Thermal Modeling for Average Size Commercial Buildings

by

Aishwarya Mukunda

California State University, Fresno

May 2019

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Abstract

In USA, commercial buildings consume around 70% of the total electricity generation. A great share of this energy and a considerable budget is spent in Heating, Ventilating and Air Conditioning (HVAC) systems. Hence, improving the energy efficiency and reliability of HVAC systems is crucial. Respective energy efficiency and diagnostic applications usually rely on a thermal model. A thermal model predicts the thermal response of the building based on various factors such as current indoor temperature, outdoor temperature, occupancy and previous heating or cooling actions. The existing thermal models are typically building specific and simplistic. This, in turn, deteriorates the performance of intelligent HVAC applications. In this work, we propose and evaluate several thermal modeling approaches for average size commercial buildings utilizing various machine learning and statistical techniques ranging from principal component analysis for feature identification to linear regression and neural networks. A comparative study is made on these methods concerning complexity, efficiency, and generality.

Chapter 1

Introduction

In this chapter we will discuss about the motivation of this research, its importance and the problem statement.

1.1 Motivation:

In the USA, commercial buildings consume 70% of the total US electricity consumption. From this share 60% is spent on Heating, Ventilating, and Air Conditioning Systems [32]. On top of that HVAC considers a great gas consumption system. Electricity and natural gas consumption are increasing at a rapid scale [34]. Hence, improving the energy efficiency and reliability of HVAC systems is crucial. The cost and energy spent on these systems can be reduced by many ways such as improving insulation, bettering the HVAC technology and/or incorporating intelligent systems. Importantly, energy efficiency improvements that rely on intelligent systems do not require considerable retrofitting. Now, these intelligent systems rely on thermal models.

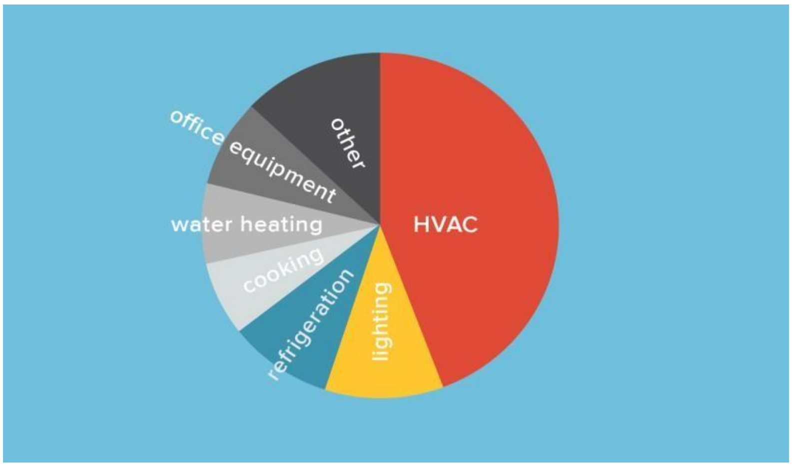


Figure 1.1 Energy consumption in commercial buildings

1.2 Research Objective:

As discussed, thermal modelling of buildings is crucial in many applications, including advanced (efficiency or economic) control of Heating, Ventilating and Air Conditioning

(HVAC) systems, intelligent HVAC diagnostics and building design [1]. A thermal model predicts the thermal response of a building based on several factors such as indoor temperature, outside temperature, occupancy, and humidity [2]. Various thermal modelling approaches have been proposed over time (e.g., [3,5,6,7]) that differ in several criteria such as generality, efficiency, and robustness. Depending on their differences, such modeling approaches can support different applications in different buildings. However, this divergence is a bottleneck for the widespread adoption of intelligent thermal applications in today's buildings. This is the case since a building-specific, and application-specific thermal model needs always to be identified before the incorporation of a thermal application. Against this background, in this project, we focus on identifying a thermal model that can support the widespread adoption of intelligent thermal applications at least in today's small size USA commercial buildings.

Figure 1.2 Commercial buildings

Arguably such a thermal model needs to meet the following criteria:

i. general enough to adequately model most (if not all) buildings considered,
ii. adequate efficiency to support sensitive advanced economic control applications potentially driven by demand response signals,
iii. relies only on data that are readily available in buildings without excessive additional instrumentation requirements, such as thermostat temperature readings and weather forecasting reports (as opposed to window shutter sensors and pyranometer readings)
iv. able to cope with intermittent and missing data that is usually the case in real settings applications,
v. robust enough and consistent to ensure the reliable operation of control algorithms,
vi. favourably has complexity that allows fast linear or convex programming optimization techniques to be utilized for control,
vii. can be trained in real operation time, and

viii. requires minimum, and readily available, knowledge about the building structure.

In this context, in this project, we evaluate different thermal modelling approaches in buildings across California against the requirements. Notably, we also discuss any unavoidable tradeoffs among the above requirements and possible workarounds.

Figure 1.3 Smart thermostat

1.3 Contributions:

In this project we propose a thermal model using different techniques like linear regression, lasso regression, neural networks and principal component analysis for feature identification. The model is built for both first order and second order dataset to consider time series autocorrelation. The above models are evaluated using real data and a thorough comparison is made against generality, efficiency and performance. A thermal model is identified that meets the above requirements for average size commercial buildings.

We now continue to discuss about related work and required background in chapter 2, then describe our approach in chapter 3. The evaluation of model and the results are described in Chapter 4. At the end we talk about the conclusion and future scope of the project.

Chapter 2

Literature Review

In this chapter we discuss in detail about the existing models and the methods used to build these models.

2.1 Thermal Modeling:

Like almost any modelling, thermal modelling can also be broadly classified into the following three categories [1]:

- ***White box***: It requires prior knowledge about the structure of buildings and the physics of thermal dynamics. This is a stringent requirement that renders such approaches impractical in real settings.
- ***Black box***: It is a data-driven approach which uses statistics and Machine Learning techniques. They do not require prior knowledge about the structure of buildings and physics of thermal dynamics but generally require a large amount of data and cannot be interpreted in physical terms. The latter can lead to non-physical performance (such as predicting a temperature increase while cooling if undertrained).
- ***Gray box***: It overcomes these drawbacks and hence, is the most suitable approach for our identification. In more detail, grey-box approaches are based on derived Equivalent Thermal Parameter (ETP), which are determined by methods like Kalman Filters and least square methods.

In the following paragraphs we discuss more in detail on Gray box thermal modelling because it does not require data about the structure of the building and the model has a physical meaning which can be generalised over the buildings.

2.2 Gray Box Thermal Modeling:

For instance, the simplest thermal equation can be defined as [1],

$$T(t+1) = c \times T(t)$$

where c represents collectively the thermal dynamics.

This constant is determined by training a large dataset with supervised learning and statistical techniques.

2.3 Supervised Learning:

Supervised Learning is a data mining task where it makes predictions for given input features based on the dataset containing input-output pairs. The following are the steps generally followed to solve a supervised learning problem:

1. ***Data Preprocessing***: It is used to convert a raw dataset to clean dataset like handling the missing data, identifying the outliers and correct data inconsistencies.
2. ***Feature Identification:*** The features that determine the target variable is only identified and remaining are removed. Principal Component Analysis is one of the techniques used to find the most uncorrelated features to predict the target variable [4].
3. ***Training Algorithms***: Appropriate algorithms are chosen to build a model that predicts the target variable. The model is built by applying the algorithm on the training dataset.
4. ***Evaluation***: The model is applied to a test dataset to evaluate the accuracy of the predictions made. The model can be evaluated based on root mean squared error [5,7].

After training several models, the most appropriate model is selected. In the following paragraphs we will discuss in detail about the methods for feature identification and training algorithms.

2.4 Principal Component Analysis (PCA):

It is a famous technique for dimension reduction, feature extraction, and data visualization. It transforms high dimensional dataset into a low dimensional dataset such that it contains only linearly uncorrelated variables called the principal components. The first principal component has the maximum variance of the dataset, and the following principal components have high variance and are orthogonal to the previous principal components. It helps in improving performance without losing important information [17].

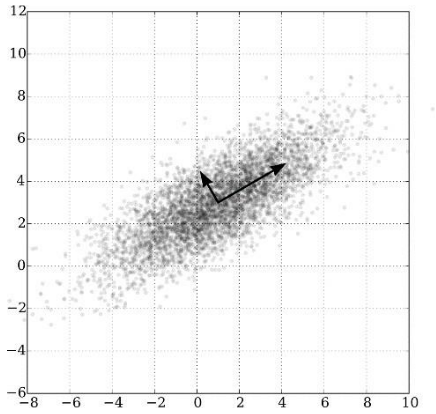


Figure 2.1 Principal component analysis

2.4.1 Procedure:

Consider a dataset represented in the form of a matrix X, where m columns represent features, and n rows represent samples. [33]

1. Normalize the dataset.
2. Find the covariance of the matrix(A).
3. Calculate eigenvectors and corresponding eigenvalues.
4. Sort the eigenvectors and select the first k dimensions.
5. Transform the original dataset into k dimensions.

The dataset is normalized before performing PCA because all the features must be of same scale in order to compare them. If not, it will result in misleading components.

The covariance of the matrix can be calculated as,

$$C_X = (X - \overline{X})(X - \overline{X})^T / n - 1$$

Where X^T is transpose of X.

In covariance matrix, main diagonal elements contain variance of dimensions and off diagonal elements contain covariance of dimensions. So, the goal of PCA is to find and remove correlated dimensions and variance along dimensions must be high. This implies values in diagonal must be high and that in off diagonal should be zero. This type of matrix is called diagonal matrix. Hence the original dataset must be transformed such that its covariance matrix is a diagonal matrix. This method is called diagonalization.

$$
\begin{aligned}
&C_x = covariance\ matrix\ of\ original\ data\ set\ X \\
&C_y = covariance\ matrix\ of\ transformed\ data\ set\ Y \\
&such\ that, \\
&Y = PX \\
&For\ simplicity,\ we\ discard\ the\ mean\ term\ and\ assume\ the \\
&data\ to\ be\ centered.\ i.e.\ X = (X - \bar{X}) \\
&So,\ C_x = \tfrac{1}{n}XX^T \\
&C_y = \tfrac{1}{n}YY^T \\
&= \tfrac{1}{n}(PX)(PX)^T \\
&= \tfrac{1}{n}PXX^TP^T \\
&= P(\tfrac{1}{n}XX^T)P^T \\
&= PC_xP^T
\end{aligned}
$$

The eigen vector matrix (P) for C_x is calculated such that C_y is a diagonal matrix and Y becomes the transformed matrix.

In [4], the dataset had eight variables of temperature features. On applying PCA, it reduced down to 1 parameter. This proved better than the aggregated temperature when plotted on a graph.

2.5 Linear Regression:

Linear regression is a supervised learning technique used to predict continuous values. It models a linear relationship between two variables, namely explanatory or independent variable and dependent variable. The value in the dependent variable is determined based on the independent variable [13]. Consider a graph, where the X axis is represented as the independent variable and Y-axis is represented as the dependent variable. The linear relationship between the variables is described in terms of a line called Regression line. The relationship is positive if the dependent variable increases, with an increase in the independent variable. The relationship is negative when the dependent variable decrease, with an increase in an independent variable. Observation is an entity indicating the value of a dependent variable based on the given independent variable. The regression line is constructed by trying to fit the observations on it, by the method of least squares. The aim of the least squares method is to minimize the errors that are, the difference between the estimated value and actual value [12]. The regression line is defined as

$$y = b_0 + b_1 \times x$$

Where y is the estimated value of the dependent variable

x is an independent variable

b_0 is y-intercept

b_1 is slope

The slope is positive(+) if the relationship is positive else, it is negative(-).

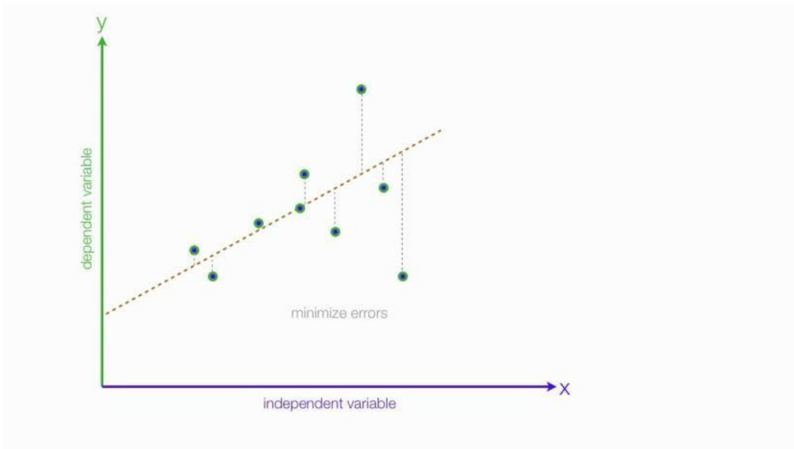


Figure 2.2 Linear regression

2.5.1 Procedure:

1. Calculate the mean $(\overline{x}, \overline{y})$ of independent (x) and dependent (y) variable respectively.
2. Find the distance between each observation and its mean $(x - \overline{x}, y - \overline{y})$
3. nd the squares of distances calculated for the independent variable $((x - \overline{x})^2)$
4. Multiply $(x - \overline{x})$ and
5. $b_1 = sum(x - \overline{x} \times y - \overline{y}) / sum((x - x)^2)$
6. Substitute b_1 in $y = b_0 + b_1 \times x$,where and $x = \overline{x}$ $y = \overline{y}$ and solve for b_0.

If there is more than one independent variable, then it is called a multiple linear regression model. It is defined as,

$$\hat{y} = w[0] \times x[0] \ + \ w[1] \times x[1] \ + \ ... \ + \ w[n] \times x[n] + b$$

Where b is intercept, x represent independent variables, and w represents its respective weights.

The best fit line is modeled using the least squares method. It aims at minimizing the squares of errors(the difference between the predicted value and actual value). This cost function is defined as,

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} \left(y_i - \sum_{j=0}^{p} w_j \times x_{ij} \right)^2$$

In [8], the total annual heating and cooling energy requirements were predicted using a multiple linear regression model. It is easier and more practical to use a linear model than nonlinear. Many optimizations rely on linear programming. The results of this paper indicate that there is a linear dependence between the predicted variable and its features.

The linear regression makes the following assumptions about the dataset:

- There exists a linear relationship between the dependent and independent variables.
- Linear regression is sensitive to outliers; hence outliers must be handled.
- All variables are multivariate normal, or they have a normal distribution.
- There is little or no multicollinearity that is all independent variables are not highly correlated.
- There is little or no autocorrelation that is the values of a feature are independent of each other.

If the dataset contains multicollinearity and the variables are not multivariate normal, then additional regularization needs to be added to the cost function. This regularization is done in Lasso regression which is explained in detail below.

2.6 Lasso Regression:

Least absolute and shrinkage and selection operator (Lasso) is a type of linear regression model. It focuses on reducing the complexity of the model by performing variable selection using regularization. The magnitude of coefficients is considered for regularization [15]. The cost function of linear regression is modified for the cost function of lasso regression as,

$$\sum_{i=1}^{M}(y_i-\hat{y}_i)^2=\sum_{i=1}^{M}\left(y_i-\sum_{j=0}^{p}w_j\times x_{ij}\right)^2+\lambda\sum_{j=0}^{p}|w_j|$$

This follows the L1-norm regularization. When λ tends to 0, the cost function is the same as that of linear regression. The regularization parameter penalizes the weights of the variables leading to zero coefficients, and hence, some variables are neglected. Therefore, this approach helps in both feature selection and overfitting of model [9].

2.7 N Order Models:

Linear regression assumes there is little or no autocorrelation. In time series dataset, there is autocorrelation as temperature values at a given timestep does depend highly on the temperature value at previous or future timesteps. This time series autocorrelation is accommodated by incorporating N order models, where each order represents the number of previous timesteps included in modeling [11]. For instance, for the first order model, only the current timestep is included while for the second order, the current timestep along with previous timesteps is included.

First Order: $T(t+1)=c_1\times T(t)$

Second Order: $T(t+1)=c_1\times T(t)+c_2\times T(t-1)$

Third Order: $T(t+1)=c_1\times T(t)+c_2\times T(t-1)+c_3\times T(t-2)$

.

.

.

N Order: $T(t+1)=c_1\times T(t)+c_2\times T(t-1)+c_3\times T(t-2)+\ldots+c_n\times T(t-n)$

The N-order model increases the complexity of the model and hence methods like neural networks are applied to handle the complex operations over the complex data.

2.8 Neural Networks

Neural networks are a class of Machine Learning techniques that are designed based on how the human brain works [18]. Neural networks in time series forecasting consider the time series autocorrelation.

2.8.1 Feed Forward Network

Feed forward network is the first and simplest neural network designed. It consists of input layer, output layer and hidden layer. The information flows in only one direction in network that is from input to hidden to output. There is no storage in network to store the information. Hence the output cannot be further sent back to the input layer to auto forecast next prediction. Hence this type of neural network is not suitable for time series forecasting.

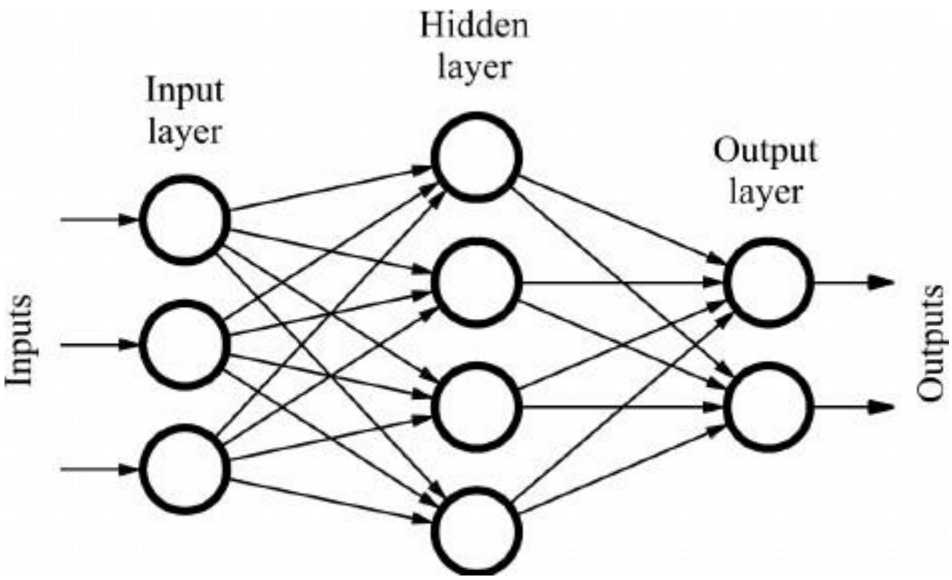


Figure 2.3 Feed Forward Network

2.8.2 Recurrent Neural Network

Recurrent neural networks are a class of neural networks that allows the flow of information in both directions in the network, that is from input to output and output to input. Its internal state acts like a memory to process the input in sequence. Hence it overcomes the shortcomings of feed forward neural network.

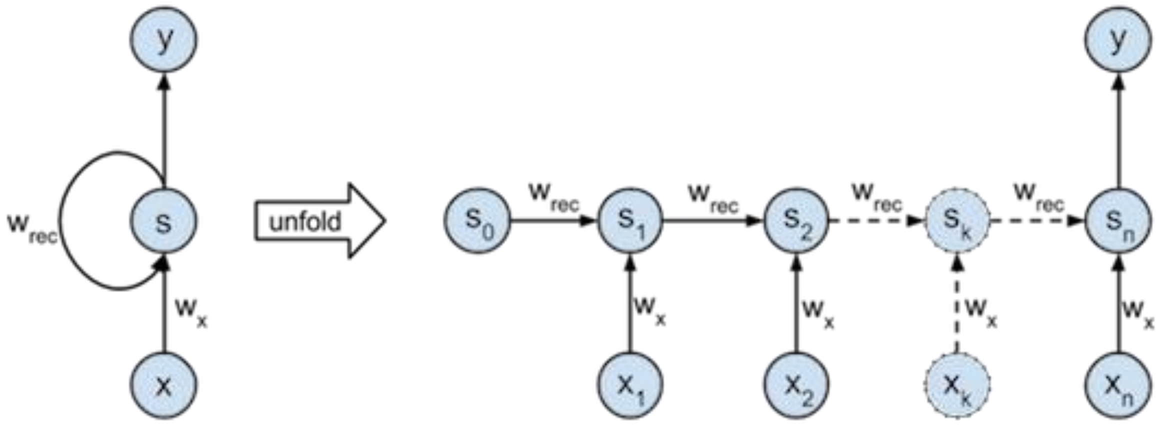


Figure 2.4 Recurrent Neural Network

The current state S_t is derived by a recursive function with previous state and input at time t.

$$S_t = tanh(W_s \times S_{t-1} + W_X \times X_t)$$

$$Y = W_y \times S_t$$

Where W_s, W_x, W_y are weights.

It uses backpropagation principle to train the dataset. Gradients are values that are used to update the weights of the network. During back propagation, the gradients become smaller and smaller. The weights in earlier layers become negligible, which doesn't help in learning. Since there is no learning in the earlier layers, the information in earlier layers is lost, resulting in short term memory of the network. This problem is called the vanishing of gradients. Hence it is not suitable for time series forecasting.

2.8.3 Long Short-Term Memory Networks (LSTM):

Long Short Term Memory network or LSTM network is a machine learning technique to make predictions of time series data or data that contain long sequences. It uses Backpropagation through time to train the dataset. It is an improvisation of the recurrent neural network as it overcomes the problem of vanishing gradients. LSTMs are better because it is insensitive to the unknown lags between important events [20].

LSTM solves the problem of vanishing gradients by incorporating three gates, which controls the flow of information in the network and a cell state that holds all the relevant information and processes it further in the network. Hence, the relevant information is not lost, and the network has a long memory.

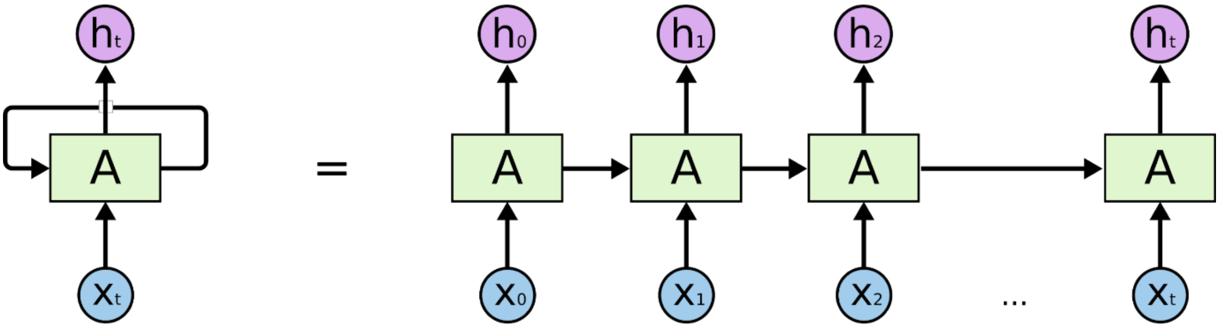


Figure 2.5 LSTM Network

Forget Gate (F_t): It decides which information to keep and which to forget from memory. Information in the previous hidden state (S_{t-1}) and current input (X_t) is passed through a sigmoid function. Sigmoid function (***σ***) is used to squeeze the values in the range 0 to 1. If the value is closer to 0, the information is forgotten. If the value is closer to 1, the information is kept to be stored in the cell state.

$$F_t = (W_{fs} \times S_{t-1} + W_{fx} \times X_t)$$

Where W_f is weights corresponding to state and input for forget gate.

Input Gate (I_t): It decides which information to keep and which to forget from the given input. Information in the previous hidden state (S_{t-1}) and current input (X_t) is passed through a sigmoid function. Values closer to 0 are forgotten, and that are closer to 1 are remembered. Information in the previous hidden state (S_{t-1}) and current input (X_t) is passed through a tanh function. Tanh function is used to squeeze the values in the range -1 to 1. The tanh output is the intermediate cell state ($\check{C}_t$). The sigmoid output is multiplied with tanh output to decide what information must be kept from the tanh output.

$$i_t = (W_{is} \times S_{t-1} + W_{ix} \times X_t)$$

$$C_t = tanh(W_{cs} \times S_{t-1} + W_{cx} \times X_t)$$

$$I_t = i_t \times C_t$$

Where W_i and W_c are weights corresponding to state and input for input gate and cell state.

Cell State (C_t): It is the memory block in the network. It holds all the relevant information. Information in previous cell state (C_{t-1}) is represented in the form of vectors and is multiplied with the sigmoid output of forget gate (F_t) to remove information from cell state that is not required. This output is added to the output of the input gate (I_t) to add the new required information into the cell state.

$$C_t = (F_t \times C_{t-1} + I_t)$$

Output Gate (O_t): It decides the output based on input and memory. Information in the previous hidden state (S_{t-1}) and current input (X_t) is passed through a sigmoid function. The new cell state (C_t) is passed through tanh function, which is multiplied with the above sigmoid output to remove information that is not required for the output. This output becomes the new state (S_t). The new state and new cell state is passed on further for the next time step.

$$O_t = (W_{os} \times S_{t-1} + W_{ox} \times X_t)$$

$$S_t = tanh(C_t) \times O_t$$

Where W_o are weights corresponding to state and input for output gate.

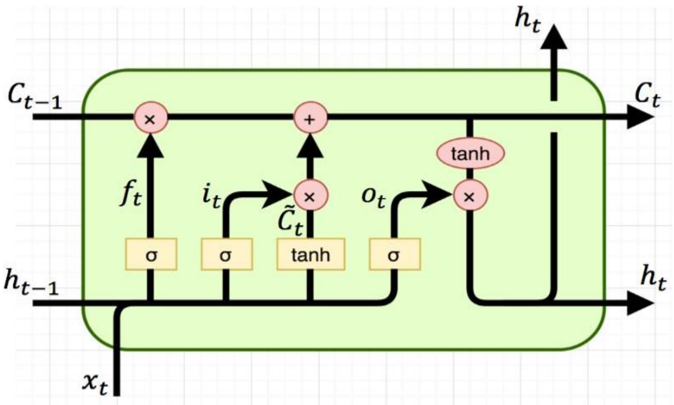


Figure 2.6 LSTM cell

In [10], the electric load was predicted for two days to mid-week. It could pass past inputs and predict sequenced outputs.

Chapter 3

Our Approach

Based on the related work and background discussed in the previous section, the following methods are investigated to build a thermal model:

- Principal Component Analysis (PCA)
- Linear Regression
- Lasso Regression
- Long Short Term Memory (LSTM)

These methods are applied to a simple model containing features of current timestep only(first order). Further, they are applied to a complex model containing features of the current and previous timestep.

3.1 Data Preprocessing:

The raw time series data was converted to a supervised learning problem by defining the target variable and feature variables that affect the target variable.

Time Index(t)	Temperature(t)

Table 3.1 Time series data set.

⇩

Time Index	Temperature(t)	Temperature(t + delta)

Table 3.2 Time series data set: Supervised learning

3.1.1 One-hot encoding: It is a technique which converts categorical feature to a binary feature(2 categories).

Time Index	Action
t1	No action
t2	heating
t3	cooling

Table 3.3 Multiple category data set

⇩

Time Index	action_heating	action_cooling
t1	0	0
t2	1	0
t3	0	1

Table 3.4 one hot encoding

3.1.2 Quality Control Test: The unrealistic values that are very high or very low values of features that act as outliers are removed by consulting the experts at the University of California Berkeley.

3.1.3 Standardization: It is a common technique to convert the dataset into a common scale so that all features are comparable. It is necessary for PCA. StandardScaler is a common technique used for standardization. It transforms the values of features to the range 0 to 1 such that the mean is 0 and variance is 1. The Standard score z is calculated as,

$$z = (x - \mu)/\sigma$$

Where x is the value of the feature, μ and σ are the mean and standard deviation of the values of that feature respectively.

3.2 Linear Regression:

A linear relationship is derived between the target variable, that is the indoor temperature at future time t + delta minutes and the features i.e. indoor temperature, outdoor temperature, action, occupancy at current time t (first order). It is also applied to a dataset that contains features at previous timesteps,t-1 (second order). The regression line is defined as

First Order:

$$T\_in(t + delta) = c_1 \times T\_in(t) + c_2 \times T\_out + c_3 \times occ$$

$$+c_4 \times action\_cooling(t) + c_5 \times action\_heating(t) + intercept$$

Second Order:

$$T\_in(t + delta) = c_1 \times T\_in(t) + c_2 \times T\_out + c_3 \times occ + c_4 \times action\_cooling(t)$$

$$+c_5 \times action\_heating(t) + c_6 \times T\_in(t-1) + c_7 \times T\_out(t-1) + c_8 \times occ(t-1)$$

$$c_9 \times action\_cooling(t-1) + c_{10} \times action\_heating(t-1) + intercept$$

3.3 Linear Regression with PCA:

PCA is applied to the dataset before performing linear regression for the feature identification. The PCA transforms the dataset to new k features from n features. Hence the regression line changes to,

$$T\_in(t + delta) = c_1 \times x_1 + c_2 \times x_2 + \ldots + c_k \times x_k + intercept$$

3.4 Long Short Term Memory (LSTM):

The first layer of LSTM performs PCA, hence PCA is not applied to dataset prior to LSTM. LSTM network consists of 3 layers: input, hidden, and output. A competent network architecture can be designed by following the rules below:

- Every neural network has one input and one output layer.
- In the input layer, the number of neurons is the same as the number of features in the dataset.
- In the output layer, the number of neurons is one as the predictor variable is a continuous value.
- One hidden layer is sufficient to solve the problem.
- The optimal size of the hidden layer is between the input and output layer. Hence, the number of neurons in the hidden layer can be considered as the mean of the number of neurons in the input and output layer.

Several architectures were built and tested, but the architecture that followed the above rules gave the best result.

3.5 Real-time applications:

All the above methods were applied to a real time dataset. This dataset was collected from the following three different stores at Corporation Yard in Berkeley:

- Electric shop
- Radio shop
- FAC Main Department

Figure 3.1 City of Berkeley-Corporation Yard

The dataset contains the following features:

- Historic and Current Indoor Temperature
- Historic and current Outdoor Temperature data
- Electric consumption
- Historic Heating and Cooling Actions
- Occupancy Schedule

The temperature data collected was an aggregation of readings from multiple nearby weather stations as they do not have a local weather station. The thermal model was built for each of the stores by training and testing the model separately for each store. The models are then evaluated based on root mean squared error (RMSE) and mean squared error (MSE) [5,7].

Chapter 5

Results

In this chapter we discuss the evaluation of the models with respect to standard error metrics and make a thorough comparison between them.

5.1 Comparison of Models

The thermal model is used to forecast the temperature for the next 7 hours for a sample data of a specific store. The average of the errors collected from 800 samples was plotted. The following graphs represent the average of RMSE values for 7 hours forecasting calculated for 800 samples.

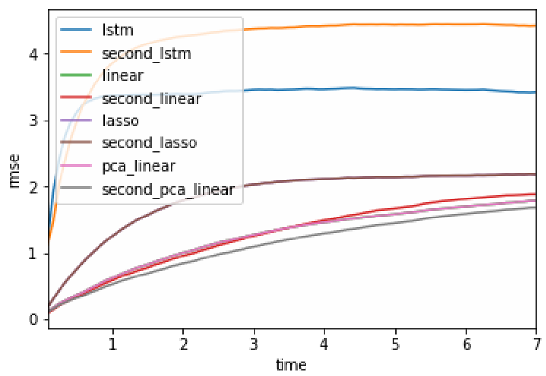


Figure 5.1 Electrical Shop

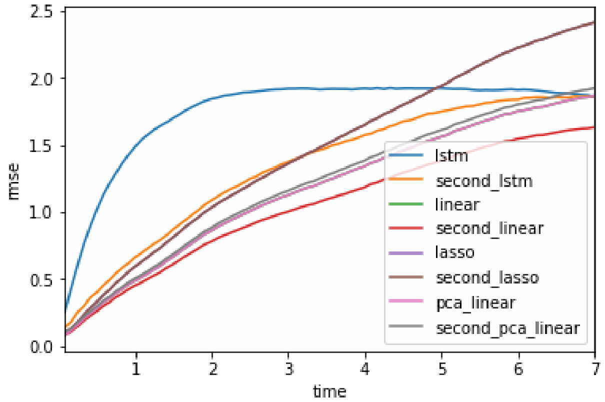


Figure 5.2 Radio Shop

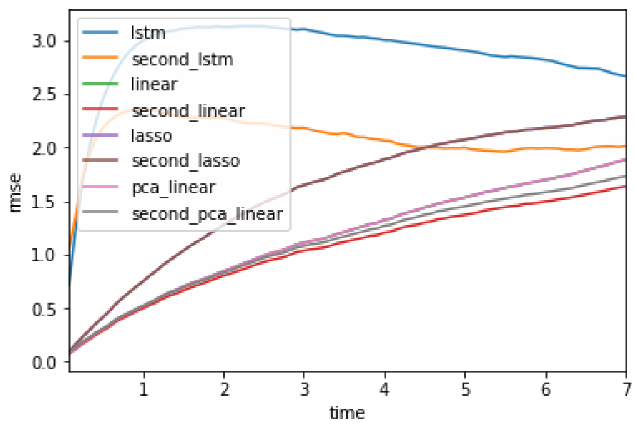


Figure 5.3 FAC Main Department

Further, the average of the RMSE values across the three stores were calculated and plotted.

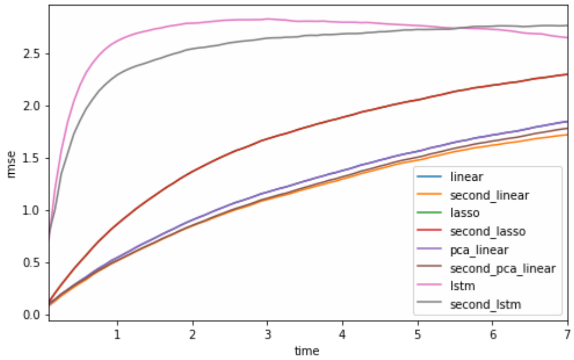


Figure 5.4 General Models

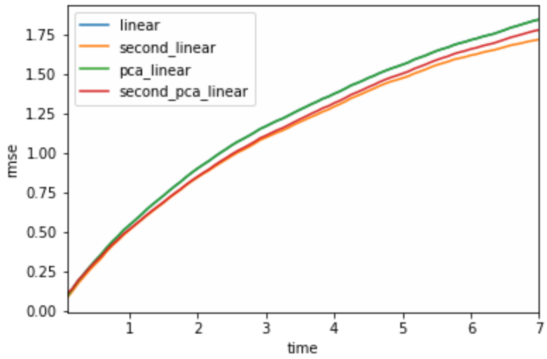
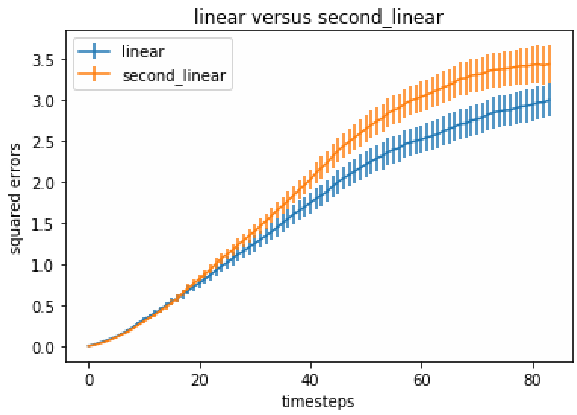


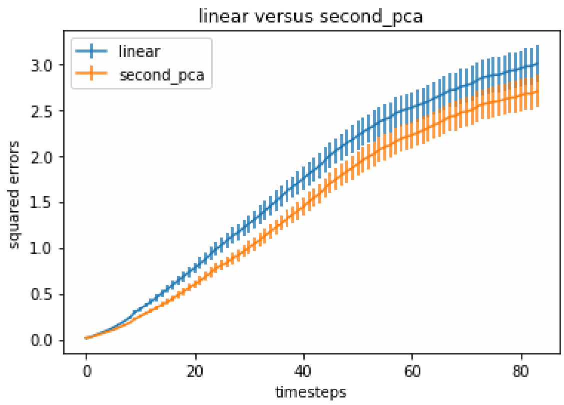
Figure 5.5 General Linear Model

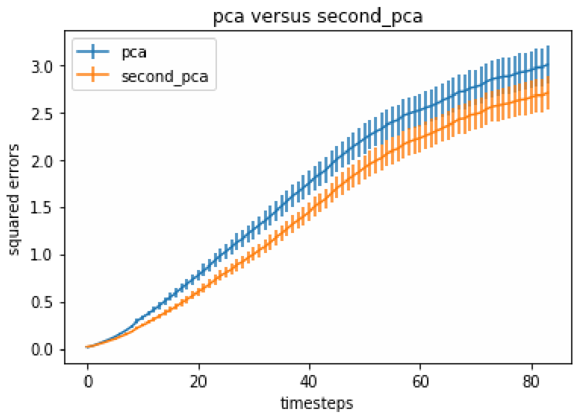
5.2 Confidence Interval

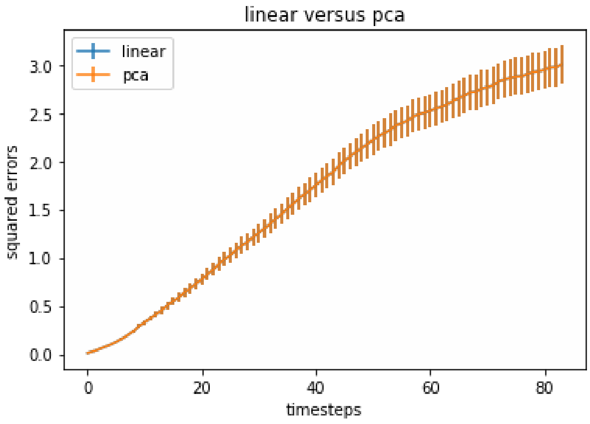
Each model is compared to each of other models by plotting their 95% confidence intervals (Fig). 95% confidence interval indicates the mean of the population lies 95% of the time in the confidence interval. If the confidence intervals of the two models do not overlap, then there is a significant difference between the two models. If they overlap, then it is not certain if the two models are significantly different from each other or not.

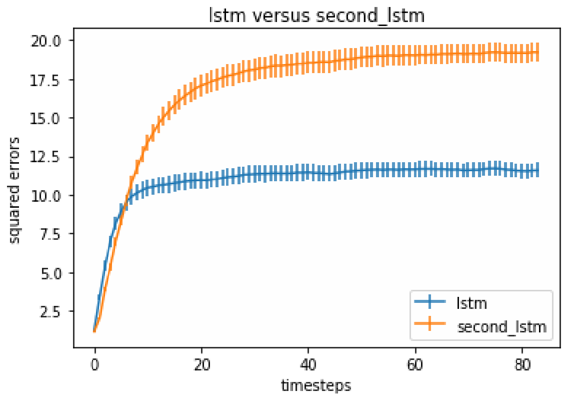
The following graphs represent the confidence interval for the electrical shop.



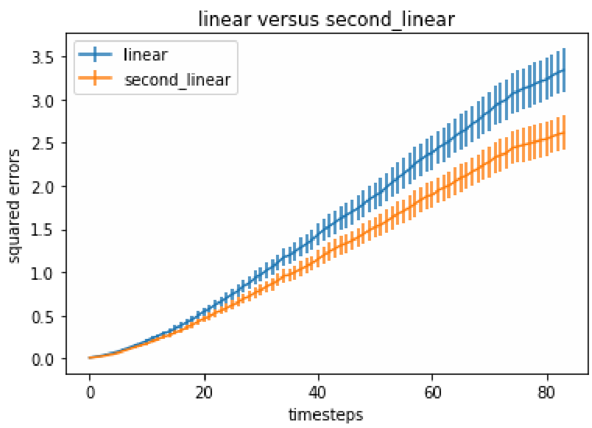


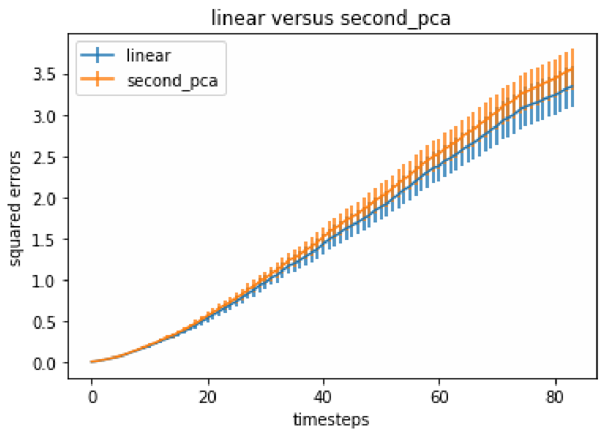


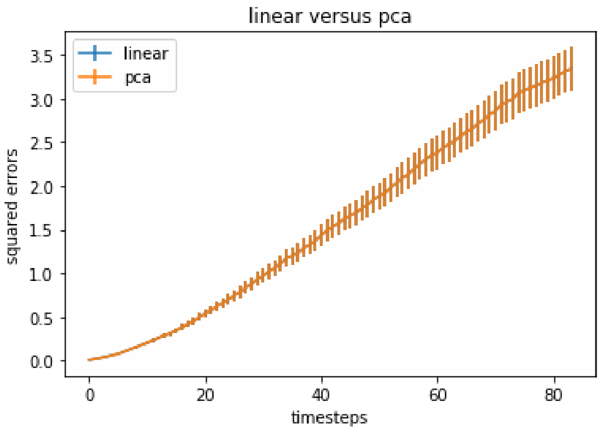


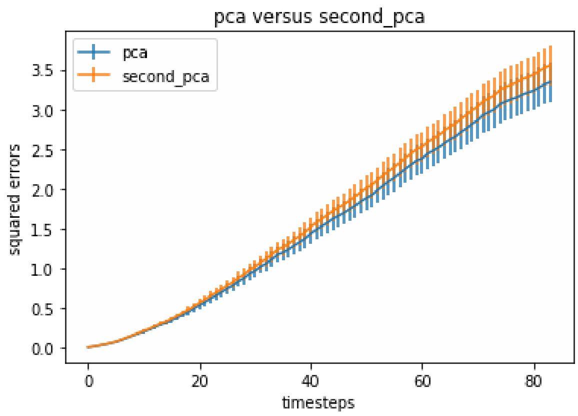


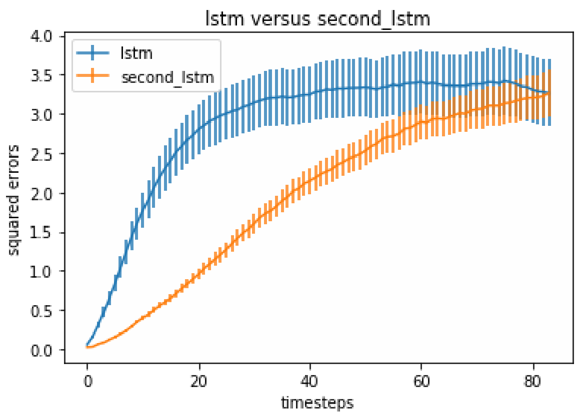
The following graphs represent the confidence interval for the radio shop.



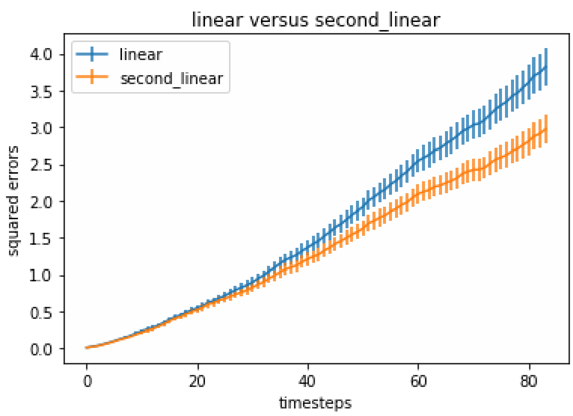


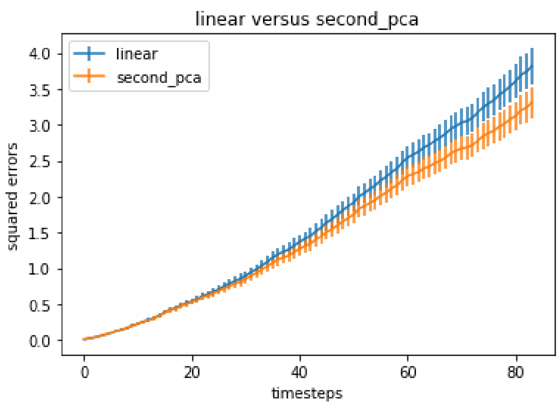


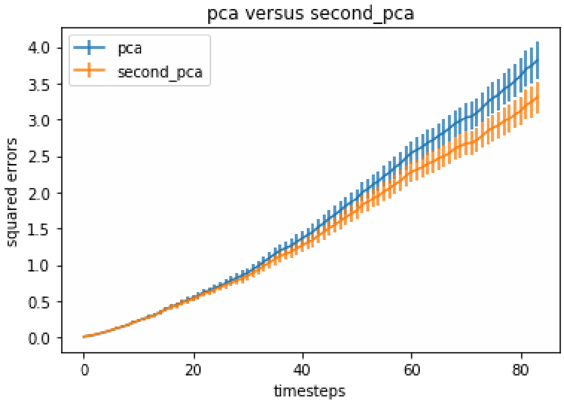


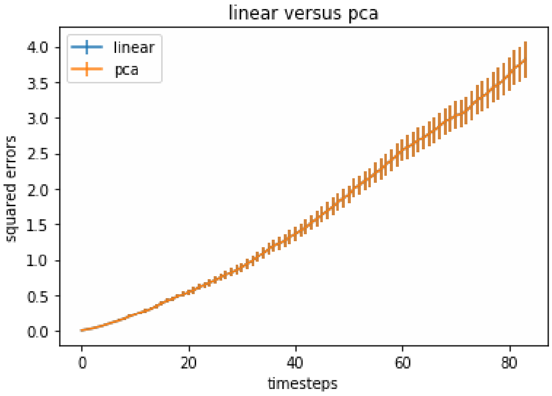


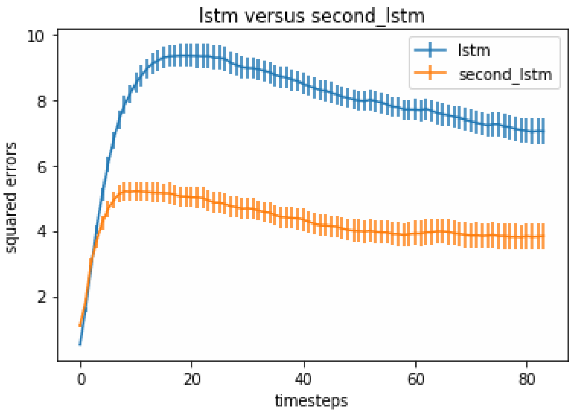
The following graphs represent the confidence interval for the FAC Main department.



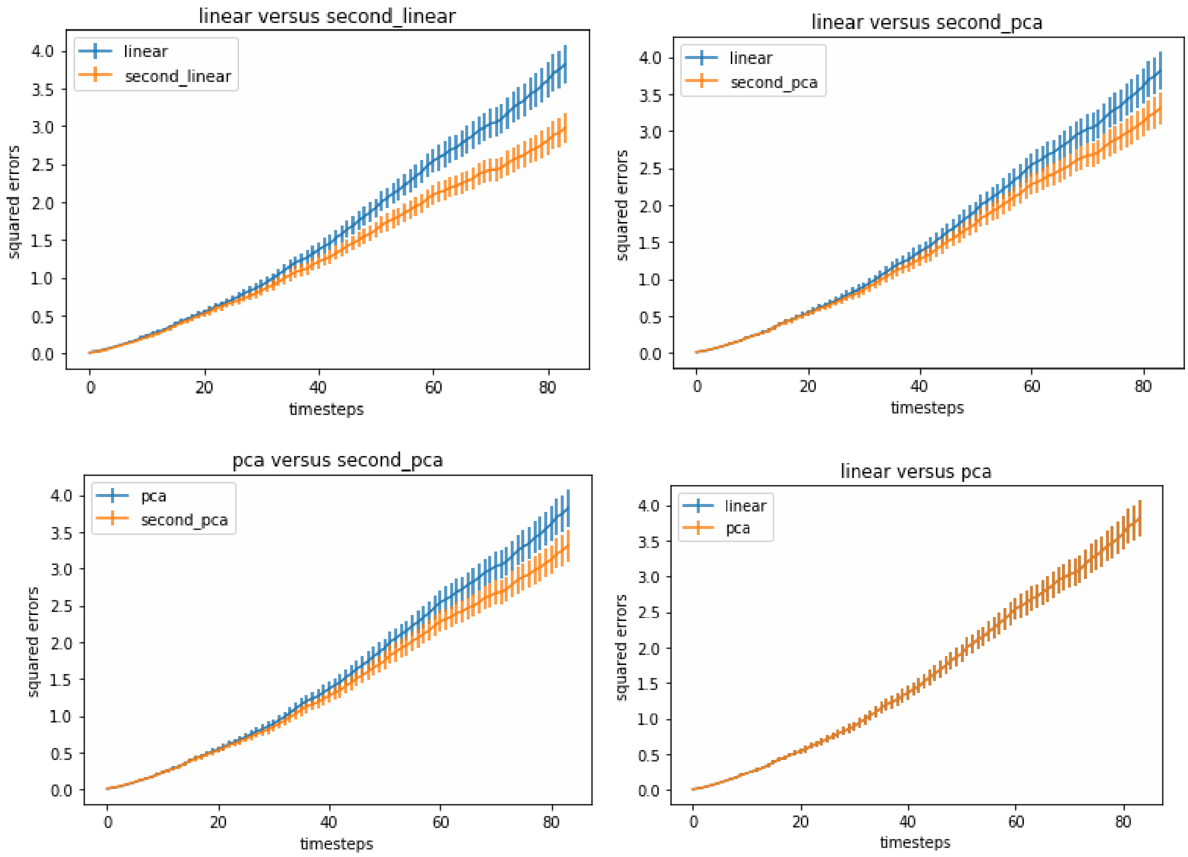


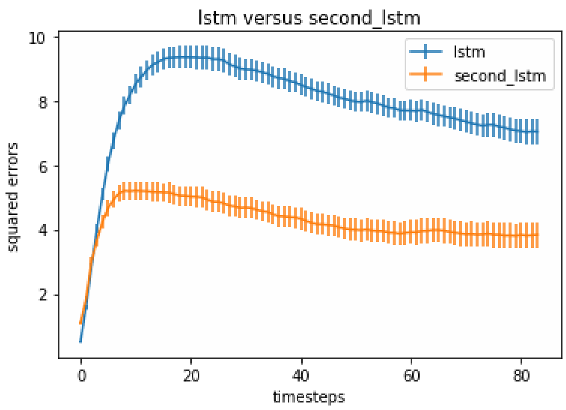






The following graphs represent the average of confidence intervals of the above stores.





5.3 T-Test:

Inferential statistics uses a random sample from a population to describe and make inference about that population. The t-test is a type of inferential statistics. It is typically used to determine if there is a significant difference in the mean between two populations or the difference observed is by a random chance [24]. There are three basic types of t-test:

- Independent t-test: It compares the mean of two different populations.

- Paired t-test: It compares the mean of the same population at different times.
- One-sample t-test: It compares the mean of a group against the known mean.

In this research, the comparison is made between two models against the same sample data. Hence, the paired t-test is used for the comparison. For this research, the null and alternate hypothesis is stated as follows:

Null Hypotheses: The mean of the difference between the means is zero.

Alternate Hypotheses: There is a difference in mean between the two given models.

We defined in advance, that if the p-value is smaller than 0.05, then the null hypothesis is rejected which means that there exists a statistical significance in the difference between the two models.

Table 5.1 shows the effect of the difference in the means of errors of different models.

Significant	**Insignificant**	**Both**
Linear and LSTM (both first and second order combinations)	First Order Linear and First Order Linear with PCA	First and Second Order Linear
First Order Linear with PCA and Second Order Linear with PCA		Second Order Linear and PCA
		First Order LSTM and Second Order LSTM

Table 5.1 T-test results

Chapter 6
Conclusion

The dataset was collected from the existing information available and did not require extra instrumentation. It was preprocessed to take care of gaps in time intervals and missing data. Different training algorithms were applied to the preprocessed dataset and evaluated. The first order and second order linear model showed almost similar results, hence on the grounds of simplicity, we suggest using the first order linear model. Since only important and limited features that were readily available were used, PCA and Lasso did not help much in improving the accuracy. Surprisingly, LSTMs did not give better results. The linear model gave the lowest error rate, and it fit well for the above-mentioned buildings. Therefore, the linear model can be regarded as a general and efficient model. The predictions made were fast and in real time operation. We can conclude based on the rule, "Keep it simple," linear regression works best.

6.1 Future scope

As a part of future work, this model could be incorporated in intelligent application systems which helps in managing the cost and consumption of energy and evaluated in this context. Adaptive thermal modelling could be used and evaluated so that the model is trained daily instead of trained only for one time. Further, more data could be collected and utilized in evaluating different machine learning techniques to improve the results.

References

1. Panagopoulos A.A, “Efficient Control of Domestic Space Heating Systems and Intermittent Energy Resources.” Section: Thermal Modelling of House, pp. 27-29 Ph.D. thesis, University of Southampton, October 2016. [1]

2. A.A. Panagopoulos, M. Alam, A. Rogers, N.R. Jennings, “AdaHeat: A General Adaptive Intelligent Agent for Domestic Heating Control,” Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, 2015. [2]

3. M. Alam, A. A. Panagopoulos, A. Rogers, N. R. Jennings, and J. Scott, “Applying extended kalman filters to adaptive thermal modelling in homes: Poster abstract.” In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, BuildSys ’14, pages 214–215, NY, USA, 2014.

4. Thavlov, Anders and Henrik W. Bindner, “Thermal Models for Intelligent Heating of Buildings,” Proceedings of the International Conference on Applied Energy, ICAE 2012. 2012. A10591. [4]

5. Hassan Harba, Neven Boyanova, Luis Hernandez, Rita Streblowa, Dirk Müller, “Development and validation of grey-box models for forecasting the thermal response of occupied buildings”, Energy and Buildings, Volume 117, 1 April 2016, Pages 199-207 [5]

6. Lorenzo Nespoli, Vasco Medici, Roman Rudel, “Grey-box system identification of building thermal dynamics using only smart meter and air temperature data,” 14th Conference of International Building Performance Simulation Association, Hyderabad, India, Dec. 7-9, 2015

7. Petri Hietaharju, Mika Ruusunen, Kauko Leiviskä, “A Dynamic Model for Indoor Temperature Prediction in Buildings,” June 2018 [7]

8. Navid Aghdaei, Georgios Kokogiannakis, Daniel Daly, Timothy McCarthy, “Linear regression models for prediction of annual heating and cooling demand in representative Australian residential dwellings,” Energy Procedia Volume 121, September 2017, Pages 79-86 (Linear regression)

9. Zhang, Kaicheng, Guliani, Akhil, Ogrenci-Memik, Seda, Memik, Gokhan, Yoshii, Kazutomo, Sankaran, Rajesh, and Beckman, Pete., “Machine Learning-Based Temperature Prediction for Runtime Thermal Management Across System Components,” IEEE Transactions on Parallel and Distributed System 2017-07-28 (LAsso)

10. Salah Bouktif, Ali Fiaz, Ali Ouni and Mohamed Adel Serhani, “Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches,” 22 June 2018 (Lstm)

11. G. C. Tiao; G. E. P. Box, “Modeling Multiple Times Series with Applications,” *Journal of the American Statistical Association,* Vol. 76, No. 376., Dec., 1981, pp. 802-816 (n order)

12. David A. Freedman, *Statistical Models: Theory and Practice*. Cambridge University Press, 2009

13. Rencher, Alvin C.; Christensen, William F., "Chapter 10, Multivariate regression,” in *Methods of Multivariate Analysis*. 3rd ed. Wiley Series in Probability and Statistics, 2012

14. Santosa, Fadil; Symes, William W., “Linear inversion of band-limited reflection seismograms”. *SIAM Journal on Scientific and Statistical Computing,* 1986
15. Tibshirani, Robert, "Regression Shrinkage and Selection via the lasso". *Journal of the Royal Statistical Society. Series B (methodological), 1996.*

16. Tibshirani, Robert, “The lasso Method for Variable Selection in the Cox Model”. *Statistics in Medicine, 1997.*

17. Aurelien Geron, *Hands -On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and techniques to Build Intelligent Systems*, O'Reilly Media, 2017

18. Christopher Bishop, *Neural networks for pattern recognition,* Oxford University Press, 1995

19. Woodward, W. A., Gray, H. L. & Elliott, A. C., *Applied Time Series Analysis*, CRC Press, 2012.

20. Sak, Hasim; Senior, Andrew; Beaufays, Francoise, *"Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling,"* 2014

21. Yuyu Zhang, Hanjun Dai, Chang Xu, Jun Feng, Taifeng Wang, Jiang Bian, Bin Wang, Tie-Yan Liu, "Sequential Click Prediction for Sponsored Search with Recurrent Neural Networks," Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014

22. Gabor Petnehazi, Recurrent Neural Networks for Time Series Forecasting, Cornell University, Tech. Report, Jan 2019

23. Klaus Greff; Rupesh Kumar Srivastava; Jan Koutník; Bas R. Steunebrink; Jürgen Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, 2015

24. Sarah Boslaugh, "Chapter 6: The t-test," in *Statistics in a Nutshell*, 2nd ed. O'Reilly Media, 2012

25. M. Zamo, O. Mestre, P. Arbogast, and O. Pannekoucke. "A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production," Solar Energy, 2014.

26. S. Wang and Z. Ma., "Supervisory and optimal control of building HVAC systems: A review," HVAC&R Research, 2008.

27. A. Panagopoulos, "A novel method for predicting the power output of distributed renewable energy resources," Technical University of Crete, Diploma thesis, 2013.

28. A. Panagopoulos, M. Alam, A. Rogers, and N. R. Jennings, "AdaHeat: A general adaptive intelligent agent for domestic heating control," In Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), pages 1295–1303, 2015a.

29. A. Panagopoulos, G. Chalkiadakis, and N. R. Jennings, "Towards optimal solar tracking: A dynamic programming approach," In Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI 2015), pages 695–701, 2015b.

30. Haiad, J. Peterson, P. Reeves, and J. Hirsch, "Programmable thermostats installed into residential buildings: Predicting energy savings using occupant behavior & simulation," Southern California Edison, Tech. Report, 2004

31. P. X. Gao and S. Keshav, "SPOT: A smart personalized office thermal control system," In Proceedings of the 4th ACM International Conference on Future Energy Systems (e-Energy '13), pages 237–246, 2013.

32. Sarah Baker. "Energy Management: simple Steps for Smart Saving." 75f.io https://www.75f.io/blog/energy-management-simple-steps-for-smart-savings

33. Rishav Kumar. "Understanding Principal Component Analysis." medium.com https://medium.com/@aptrishu/understanding-principle-component-analysis-e32be0253ef0

34. "Global Energy Statistical Yearbook 2018." Enerdata.net https://yearbook.enerdata.net/electricity/electricity-domestic-consumption-data.html