
##Creating data set with lags for each regressors with current and historical data
```

```
        X = list()
        for t in range (n_var):
            var = data_back_pass[:,t]
            var = var.reshape( len(var),1)
            seq = split_behind(var, n_steps)
            del (var)
            seq = seq.reshape(len(seq), n_steps)
            X.append(seq)

        data_training_reg = array(X)
        data_training_reg = np.hstack(data_training_reg)
        del(X,seq,t, data_back_pass)

## Import targets training
        xls = pd.ExcelFile(r'C:\Users\NAME\Desktop\1d. Monolithic Forecasting results.xlsx'')
        data_tar_full = pd.read_excel(xls, 'T-1')
        del data_tar_full['Date and time']
        n_var_tar = data_tar_full.shape[(1)]
        data_for_pass = data_tar_full.values

        X = list()
        for t in range (n_var_tar):
            var = data_for_pass[:,t]
            var = var.reshape( len(var),1)
            seq = split_ahead(var, n_steps)
            del (var)
            seq = seq.reshape(len(seq), 24)
            X.append(seq)

        data_training_tar = array(X)
        data_training_tar = np.hstack(data_training_tar)
        del(X,seq,t, data_for_pass)

## Align length of data backwards and forewards data. Then reshape arrays to prepare for joining
        data_training_reg = data_training_reg[:len(data_training_tar)]
## Normalize data
        scaler = preprocessing.MinMaxScaler()
        data_nreg = scaler.fit_transform(data_training_reg)
        data_ntar =  scaler.fit_transform(data_training_tar)
## Deleting excess data
        del ( n_var_tar, data_reg_full, data_tar_full)
        #del (data_training_reg, data_training_tar)

## TESTING DATA IMPORT AND PREPARATION
## Import regressors Validation
        xls = pd.ExcelFile(r'C:\Users\NAME\Desktop\1d. Monolithic Forecasting results.xlsx')
        data_reg_full = pd.read_excel(xls, 'R-1F')
        del data_reg_full['Date and time']
        n_var = data_reg_full.shape[(1)]
        data_back_pass = data_reg_full.values
## Creating data set with lags for each regressors with current and historical data
        X = list()
        for t in range (n_var):
            var = data_back_pass[:,t]
            var = var.reshape( len(var),1)
            seq = split_behind(var, n_steps)
```

```
        del (var)
        seq = seq.reshape(len(seq), n_steps)
        X.append(seq)

    data_test_reg = array(X)
    data_test_reg = np.hstack(data_test_reg)
    del(X,seq,t, data_back_pass)
## Normalize data
    scaler = preprocessing.MinMaxScaler()
    data_nreg_test = scaler.fit_transform(data_test_reg)
    #del (data_test_reg, data_test_tar)

#################### PART 3: Training AE, extracting encoder, compressing data
## Encoder 1
### Training the encoder
    model_AE = Sequential()
    model_AE.add(Dense(n_steps*n_var,activation='tanh',input_dim=n_steps*n_var))
    model_AE.add(Dense(h1,activation='tanh'))
    model_AE.add(Dense(h2,activation='tanh'))
    model_AE.add(Dense(h1,activation='tanh'))
    model_AE.add(Dense(n_steps*n_var,activation='tanh'))
    model_AE.compile(loss=keras.losses.mean_squared_error,optimizer=keras.optimizers.RMSprop(lr=0.0001
    , rho=0.9, epsilon=None, decay=0.0), metrics = ['accuracy'])
    model_AE.fit(data_nreg, data_nreg, epochs=500, verbose=0)
### Extracting the encoder
    model_EN = Sequential()
    model_EN.add(Dense(n_steps*n_var,input_dim=n_steps*n_var,activation='tanh',
    weights=model_AE.layers[0].get_weights()))
    model_EN.add(Dense(h1,activation='tanh', weights=model_AE.layers[1].get_weights()))
    model_EN.add(Dense(h2,activation='tanh',weights=model_AE.layers[2].get_weights()))
### Compressing data
    Encoder_compressed_train_data = model_EN.predict(data_nreg, verbose=0)
    Encoder_compressed_train_data=Encoder_compressed_train_data.reshape(len(Encoder_compressed_train
    _data),h2, 1)
    print ("Trained the AE")

#################### PART 4: Training 4 LSTM models for the EN
    n_features = 1
## LSTM Group 1 using encoder 1
#### Model 1
    value = randint(50, 200)
    model_LSTM_1 = Sequential()
    model_LSTM_1.add(LSTM(value, activation='relu', input_shape=(h2, n_features)))
    model_LSTM_1.add(Dense(48))
    model_LSTM_1.compile(optimizer='adam', loss='mse')
    # fit model
    model_LSTM_1.fit(Encoder_compressed_train_data, data_ntar, epochs=300, verbose=0)
    print ("Trained LSTM 1 ")
#### Model 2
    value = randint(50, 200)
    model_LSTM_2 = Sequential()
    model_LSTM_2.add(LSTM(value, activation='relu', input_shape=(h2, n_features)))
    model_LSTM_2.add(Dense(48))
    model_LSTM_2.compile(optimizer='adam', loss='mse')
    # fit model
    model_LSTM_2.fit(Encoder_compressed_train_data, data_ntar, epochs=300, verbose=0)
```

```
        print ("Trained LSTM 2 ")
### Model 3
        value = randint(50, 200)
        model_LSTM_3 = Sequential()
        model_LSTM_3.add(LSTM(value, activation='relu', input_shape=(h2, n_features)))
        model_LSTM_3.add(Dense(48))
        model_LSTM_3.compile(optimizer='adam', loss='mse')
        # fit model
        model_LSTM_3.fit(Encoder_compressed_train_data, data_ntar, epochs=300, verbose=0)
        print ("Trained LSTM 3 ")
### Model 4
        value = randint(50, 200)
        model_LSTM_4 = Sequential()
        model_LSTM_4.add(LSTM(value, activation='relu', input_shape=(h2, n_features)))
        model_LSTM_4.add(Dense(48))
        model_LSTM_4.compile(optimizer='adam', loss='mse')
        # fit model
        model_LSTM_4.fit(Encoder_compressed_train_data, data_ntar, epochs=300, verbose=0)
        print ("Trained LSTM 4 ")

        del (Encoder_compressed_train_data)

####################  PART 5: Applying encoder-LSTM models to testing data
### Compressing data
        Encoder_compressed_test_data = model_EN.predict(data_nreg_test, verbose=0)
        Encoder_compressed_test_data=Encoder_compressed_test_data.reshape(
        len(Encoder_compressed_test_data),h2, 1)
### Forecasts
        ### AE-LSTM Group 1
        forecast_LSTM_1= model_LSTM_1.predict(Encoder_compressed_test_data, verbose=0)
        forecast_LSTM_2= model_LSTM_2.predict(Encoder_compressed_test_data, verbose=0)
        forecast_LSTM_3= model_LSTM_3.predict(Encoder_compressed_test_data, verbose=0)
        forecast_LSTM_4= model_LSTM_4.predict(Encoder_compressed_test_data, verbose=0)
        LSTM_ensemble_forecasts= np.mean([forecast_LSTM_1, forecast_LSTM_2, forecast_LSTM_3,
        forecast_LSTM_4], axis=0)
```