Evaluation of whole-building energy baseline models for Measurement & Verification of commercial buildings

WP Robbertse

21076774

Dissertation submitted in partial fulfilment of the requirements for the degree Magister in Mechanical Engineering at the Potchefstroom Campus of the North-West University

Supervisor: Mr. WH Kaiser

November 2016

EXECUTIVE SUMMARY

Title: Evaluation of whole-building energy baseline models for Measurement & Verification of commercial buildings

Author: WP Robbertse

Promoter: Mr. WH Kaiser

Keywords: Measurement & Verification, Whole-building, Energy Baseline, Low Uncertainty, Energy Savings

This dissertation documents the research, development and application of whole-building energy baseline models for the determination of energy savings in commercial- and hospital buildings through the process of measurement and verification (M&V). This study compliments the principles addressed in resources such as SANS 50 010 and the International Performance Measurement and Verification Protocol (IPMVP). The methodology developed in this study requires fitting a baseline model to data from a baseline period and using it to predict what the energy consumption would have been during a subsequent period. The main focus is placed on finding/developing whole-building energy baselines with low uncertainty as it directly affects the size of the savings determined. It was also required to find baseline models that can be used to track performance on small time intervals.

As the baseline model needs to adjust to changing conditions (energy governing factors), it was required to firstly determine the major energy governing factors in commercial buildings. From literature it was found to be ambient temperature as heating, ventilation and air conditioning (HVAC) systems are typically the largest energy consumers. However time-of-day also plays a vital role as occupancy differs throughout the day and different modes of HVAC operation may occur at different times of the day.

Regression modelling is the most common method used to develop baselines for energy use. This creates a relation between the energy use and the energy governing factor. A regression model therefore expresses energy use as a function of the ambient temperature. This single regression of energy use versus ambient temperature delivers very poor results as it does not consider the time factor.

The Day-Time-Temperature (DTT) model was the best performer found in literature. This model addresses both the influences of temperature and the time factor by introducing multiple linear regressions for specific time categories. By analysing the daily load profiles in building energy use, it is possible to identify the different load profiles at different times of the day, which is a good indication of the number of multiple regressions required. Analysing a complete year's profile can become rather intensive especially if the load profiles aren't constant throughout. For this reason it is more effective to create regression models for the smallest possible time categories, in this case hourly.

In literature the DTT model categorized load profiles for each hour of each weekday type to produce 24 x 7 regression models. In this study the load profiles were categorised even further in the search for more accurate models. The load profiles were categorised for each hour of each day type per season (24 x 7 x 4 regression models) and also for each hour of each day type per month (24 x7 x12) as different energy use is expected for cooling and heating requirements each month or each season.

The performance of the 3 DTT model variations were evaluated against a set of testing data from 5 buildings spanning a diversity of building types, climate zones and sizes. The first important finding from the results is that accuracy greatly increases by introducing the time factor. Better correlations were found when regressing in time categories during which only a specific heating or cooling energy is required.

From the results, it is evident that the DTT 24x7x12 model performed the best when using the temperature data from which the regression model was developed, as input into each energy use function. This model performed poorer when using a completely new set of input data which indicates some model instability.

Both the DTT 24x7x1 and the DTT 24x7x4 model remained stable when using a second set of data although the DTT 24x7x1 model fails to properly adjust to temperature variations due to the limitations of a linear regression that was applied to a polynomial correlation. This, of course, leaves room for future work to reduce the number of time categories and instead apply more complex regression models on identifiable cyclical patterns of energy use.

In M&V practice the uncertainty of a baseline always refers to the uncertainty of a model with regards to the actual values from which that model was created, and therefore the uncertainties as determined from year 1's data will typically be used to define baseline uncertainty and reduce the savings achieved. Therefore by knowing which models remain stable with the use of new data it is concluded that the best performing model was in fact the DTT24x7x4 model. This model performed second best in year one, but performed best during year two whilst being able to adjust well to temperature changes and still remain a stable model.

ACKNOWLEDGEMENTS

Thank you to Veritek (Pty) Ltd and the NWU for the opportunity, financial assistance and support to complete this study.

I would like to thank my managers and mentors, Christo van der Merwe and Jan Wiid for all the guidance in measurement and verification and sharing their wealth of knowledge to aid in the completion of this dissertation.

Mr. Werner Kaiser, thank you for the support and guidance, your input on the academic level was immense and influenced the success of this document.

Thank you to my Heavenly Father for the opportunity, the power and perseverance to complete this study.

Table of Contents

List of Tables

List of Figures

NOMENCLATURE

Abbreviation	Meaning
(med(absRTE))	Median of the Absolute Relative Total Error
ASHRAE	American Society of Heating Refrigeration and Air-conditioning Engineers
BMS	Building Management System
CMVP	Certified Measurement & Verification Professional
CV-RMSE	Coefficient of Variation of the Root Mean Square Error
DTT	Day-Time-Temperature
EE	Energy Efficiency
ESM	Energy Savings Measure
FEMP	Federal Energy Management Program
GSEP	Global Superior Energy Performance
HVAC	Heating Ventilation and Air Conditioning
IPMVP	International Performance Measurement & Verification Protocol
LBNL	Lawrence Berkeley National Laboratories
M&V	Measurement & Verification
MBE	Mean Biased Error
nRMSE	Normalised Root Mean Square Error
NWU	North West University
R^2	Coefficient of Determination

SANAS	South African National Accreditation System
SANEDI	South African National Energy Development Institute
SANS	South African National Standard
SARS	South African Revenue Service
SAWS	South African Weather Services
USDOE	United States Department of Energy

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

The South African electricity grid is extremely constrained and requires the aid of the end user to reduce the energy demand to relief some of the load on the National Power Utility's (Eskom) power stations.

Energy awareness and energy efficiency (EE) initiatives and projects are some of the easiest contributions that commercial users (including hospitals) can make. Environmental issues arising from energy use are also reduced by being more efficient in energy use and many companies aim to implement these initiatives to ensure compliance in both aspects.

This has led to the development of several initiatives to encourage industries to use energy more efficiently. One of these initiatives, the 12L Tax incentive which awards energy efficient users with 95 cents per kWh saved, brought forth a market for the evaluation of energy use in order to prove an entity is using energy more efficiently. Energy efficiency is proven by performance tracking one's energy use against an energy baseline. Any difference between the baseline and new energy usage after the implementation of EE projects is considered to be energy efficiency.

Energy-aware users need to prove that their energy conservation measures are however successful not only to justify the costs regarding energy efficient projects and to improve the corporate image but also to qualify for certain national tax incentives which they could greatly benefit from. Energy savings reports should be submitted to the South African National Energy Development Institute (SANEDI), who manages the 12L program on behalf of National Treasury.

Certified Measurement and Verification Professionals (CMVP) are tasked to prove energy savings of customers/end users. Energy savings can be determined by comparing actual energy use, after energy efficiency projects are implemented, with energy baselines that represent what the energy use would have been if no energy efficiency measure were taken.

Baselines are developed from mathematical models that enable the baseline to be adjusted as the independent variable (energy governing factor) changes during the evaluation period. Baseline models are mostly developed from regression modelling as this method shows the correlation between energy usage and certain energy governing factors. CMVP's have to analyse large amounts of building data for the development of commercial building baselines. The accuracy of baselines directly affects the quantity that an end user can claim from the 12L Tax incentive as large model uncertainties decrease the savings.

In this project, the accuracy of various modelling methods will be evaluated for the purpose of integrating the best models with information technologies. This is done to streamline the whole-building M&V process for commercial buildings and hospitals by reducing cost through automation and accuracy.

1.2 PROBLEM STATEMENT

The development of energy baselines for commercial buildings and hospitals can become a very intensive and costly exercise once the building management starts to implement various EE projects which have to be evaluated.

A cost effective whole-building M&V method exists within the specified methodologies of SANS 50 010 to look at the buildings holistically [1]. In commercial buildings and hospitals, it is often difficult to determine the energy governing factors and patterns in operation. In order for M&V professionals to evaluate energy usage, a proper relation between energy use and energy governing factors has to be established.

This study evaluates the factors that drive the energy use in commercial buildings and hospitals and also evaluates various energy modelling methods for baseline development. There is also an ever increasing requirement from clients to develop baselines for very small intervals as energy data is available for half-hourly intervals in this modern age. Baseline models will always have a certain modelling error that affects savings evaluation results as SANEDI deducts the uncertainty percentage from the savings to report conservatively. Clients are already reluctant to appoint independent M&V bodies as this involves an extra project cost just to prove energy efficiency. This study will therefore evaluate various baseline models in order to determine which model(s) provide the best fit and lowest uncertainty in order to maximise the savings gained from EE project implementation.

1.3 AIM OF THIS STUDY

The scope of this study is to review literature specific to the M&V of commercial building's- and hospitals' energy usage. The main focus is placed on the development of whole building energy baselines with low uncertainty, applicable to the aforementioned buildings, whilst complying with national standards and the requirements of SANEDI for the qualification of tax returns related to energy efficiency projects.

1.4 RESEARCH METHODOLOGY

The following methodology is followed to achieve the above-mentioned scope:

Literature review

Literature is reviewed to give information on the commercial sector energy use and what energy governing factors and baseline models can be considered when developing energy baselines for commercial buildings and hospitals. The SANS 50 010 standard for M&V is also reviewed along with the metrics which can determine baseline model accuracies.

Baseline development methodology and evaluation

Various baselines are chosen for evaluation based on the findings from the literature study. The available test data is discussed at the hand of the type of building and climate region it is located in. The baselines are evaluated against statistical metrics to determine the model with lowest uncertainty.

Results & Conclusion

The results obtained from each baseline model with the different sets of available data are compared and a conclusion is drawn to determine the most suitable whole-building baseline model for the measurement & verification process of commercial buildings.

CHAPTER 2: LITERATURE REVIEW

This chapter contains an in-depth literature study that provides background to sudden interest of the commercial sector of South Africa to partake in energy efficiency initiatives. Literature is reviewed to study the various baseline models that are in use for whole-building M&V while still complying with the national standards and legislative requirements. To address the problem statement of this study it is required to find/develop techniques that can provide more accurate baseline models that can be used to draw maximum benefit from any incentive. The main focus of this literature study is to find relevant energy governing factors as well as various baseline modelling techniques and evaluation criteria for model accuracy.

2.1 COMMERCIAL SECTOR ENERGY USE

The commercial sector consists of a range of buildings with varying energy usage. Office buildings are the most common, but the portfolio also includes hotels, hospitals, shopping centres, restaurants, schools, and community centres amongst others.

Energy efficiency improvement opportunities are typically sought where electricity consumption is the highest and improvement is easily achievable. According to the Eskom 2011 integrated report [2], the commerce industry electricity consumption makes up 10% of the total consumption as illustrated in Figure 1.

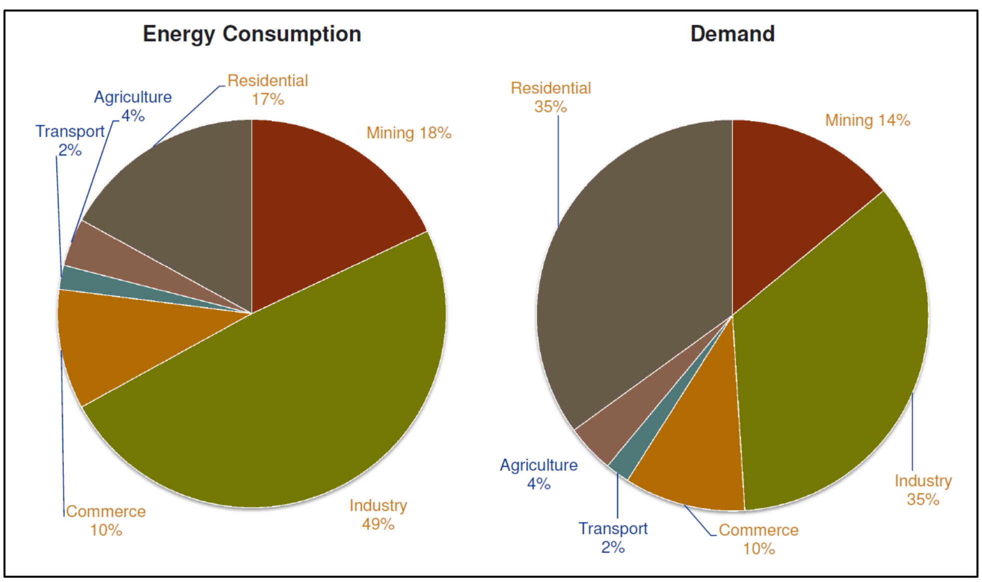

Figure 1 *South African electricity consumption by sector [2]*

A lot of energy efficiency projects have already been implemented in the industry sector as one would expect of the largest consumer. The commercial industry can however also contribute to energy efficiency as they make up a significant portion of the country's energy use, and with the incentive programmes they are urged to participate even more. The major opportunities energy efficiency improvement exist in commercial buildings.

2.2 ENERGY GOVERNING FACTORS IN COMMERCIAL BUILDINGS

In order to develop an energy use baseline one has to determine the factors or independent variables that govern the use of energy in commercial buildings.

The national power utility of South Africa (Eskom) has developed an integrated demand-side management program to promote energy efficiency across all sectors and activities. Various energy saving opportunities were identified in the residential, commercial and industrial sectors.

Energy saving opportunities as identified by Eskom, are used as a starting point for the identification of the energy governing factors in commercial buildings.

As seen in Figure 2 [2], various energy users have been identified as the main drivers of energy use in commercial buildings:

- HVAC
- Lighting
- Motors
- Water heating

The largest single end-use contribution to energy consumption within the Commercial Sector is from Heating Ventilation and Air Conditioning. In addition motors, water heating and lighting contribute significantly to the commercial building energy use.

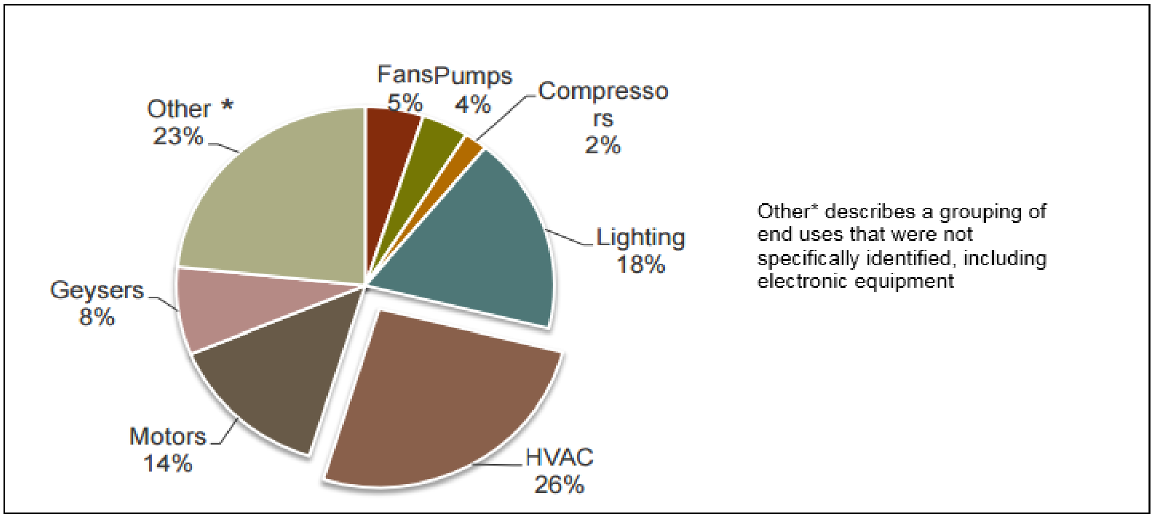

Figure 2 *Commercial end use analysis [2]*

Energy usage in the commercial building sector largely affects the energy efficiency strategies for these buildings. In the U.S. the most significant energy end-use in commercial buildings is lighting, followed by heating and cooling, each contributing to one –seventh of the total. Ventilation uses another 7% of the energy making HVAC as a whole the largest user of energy in commercial buildings at nearly 32% [3]. Although these statistics are not applicable to South Africa due to the difference in climatic conditions, it was selected to emphasize that the HVAC systems in commercial buildings tend to be the largest energy consumer across the board.

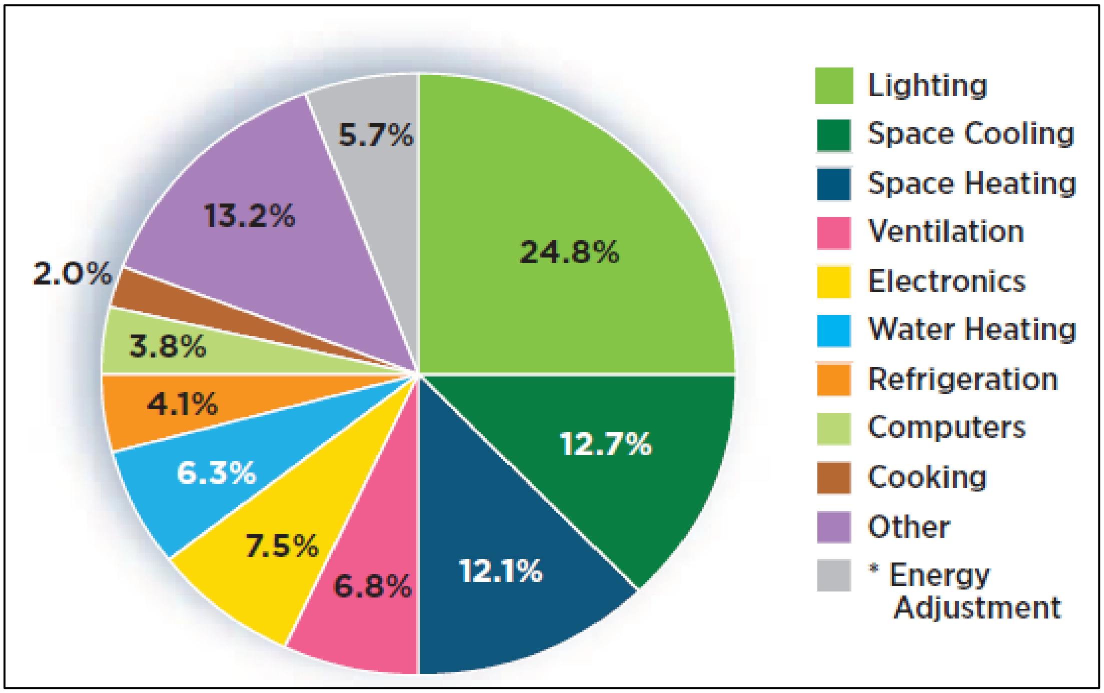

Figure 3 *Commercial primary energy end-use splits [3]*

With HVAC being the largest contributor to energy use in the commercial sector it suggests that ambient temperature is a key energy driver to consider in the whole-building approach. This is in agreement with research done by NorthWrite for the United States Department of Energy (USDOE) [4]. It is common practice to use Degree Days to relate HVAC-driven energy use to environmental energy governing factors, but for this study it is required to develop baselines that can accommodate intervals smaller than daily values as per the problem statement in section 1.2.

In Mathieu, Johanna L. et al [5] it is seen that other important independent variables to consider are time factors. Commercial buildings are typically occupied during certain hours of each day. In many modern buildings, with Building Management Systems (BMS), the HVAC is typically switched on during daily occupancy hours and switched off at night. Therefore building operations would be different at various times of the day and even different days of the week.

In Figure 4 a building load profile for a typical commercial building is shown. In most cases [5], a building will have a clear base load which is seen during the night hours. The power consumption rarely falls below this base load. If a BMS controls the HVAC system it will be switched from night time to daytime operation and depending on the building interior cooling/heating that occurred during the night, the HVAC may be switched on at high power (morning start-up) to bring interior temperatures to set point. As the occupancy increases and the outdoor temperature rises (in the cooling season) the load will increase on the HVAC (morning ramp-up) and at some point of the day peak load will be reached. The HVAC system will switch back to night time operation somewhere during the afternoon or evening (evening set-back) and the energy consumption will decrease accordingly. In some cases, excess consumption persists above the base load leading to an evening shoulder.

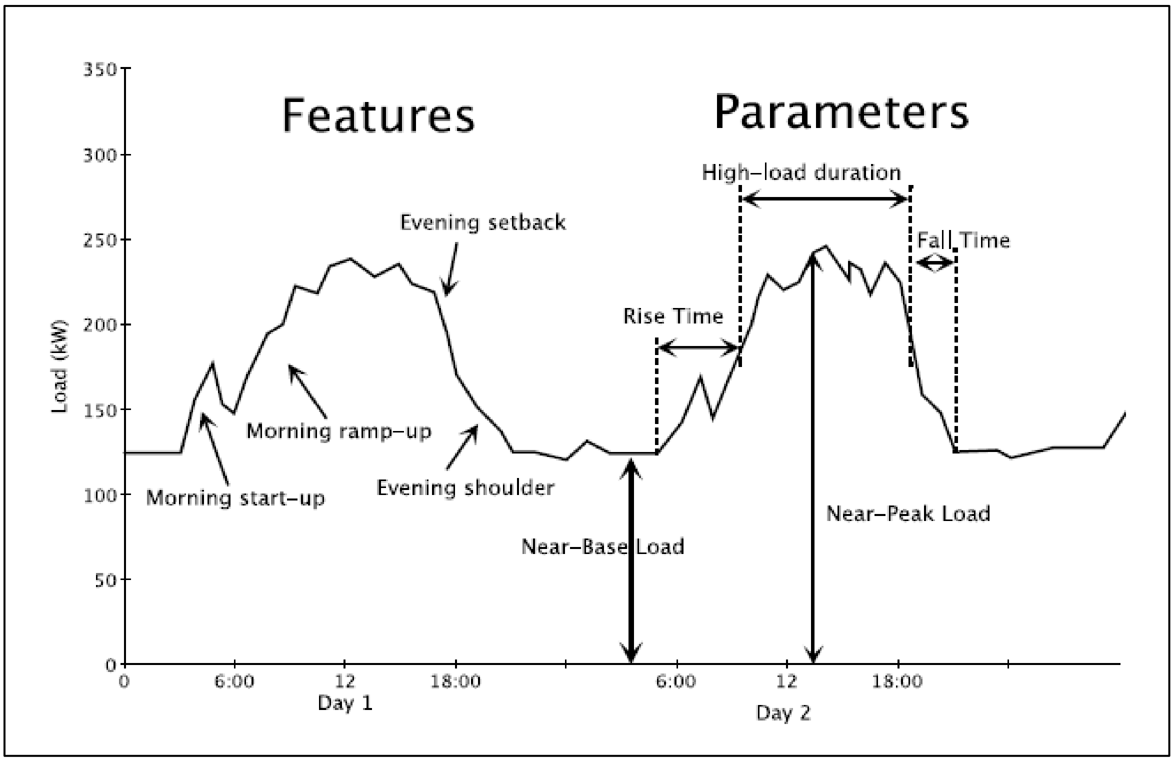

Figure 4 *Building load profile features and parameters [5]*

In the figure above the Near Base Load parameter is defined as the 2.5th percentile of daily load and Near Peak Load parameters is defined as the 97.5th percentile of daily load. High-Load Duration is the time period for which load is closer to near-peak than near-base load. Rise Time is defined as the duration for load to go from near-base load to start of high-load period whereas Fall-Time is the duration for load to go from end of high-load period to near-base load [5].

The above parameters are direct results of building operation and occupation. Occupancy would also govern the use of lighting and other electronic equipment unless a building is controlled by a BMS. In either case the energy use is definitely also time dependent as it relates to operational hours.

The identification of major energy governing factors provides the parameters for baseline development which is a requirement for the evaluation of energy savings and to enable benefits from tax incentives.

2.3 ENERGY EFFICIENCY INITIATIVES: SECTION 12L OF THE INCOME TAX ACT

The constrained South African electricity grid lead to the development of certain legislation to promote efficient use of energy. Efficient energy use will also reduce the effect of greenhouse gasses as most of the South African energy is generated from coal.

Therefore the purpose of the 12L tax incentive is to address the efficient use of energy more than generation of energy [6]. As part of this program proof of savings is required in the form of a SANEDI certificate for the intended purpose of claiming tax returns as provided in section 12L of the Income Tax Act.

The regulation stipulates that an allowable income deduction should be granted in respect of energy efficiency savings by a tax entity in respect of a year of assessment [7].

2.3.1 Objective

The regulation for 12L sets out the process for determining the quantitative energy efficiency savings, and requirements for claiming the proposed tax allowance. The regulation also stipulates a prerequisite that energy savings reports have to be compiled by M&V professionals performing M&V in compliance with SANS 50 010:2011 and the savings certified by SANEDI through issuing of a certificate [8].

SANEDI is a state-owned entity with the main function to direct, monitor and conduct applied energy research and development, demonstration and deployment as well to undertake specific measures to promote the uptake of Green Energy and Energy Efficiency in South Africa [9].

2.3.2 Offerings

According to the regulation, the deduction will be calculated at 45 cents per kilowatt hour of energy efficiency savings, however, the 45 cents per kWh has recently been increased to 95 cents per kWh (2015 budget speech) [10].

The energy efficiency savings have to be measured and confirmed by a South African National Accreditation System (SANAS) certified M&V body and work must be done in compliance with SANS

50 010:2011 as prescribed by regulation. No deduction is allowed if the taxpayer receives a concurrent government benefit in respect of energy efficiency savings.

2.4 ENERGY EFFICIENCY EVALUATION REQUIREMENTS

In this section all requirements for the evaluation of EE projects are discussed in terms of national standards and data credibility. It order to qualify for tax incentives, the South African Revenue Services (SARS) requires that any EE project be evaluated by a SANAS certified M&V body. SANAS certification is achieved by complying with the SANS 50 010 M&V standard.

2.4.1 SANS 50 010:2011

The SANS 50 010:2011 provides the guideline and standard for all M&V activities that would take place during the evaluation of EE projects.

2.4.1.1 Definitions

Measurement and Verification:

According to SANS 50 010:2011 [1] measurement & verification is "the process of quantifying EE savings or the impacts by determination of actual consumption and relevant energy-governing factors, and to develop baselines and baseline adjustments."

Energy Efficiency Savings:

SANS 50 010:2011 [1] describes EE Savings as "the difference between the actual amount of energy used in the carrying out of any activity in a specific period and the amount of energy that would have been used in the carrying out of the same activity during the same period under the same conditions if the energy-savings measure was not implemented." More detail on how M&V is applied to determine energy savings, is given in the section that follows.

2.4.2 M&V APPROACH

In Figure 5 below the baseline period energy refers to the energy that was measured before the implementation of an EE project. The 12L Tax Incentive requires this baseline period to consist of a complete year before implementation. Once a baseline has been developed the Energy Savings Measure (ESM) can be implemented and a decrease in energy use would be expected. After implementation, it is again required to measure the actual energy use during a reporting period of another complete year. The actual energy use measured during the post-implementation stage is then compared to the adjusted baseline. Baseline adjustments are made to simulate what the energy use would have been due to the changes of the most relevant energy governing factors. These adjustment functions are mathematical models which are mostly produced by means of regression modelling.

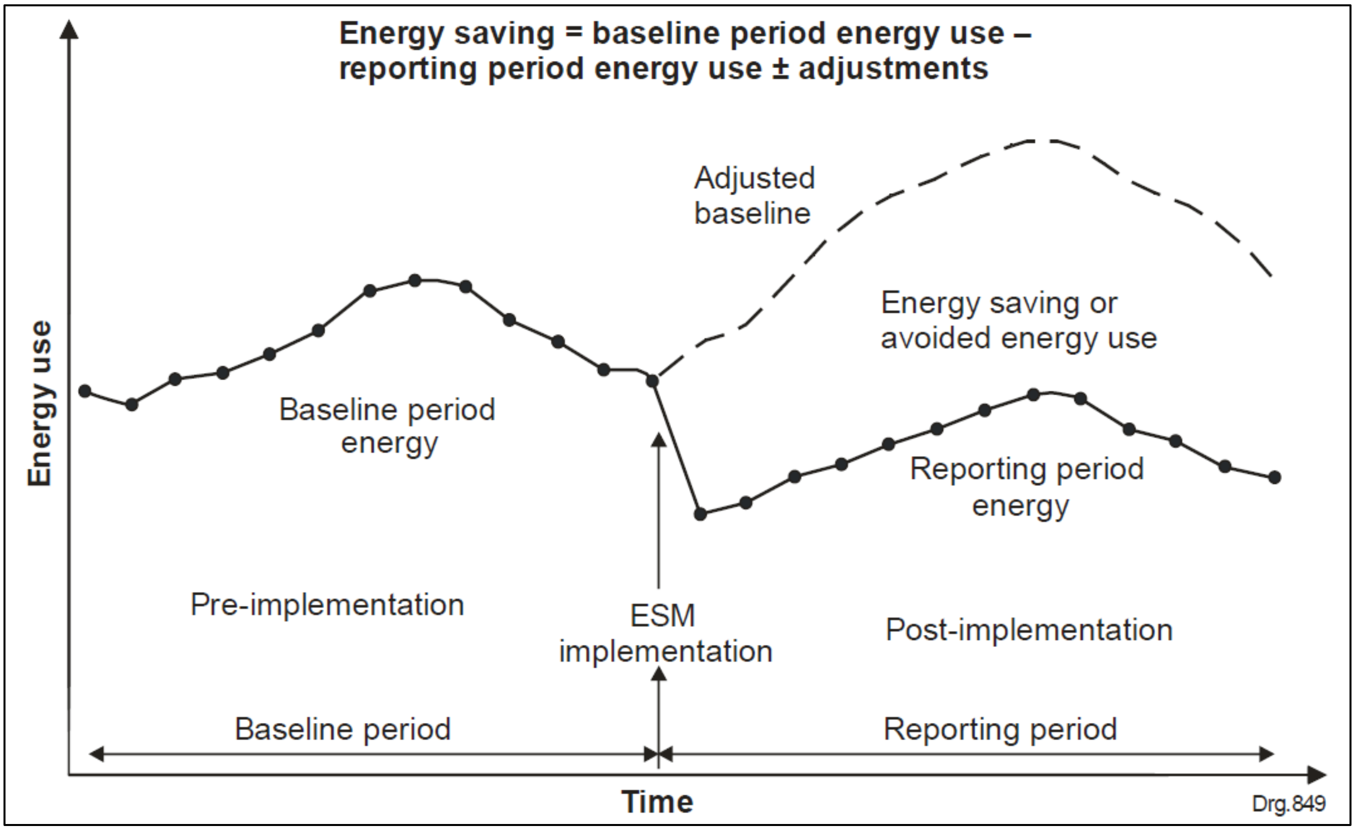

Figure 5 *Overall approach to energy efficiency baseline determination [1]*

Various M&V Approaches are specified by the SANS 50 010:2011 standard. Each approach is developed to suit a system based on the chosen boundary for measurement. The boundary for measurement is normally agreed upon by the CMVP performing the M&V, and the client based on the specific equipment/system that has to be evaluated and the costs involved.

2.4.3 BOUNDARY FOR MEASUREMENT

The SANS 50 010:2011 standard [1] provides four options for determining savings. The option to be used is mainly influenced by the location of the measurement boundary. If only specific equipment has to be evaluated, one of the retrofit isolation options will be used. If a total facility's energy performance has to be evaluated the whole facility option will be used. In the case of unreliable or unavailable data, the calibrated simulation option can be used. The four options for boundary measurements are:

Retrofit isolation: Key-parameter measurement

In the retrofit isolation key-parameter approach, only parameters related to energy use or energy drivers (or both) are measured.

Retrofit isolation: All-parameter measurement

The all-parameter measurement approach measures energy use and all parameters relating to energy drivers.

Whole facility

The whole facility option approaches a facility holistically. A measurement boundary is drawn around the whole facility. Performance and savings will be assessed for the whole facility.

Calibrated simulation

This option is used when baseline or reporting period data are unavailable. A calibrated simulation program shall generate simulated data for either part or all the facility.

This study focusses on the whole facility approach only as it is often the most inexpensive M&V approach.

2.4.3.1 Measurement periods

A measurement period should span a full operating cycle from maximum energy use to minimum. This period should represent all operating modes of the facility.

This measurement period should include only time periods for which all fixed and variable energy-governing factors are known and it should fairly represent all operating conditions of a normal operating cycle.

Measurement periods should coincide with the period immediately before implementation of the energy savings measure to provide a proper baseline for measuring the effect of just the energy savings measure.

The length of any reporting period shall be determined with due consideration of the life of the EE project and the likelihood of degradation of originally achieved savings over time.

2.4.3.2 Calculation of the baseline

Energy savings cannot be measured directly since savings represent the absence of energy use. Energy savings is calculated by comparing measured energy use before and after implementation of energy efficiency projects and making suitable adjustments for changes in energy governing factors.

If an intervention is successful the energy use will be reduced by a certain amount. Comparison between the situation before and after an intervention is achieved through the use of a baseline that characterizes the energy usage based on certain known and/or measurable input variables or patterns.

The savings through energy efficiency can then be calculated by obtaining the difference between the adjusted baseline and the actual energy usage.

2.1.2.6 Basis for the baseline adjustments [1]

Routine adjustments

Energy governing factors expected to change routinely such as production volume or weather are adjusted according to a variety of techniques. Appropriate techniques may be as simple as constant values or as complex as multiple parameter non-linear equation each correlating energy use with independent variables [11].

Non-routine adjustments

Non-routine adjustments provides for energy governing factors which are not usually expected to change such as the design and operation of installed equipment, facility size etc. Throughout the reporting period these static factors have to be monitored for change. When measuring savings under reporting-period conditions, baseline-period energy will be adjusted to the reporting period conditions of energy governing factors.

Measurement of variables

Energy use will be measured either by the direct measurement of energy flow or by the direct measurement of proxies of energy use that give the direct indication of energy use.

One or more of the following techniques will be used to measure the energy quantities in the several forms of the energy savings equation:

- Measuring equipment or energy supplier invoices;
- Special meters to isolate an EE project to a system or a portion of a system or facility from the rest of the system or facility;
- Measurements will be either periodic for short intervals, or continuous throughout the baseline or reporting periods;

2.1.2.7 Uncertainty

Any uncertainty in measurements, sampling and mathematical modelling will be managed to ensure that conservative savings are reported.

Exact quantification of uncertainty will not be required, but uncertainty will be taken into account such that more accurate measurement or a more rigorous M&V process lower the savings that are reported.

Uncertainty management will include:

- any values, whether measured or estimated
- choice of methodology
- baseline period energy use
- reporting period energy use
- energy governing factors considered, and
- an estimation of interactive effects.

In all cases, uncertainty will be caused by modelling errors and the quality of data (measurement errors).

2.4.4 DATA & DATA CREDIBILITY

Data quality is a particularly important aspect for M&V purposes to ensure the credibility of reported energy performance and to ensure SANEDI requirements for valid and verified results are met.

2.4.4.1 Definitions

M&V Data Quality Definition – According to the Global Superior Energy Performance (GSEP) Data Quality Guidance Document [12] "The quality of M&V reporting should be to an accuracy level, confidence level and cost acceptable to all stakeholders involved."

2.4.4.2 M&V data quality level

Statistically, a performance with at least ±7.5% accuracy and 80% confidence is recommended by the GSEP Data Quality Guidance Document however specific project conditions may alter this recommendation. This guide is often used as SANS 50010:2011 only states that uncertainty shall be managed to deliver conservative results and thus in order to guarantee the credibility of energy performance reporting, some characteristics have to be managed. In the next few paragraphs various forms of uncertainty in the M&V process are discussed.

- Instrumentation will always have a certain error in measurement. Measurement equipment errors can be caused by inexact measurements, calibration issues, improper meter operation or installation. According to SANS 50 010:2011 all instruments shall be calibrated as recommended by the equipment manufacturer.
- Sampling errors exist as a result of the variation of values within the population. Sampling should be performed defining a number of measurements taken per unit of time (temporal) or by designating a certain number or percentage of physical energy consuming items (physical).
- Finding mathematical models that fully account for all variations in energy use can become very complex and it is therefore very important to manage. Modelling errors can be the result of an inappropriate functional model, exclusion of relevant models or the inclusion of irrelevant variables.
- Instrumentation errors, sampling errors and modelling errors are combined to deliver the total uncertainty which will decrease the savings achieved by the relevant project.

Energy efficiency projects may be implemented within a system boundary that affects the larger organizational boundary but may not be accounted for in the performance computational methodology.

Managing uncertainty means the data quality should be managed to deliver acceptable results. Mathematical and regression models may be developed and used to optimally establish measurement samples against the benefits generated.

In order to develop baselines that meet SANS 50 010 requirements, all measured data must first be verified from its source by calibration or testing certificates. The M&V whole-building approach will typically require verified building utility meter data and verified data from the relevant energy governing factors. Once the before-mentioned data is acquired and analysed, it can be used to develop the applicable building's energy baseline.

2.5 ENERGY BASELINE MODELLING

2.5.1 METHODOLOGIES USED IN LITERATURE

In a study from Berkeley Labs [13], the methodology utilized to assess the various baseline models comprises of four steps in which a statistical cross-validation approach is used. The model is fit using historical data and then used to develop the adjusted/predicted energy consumption for future periods. Statistical measures are then quantified and compared. This methodology comprises 4 steps:

Step 1: The historical data period and the building characteristics are varied to determine the predictive baseline performance of each model across a range of conditions.

1. *Historical data period – the amount of data used to build the model*

   This study developed models by using metered whole-building hourly data and compared it against weather data for 6-, 9-, and 12-month historical periods.

2. *Building characteristics*

   The data of this study is largely comprised of commercial offices, but also includes small and non-office buildings. For the 16 months of metered data, the operation of the building remained constant, meaning no energy conservation measures have been implemented during that time.

Step 2: The baseline models and adjustments are generated. Electricity consumption data and the associated independent variables are used to develop each baseline model. The data for the independent variable associated with the prediction period are then used as inputs in the models to predict electricity consumption for that prediction period. Independent variables include outside air temperature and time of day or week.

Step 3: The performance of each model is evaluated in this step by means of statistical performance metrics and prediction bounds. Two parameters are varied:

1. *Prediction period*
   In this study only 16 months of data were available. Since 6, 9, and 12 months historical data periods were used, the associated prediction periods were 10, 7 and 4 months respectively.
2. *Predicted quantities*
   Models were developed from hourly interval weather and meter data and then accumulated into daily, weekly and monthly energy use predictions.

Step 4: Statistical metrics are calculated and used to compare the performance of each model.

By following the steps as provided in the study above, it is clear that a relevant baseline modelling technique is to be selected once the data has been collected.

2.5.2 CLASSIFICATION OF BASELINE MODELS

2.5.2.1 Single-variant linear regression model

Building energy research frequently shows that regression analysis has been widely applied to whole-building energy baseline approaches. A regression analysis is a statistical technique that evaluates the relation between variables. The main aim of a regression model is to derive a function that relates a dependent variable to independent variables. The primary study in [14] uses outdoor dry-bulb temperatures to baseline whole building energy use based on monthly utility bills.

A linear function is calculated to express the regression:

$$E = a_0 + a_1 T_0 + \varepsilon \qquad \textit{Equation 1}$$

Where

T_0 = Monthly average outdoor dry-bulb temperatures

E = Energy consumption

ε = normally distributed random number with zero mean and standard deviation

α_0 = intercept

and α_1 = slope

The application of this model is very simple but it provides one linear model for a holistic system. This one model therefore covers a broad range of operations but cannot reflect any small changes and would therefore provide a model with very bad fit to the actual data.

2.5.2.2 Polynomial Models

Polynomial regression models for the whole-building approach will mostly be 2nd order as more energy will be used for high and low temperatures. Polynomial models should be limited to 2nd and 3rd order. In cases where energy use is a function of flowrate, a 3rd order might be most applicable as flow is often a function of ambient temperature and energy use is a function of flow. Considering the relevant energy drivers as identified in section 2.2, only 2nd order models will be applicable in temperature governed commercial buildings [15].

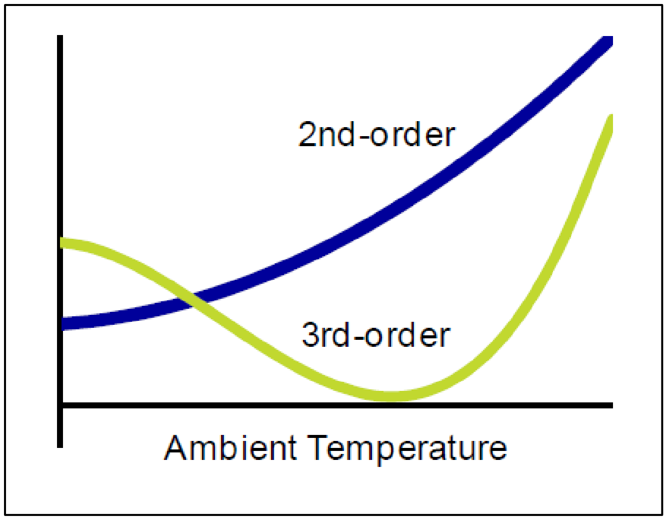

Figure 6 *Polynomial models [15]*

The application of this model is also very simple and provides one polynomial fit model for a holistic system. Although the curved model is a more accurate fit to the actual data, it is still one model to cover a broad range of operations and cannot reflect any small changes and would therefore provide a model with bad fit to the actual data.

2.5.2.3 Mean-week Model

The mean-week model utilizes only day and time to create an energy use profile for the average of the available data that for the exact time of day and day of the week. The average is then regarded as the prediction, for example, the prediction for Wednesday at 4 PM is the average of all the data for Wednesdays at 4 PM.

Once again this is a very simple model to apply but this model does not represent energy use as a function of the energy governing factor and can therefore not be adjusted. Baseline adjustments are required for each new period of evaluation as per SANS 50 010:2011.

2.5.2.4 Change-point models

Systems are often dependent on a certain energy governing factors but only within a certain range of values. In the event of cooling, energy consumption is related to high temperatures which are above the HVAC set point. Heating energy follows the same trend because low ambient temperatures, below the set point, will require heating energy that increases as the ambient temperature decreases. When neither heating nor cooling energy is required, the energy consumption will not be zero as the HVAC system will have a base load requirement to power fans and other related equipment.

In cases like these, change point regression models can be applied to improve the results of simple linear regression. Change-point regression models provide a better fit specifically for commercial buildings such as office buildings. Change points (when heating or cooling energy transitions to base load) can be selected and a linear regression model developed for each temperature dependent zone (heating/cooling or base load).

Figure 7 below shows the change-point models utilized for temperature-dependent loads. The top line illustrates 2-parameter heating and cooling models; the second row illustrates 3-parameter models; the third row shows 4-parameter models; and the last row represents a 5-parameter joined heating and cooling model.

Three-Parameter (3P) Change Point Model

When an independent variable changes over the lower or upper region of the range of the variable and remains constant over the rest of the range, it is appropriate to apply the three-parameter model to predict energy use. This is typically the case with cooling and heating energy use in buildings when temperature varies above or below a certain set point. The 3P models can be seen in row 2 of Figure 7 below and the mathematical models are:

3P Cooling Change-Point Model:

$$Yc = \beta 1 + \beta 2(X1 - \beta 3)^{+}$$ *Equation 2*

3P Heating Change-Point Model:

$$Yh = \beta 1 + \beta 2(X1 - \beta 3)^{-}$$ *Equation 3*

Where Yc/h = energy use

β1 = the intercept

β2 = the parameter defining temperature dependency (slope)

β3 = change-point

(…)- = this parenthetic term’s value is set to zero when negative

(…)+ = this parenthetic term’s value is set to zero when positive

β2 is therefore only applicable for the non-zero condition.

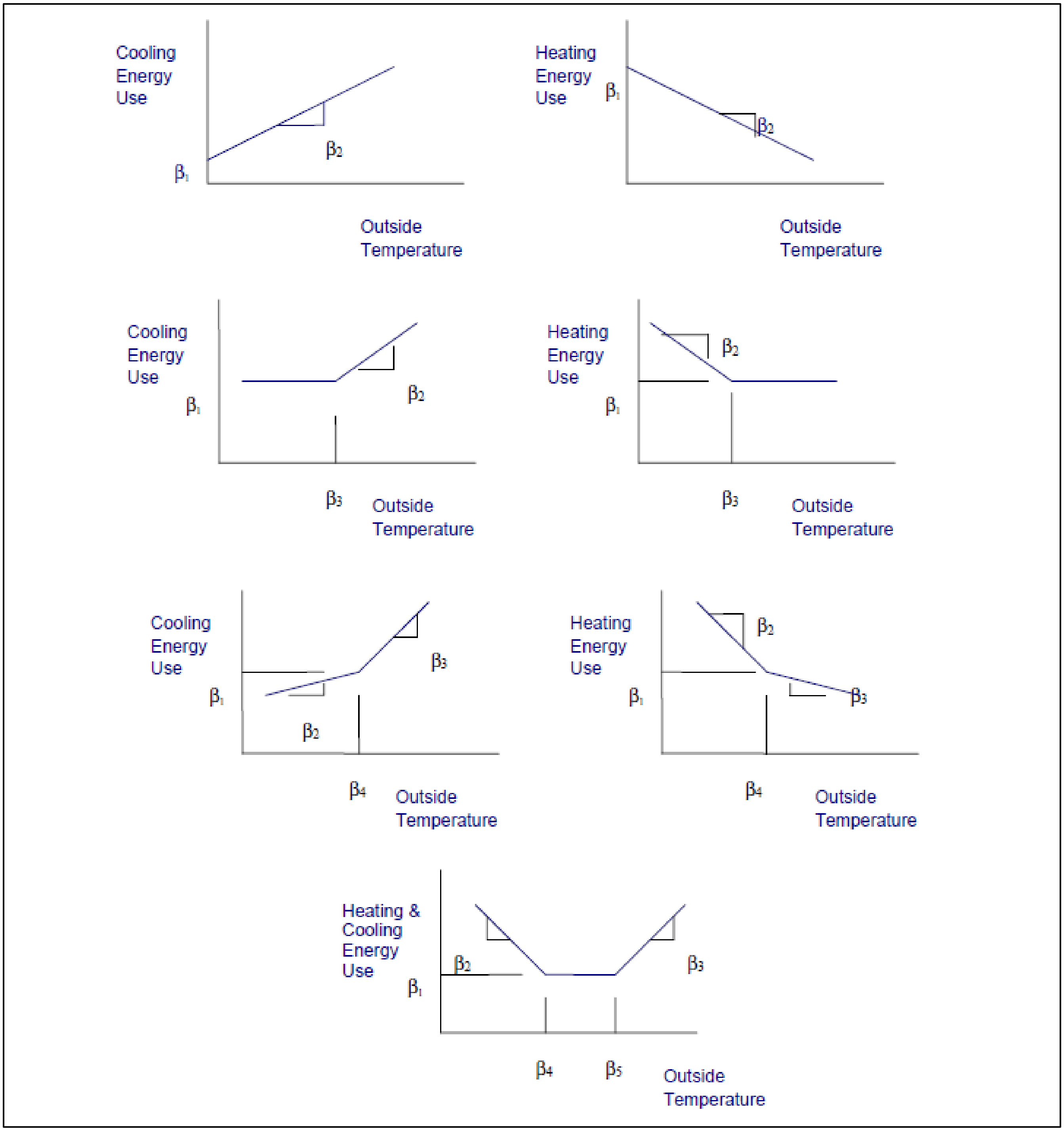

Figure 7 *Change-point linear and multiple-linear inverse building energy analysis models [15]*

Four-Parameter (4P) Change-Point Model

The 4P model, seen in row 3 of Figure 7, also incorporates a change point by adding another non-zero slope to better fit the relationship between heating and cooling energy use with ambient air temperature as an independent variable. Two slopes are often the case as HVAC cooling will include economizer cooling at certain temperatures but only consist of compressor cooling at lower ambient temperatures.

The 4P mathematical model:

$$Y = \beta_1 + \beta_2(X_1 - \beta_4)^- + \beta_3(X_1 - \beta_4)^+ \qquad \text{Equation 4}$$

Where Y = energy use

β_1 = the constant term

β_2 = left slope (heating)

β_3 = right slope (cooling)

β_4 = left change point

$(\ldots)^-$ = this parenthetic term's value is set to zero when negative

$(\ldots)^+$ = this parenthetic term's value is set to zero when positive

The different sides of the model can have either the same or opposite slopes. In most applications, the slope will have the same sign [15] [16].

Five-Parameter (5P) Change-Point Model

The 5P model is used in many of the same situations as the 4P models as it can combine slopes of opposite signs with constant (no cooling/heating) values in the middle. The use of these models is more appropriate for daily data than for hourly data.

The 5P mathematical model:

$$Y = \beta_1 + \beta_2(X_1 - \beta_4)^- + \beta_3(X_1 - \beta_5)^+ \qquad \text{Equation 5}$$

Where Y =energy use

β_1 = the constant term

β_2 = left slope (heating)

β_3 = right slope (cooling)

β_4 = left change point

β_5 = left right change point

$(\ldots)^-$ = this parenthetic term's value is set to zero when negative

$(\ldots)^+$ = this parenthetic term's value is set to zero when positive

Before the availability interval meter data, change-point models were considered the industry standard. Change-point models also relate whole-building energy use outdoor dry-bulb temperature by means of piecewise-continuous temperature response [16].

2.5.2.5 The Lawrence Berkeley National Laboratories (LBNL) Model

The LBNL developed a piecewise linear regression model includes temperature response at specific times of the week. Time of week typically refers to occupied and unoccupied hours of the day.

This load prediction method includes the 2 important energy drivers as discussed in section 2.2: (1) outdoor air temperature and (2) time of week. In this method a week is divided in 15 minute intervals for weekdays (Monday-Friday). A regression coefficient $\alpha_{i\ is}$ then determined for each time interval, which allows for the prediction of the load for each time-of-week interval. It is then also expected that the load will be temperature dependent as high temperatures require cooling energy while low temperatures will require heating energy. "Dead-band" zones will also be expected when no heating or cooling is required. Figure 8 below describes the temperature dependent load.

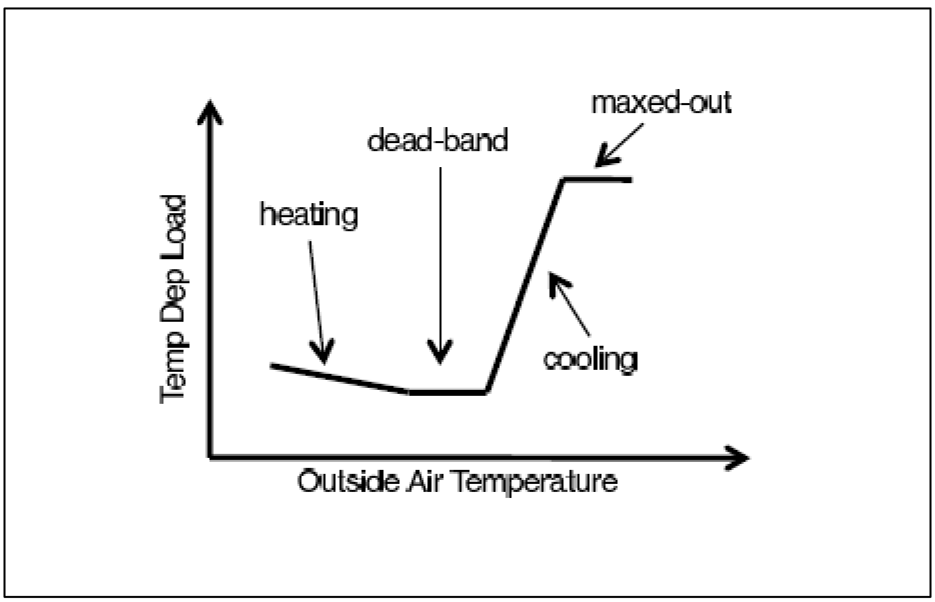

Figure 8 *Temperature-dependent load [5]*

The effect of temperature on load is usually nonlinear and therefore modelled with a piecewise continuous linear model. In this method the outdoor temperatures experienced by a building is divided into six equally-sized intervals. The temperature band (minimum – maximum) that is experienced by the building is divided to produce six temperature parameters, β_j for j= 1 to 6.

The outside temperature at time *t* (which occurs in time of week interval *i*) is broken down into six component temperatures, $T_{c,j}(t_i)$ with *j* =1 to 6, to achieve piecewise linearity and continuity. Each of the six component temperatures is then multiplied by the temperature intervals and summed to determine the temperature dependent-load.

The following algorithm is used to determine component temperatures:

Let B_k (*k*=1…5) be the bound of the temperature intervals.

1. If $T(t_i) > B_1$, then $T_{c,1}(t_i) = B_1$. Else $T_{c,1}(t_i) = T(t_i)$ and $T_{c,m}(t_i) = 0$ for m =2…6, and algorithm is ended.
2. For *n* = 2…4, if $T(t_i) > B_n$, then $T_{c,n}(t_i) = B_n - B_{n-1}$. Else $T_{c,n}(t_i) = T(t_i) - B_{n-1}$ and $T_{c,m}(t_i) = 0$ for *m = (n+1)…6*, and the algorithm is ended.
3. If $T(t_i) > B_5$, then $T_{c,5}(t_i) = B_5 – B_4$ and $T_{c,6}(t_i) = T(t_i) - B_5$.

As most commercial facilities operate different at night or during unoccupied hours therefore the temperature parameters β_j are only used in occupied mode. Occupied and unoccupied modes are manually determined by looking at average load profiles.

For all facilities this method determines occupied load ($\hat{L}o$) as follows:

$$\hat{L}o\big(t_i, T(t_i)\big) = \alpha_i + \sum_{j=1}^{6} \beta_j T_{c,j}\,(t_i). \qquad \text{Equation 6}$$

A single temperature parameter, β_u is used to predict load when the building is in unoccupied mode, since it is expected that most facilities operate at or near the dead-band at night.

Unoccupied load ($\hat{L}u$) is determined as follows:

$$\hat{L}u\big(t_i, T(t_i)\big) = \alpha_i + \beta_u T\,(t_i). \qquad \text{Equation 7}$$

This model differs from other methods that compute regressions for each time-of -day as it captures day-to-day load variation by the addition of the time-of-week indicator α_i and avoids the use of change-point models which leads to complex iterative regression [5].

2.5.2.6 Day-time-temperature models

The DTT models include the temperature energy governing factor as well as the time factor. The ambient temperature is a typical continuous variable which often leads to a 2nd order polynomial model when evaluating a year's energy consumption data as a function of temperatures. This is due to the fact that most buildings in South Africa experience low temperatures that require heating energy in winter and high temperatures that require cooling energy in summers.

In the case of half-hourly or even hourly data, poor correlation can be seen between energy consumption and temperature. When aggregating the data to daily values it becomes clear that a polynomial function is applicable however two different data clusters can be seen often, which indicates two modes of operation. In commercial buildings this occurs due to the fact that the building is used and occupied differently over weekends and holidays.

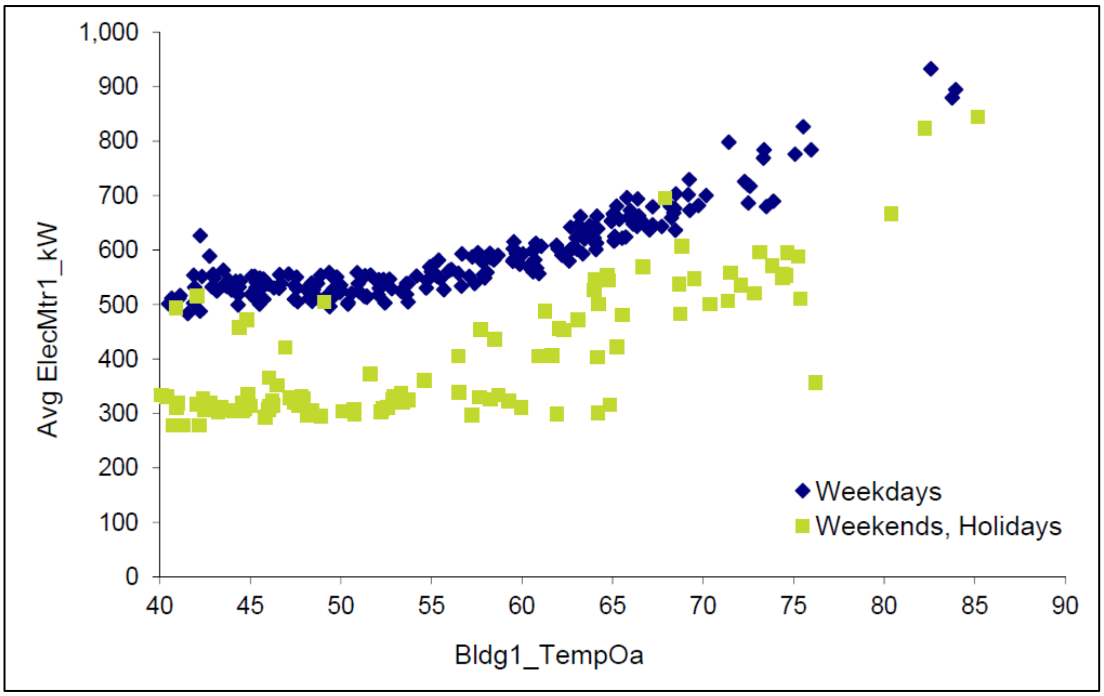

***Figure 9** Daily sample electricity consumption, showing day types [15]*

From *Figure 9* above it is clear that categorical variables (time factors) should also be defined as two different data clusters with little overlap can be seen. In a typical order of importance, the time factors are:

1. Day of week
2. Hour of the day
3. Season of the year [4]

The use of categorical values leads to multiple regression models. A weakness of multiple regression models as in the American Society of Heating Refrigeration and Air-conditioning Engineers (*ASHRAE) RP-1050 Inverse Modelling Toolkit* [16], discussed in section 2.5.2.4, is that these models apply a constant term to a categorical variable resulting in the same slope for all categories when in fact they should be different for some time categories.

The models described in *ASHRAE RP-1050* are widely accepted for M&V purposes [17], and although they consist of multiple regression models (change-point models) they have a number of weaknesses:

1. User interaction is required to select the appropriate model type.
2. HVAC economizer behaviour is not accurately represented by these models.
3. Variable refrigerant flow behaviour is not accurately represented by these models.
4. It does not detect slight increases in energy use.

To address these problems a model needs to be built on historical data from similar conditions and needs to follow the data, not any prior assumption about what shape the model should take as is the case with change point models. In essence, the prediction of "what energy use would have been" is an average of the baseline data that corresponds to those conditions. Regression models like the change point models, attempt to fit a pre-specified shape based on data far from the corresponding conditions. To develop more accurate models, the use of pre-specified shapes should be avoided and data from near the corresponding conditions should be used [4].

Since different operating modes exist at different times of day and week, data near these corresponding conditions need to be evaluated. As already mentioned linear regression model can use categories as well as continuous variables. Categorizing time can create models requiring and providing varying levels of detail.

The information in Table 1 indicates that a good model can be developed from a minimal level of time categorization. The higher the level of time categorization, the more accurate the model due to the fact that more data points are available to follow the corresponding conditions per time interval.

Table 1 *Examples of time categories and models [4].*

Reference Categorization Level	Number of Regressions	An Individual Regression includes	Each Data Point in a Regression Represents
1	1	All data	Daily Energy Use

Reference Categorization Level	Number of Regressions	An Individual Regression includes	Each Data Point in a Regression Represents
2	1	All data	Hourly Energy Use
3	1 to 7	Data for a single day type	Daily Energy Use
4	1 to 4	Data for single Occupancy Period (Occupied, Unoccupied, start-up, shutdown)	Hourly Energy Use
5	7	Data for a single day of the Week	Daily Energy Use
6	24 to 24x7	Data for a single Hour of the Day, for a single day type	Hourly Energy Use
7	7 x 24	Data for a single Hour of the Day, for a single Day of the Week	Hourly Energy Use

Typical regression modelling involves simple linear or even polynomial regression models. Linear regression is by far the simplest method of modelling (apart from the mean-week model) although a linear relationship is not often seen between whole building energy use and ambient temperatures when analysing a whole year's data as this mostly involves slopes with different signs for heating and cooling during winter and summer months. In this case, multiple linear regressions in the form of change-point models might be more suitable, or even polynomial models as this will provide a trend line with an even better fit.

According to the NorthWrite Energy Modelling methods [4] the Day-Time-Temperature models also deliver more accurate results than typical regression:

Table 2 *NorthWrite energy model DTT results [4]*

Time Period	Statistic	Typical Regression	DTT Model
Occupied Hours	R^2 (Higher is better)	0.776	0.833
	CV-RMSE(Lower is better)	6.0%	5.2%
Unoccupied Hours	R^2 (Higher is better)	0.211	0.645
	CV-RMSE(Lower is better)	22.5%	15.1%

As seen in Table 2 above, a baseline model's accuracy or performance will typically be measured at the hand of one or more performance metrics.

2.5.1 BASELINE MODELS PERFORMANCE METRICS

The evaluation criterion in [13] involves statistical performance metrics referred to as 'goodness-of-fit' metrics. Both of these studies were done in compliance with the ASHRAE Guideline 14 [18] and used the metrics deemed most relevant by the guideline for whole-building M&V applications.

In order to select the most appropriate regression model, the goodness of fit is maximized by evaluating the model according to the coefficient of determination (R^2) and the normalized root-mean-squared-error (nRMSE) also known as the coefficient of variation of the root mean squared error (CV-RMSE).

The R^2 value is defined as the Pearson coefficient of correlation which indicates the fraction of the variation in the dependent variable about its mean value [19].

The normalized root mean squared error nRMSE is a metric that quantifies error size relative to the mean of the data. An nRMSE value of 10% indicates that the mean variation in total energy consumption not explained by the regression model is only 10 % of the mean value of total energy consumption [19].

The nRMSE is defined by the equation below:

$$nRMSE = \frac{\sqrt{\frac{\sum_{i-1}^{n}(E_i - \widehat{E_i})^2}{n}}}{\frac{\sum_{i-1}^{n} E_i}{n}}$$

Equation 8

Where E_i is the actual metered energy per unit time, $\hat{E}_i$ is the prediction, and n is the total number of predictions in the prediction period.

In the Berkeley Labs study the median of the absolute relative total error (med(absRTE)) is used to better understand the model prediction error of the total energy consumption. The median tends to be less sensitive to extreme values arising from pathological or unusual cases.

The equation for med(absRTE) is given below:

$$med(absRTE) = median\left\{\frac{|E_i - \hat{E}_i|}{E_i}\right\} \qquad \text{Equation 9}$$

Where E_i is the actual metered energy per unit time, $\hat{E}_i$ is the prediction, and *i* indicates the prediction period.

Further evaluation criterion of Whole-Building M&V are specified by various guidelines such as the Federal Energy Management Program (FEMP) 3.0 [20] and ASHRAE Guideline 14 [18] for both hourly and monthly data, simulated from the developed baseline model.

Table 3 *Baseline evaluation criteria for simulated data*

Data Type	Indices	FEMP 3.0	FEMP 2.2	ASHRAE G14
Monthly	nRMSE	+/- 10 %	+/- 15 %	+/- 15 %
Hourly	nRMSE	+/- 30%	+/- 25 %	+/- 30 %

According to the IPMVP [21], which is the guideline from which the SANS 50 010 was developed, Regression models should be evaluated by R^2 and CV(RMSE) and then the precision range of the predicted values should be used to determine the range of savings.

The relative precision is:

$$Relative\ Precision = \frac{t\ x\ SE}{estimate} \qquad \text{Equation 10}$$

where t = t-value from the t-distribution

SE = Standard Error

To summarize, literature shows that 3 metrics are mainly used to evaluate regression models whilst the study from Berkeley Labs uses a fourth metric.

The statistical metrics are:

Table 4 *Statistical metrics for the evaluation of energy baseline regression models*

Statistical Metric	Description
R^2	Coefficient of determination
nRMSE/CV(RMSE)	Normalised root mean square error
Relative Precision	Absolute precision divided by estimate
med(absRTE)	Median of the absolute relative total error

The above-mentioned statistics are important to evaluate the "goodness-of-fit" of a model but the relative precision should also be evaluated as this is value directly influences the savings achieved and the aim of this project is to reduce the precision interval.

2.5.2 COMPARING VARIOUS BASELINE MODELS

The evaluation of energy baselines is a subject on which more and more studies are being conducted in order to improve the accuracy of baselines and decrease the error percentage for more allowable savings granted by incentives. Various different baselines models already exist for the whole-building M&V approach. Literature provides certain evaluation methods for commercial buildings energy baselines.

The study by the Lawrence Berkeley National Laboratories [13] introduces a statistical metric to compare the accuracy of energy baseline models for M&V baseline energy savings. This methodology requires fitting a baseline model to historic data and using the model to predict energy use during a subsequent period in the future. This prediction represents what the energy use would have been under the same conditions should no energy efficiency measures have been implemented.

In this study 5 baseline models were evaluated using data from 29 buildings in California. These 5 models include the mean-week model, the change point models, day-time-temperature model, a certain proprietary model and the LBNL model. Several statistical metrics were used to characterize the accuracy of the predictions. The nRMSE and the median absolute relative total error were deemed the most relevant to understanding the error in M&V of building energy savings. In some cases, the best-performer as evaluated by one metric was not the best performer when evaluated by another. This study concluded that the LBNL, Proprietary Model and the DTT were the best performing models with regards to nRMSE.

Table 5 *nRMSE of the best performing models in the LBNL study*

Model	nRMSE
DTT	8-18%
LBNL	8-19%
Proprietary model	10-19%
Change-point models	11-25%
Mean Weak model	12-20%

The DTT model ranged from 8-18%, the LBNL model ranged from 8-19% and the Proprietary model ranged from 10-19% of the mean. The DTT model was also the best performing model with regards to median relative total error. The median absolute percent errors for the models can be seen in the table below.

Table 6 *Med(absRTE) of the models evaluated in the LBNL study*

Model	Med(absRTE)
DTT	3-5%
LBNL	3-6%
Proprietary	4-6%
Mean-week model	3-7%

Model	Med(absRTE)
Change point model	5-7%

2.6 LITERATURE DISCUSSION

The literature review delivered various approaches for the development of holistic energy baselines. In the South African M&V industry, baselines have to be developed in accordance with the SANS 50 010 standard which is derived from the IPMVP. The SANS 50 010 requires that baselines be developed from credible data if a project applies for the 12L tax incentive. Credible data includes energy and weather data from calibrated metering systems.

From the literature it is clear that the major energy governing factor in commercial buildings and hospitals is ambient temperature. It is by far not the only energy governing factor but is definitely the largest factor that can influence a building's energy use holistically. Most other factors will require specific metering which adds to the M&V costs and defeats the purpose of a holistic cost-effective approach. The effects of most other contributors can however be captured by also evaluating the energy use in time categories as building occupancy also affects energy use and specifically the load on HVAC systems.

Various baseline modelling methods have already been developed and with the increasing amounts of available energy data, more accurate models can be developed. In literature the traditional regression methods have been evaluated alongside a few new methods. The new models such as the DTT models and the LBNL model outperformed the traditional models as by having lower uncertainty.

The DTT model focusses on multiple temperature dependent regressions for specific time categories. Similarly the LBNL model makes use of multiple regressions for occupied and unoccupied time categories, but also for specific temperature ranges. The multiple regressions stem from the methods as used in the early Change-Point models but can now be done with far greater accuracy because of the availability of energy data for smaller time categories such as half-hourly billing measurements. From these developments it is obvious that multiple regressions for various time categories and temperature ranges deliver the best results.

The SANS 50 010 specifies that uncertainty should be managed as this influences the range of energy savings. As the SANS 50 010 is derived from the IPMVP and does not provide a specification for the calculation of uncertainty, it is therefore best to determine uncertainty as documented in the IPMVP. This study will however also look at the median of the *absolute relative total error* as this metric better describes the typical error in the prediction of the entire evaluation period and is insensitive to extreme values that often occur from out-of-normal operations or measurements.

The following metrics should be used to evaluate a model uncertainty:

- R^2
- CV(RMSE)
- Med(absRTE)
- Precision interval

It is also worth noting that a project's performance can only be evaluated for the same time intervals as was used in the development of the baseline i.e. if weekly energy values were used to develop the baseline, the performance can only be evaluated weekly. Therefore smaller time intervals can produce more accurate tracking to allow clients to manage their energy consuming activities/systems better.

The performance of an EE project is evaluated against an adjusted baseline. This means that the "goodness-of-fit" of a specific baseline model should also be evaluated after the adjustments have been made in the second year of evaluation. Adjustments would include only the performance period's ambient temperatures.

From the literature, it is evident that the way forward is to develop multiple linear regressions of electricity consumption against ambient temperatures for various time categories and/or temperature ranges and to evaluate each model's performance at the hand of statistical metrics as provided in the IPMVP guidelines for M&V.

The best performer from literature, the DTT method, will be used and further developed in terms of time categories for multiple linear regressions, in search of more accurate baselines.

CHAPTER 3: METHODOLOGY

In section 2.5 it was determined that the best performing energy baseline model is the Day-Time-Temperature model in which multiple linear regressions of energy versus temperature are developed for hourly time categories for each day type. This approach will be followed and furthered by introducing more time categories to increase the accuracy of these energy baseline models. The extra time categories include seasonal and monthly time categories as it is expected to find better correlations between temperature and energy use in these categories because a certain heating or cooling range is isolated.

The methodology developed to assess baseline modelling accuracy comprises of 5 phases as illustrated in Figure 10.

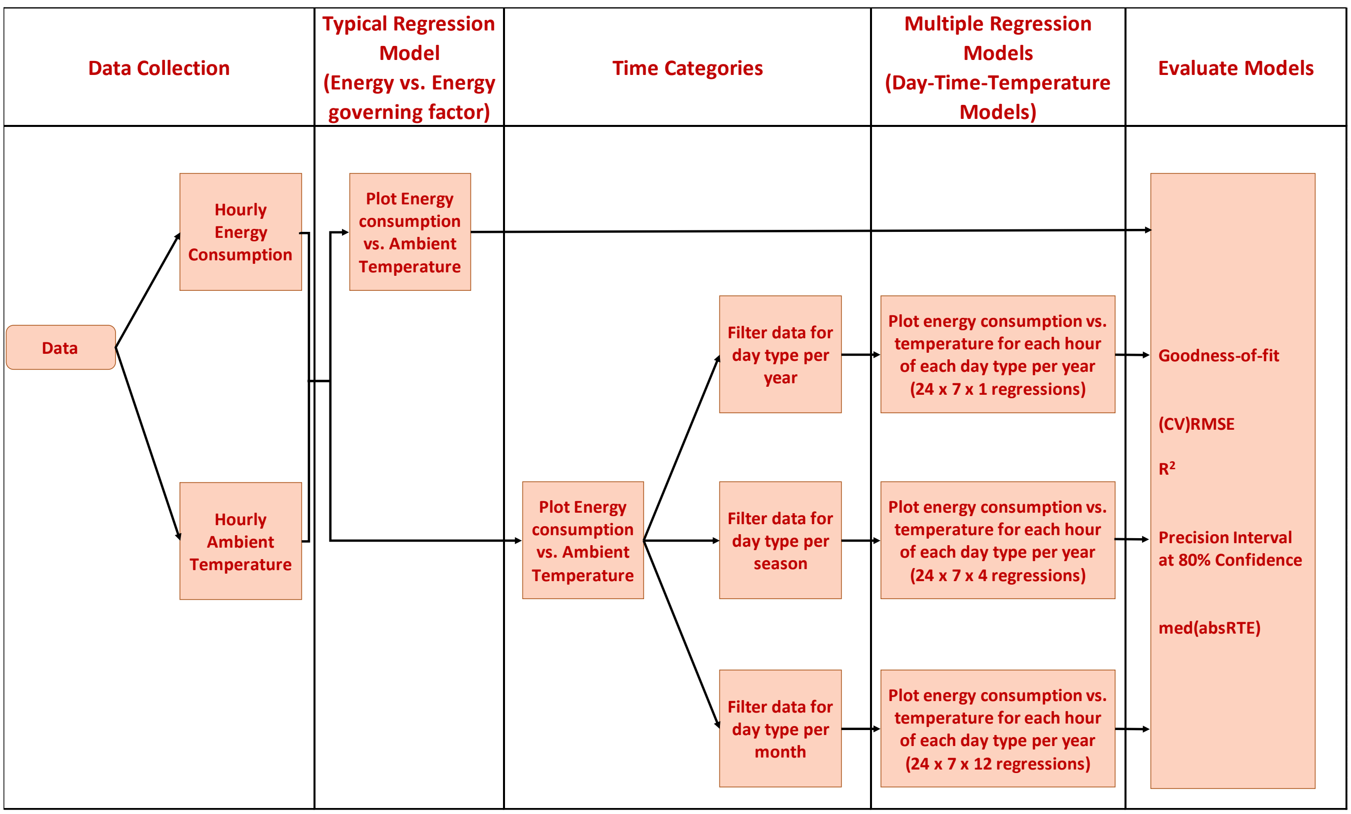

Figure 10 Schematic representation of the 5 phases in which baselines were developed and evaluated.

3.1 DATA COLLECTION

In the data collection phase, two sets of data is entered into a database – hourly electricity consumption in kWh readings as measured by the municipal metering systems and hourly ambient temperatures in °C as measured by local South African Weather Services (SAWS) weather stations.

Both of these data sets are referenced to date and time to ensure that the electricity consumption corresponds to ambient temperature that occurred at any specific date and hour. Data has been collected from five buildings of which three are commercial buildings and two are hospitals. The building types and locations are given in Table 7 below:

Table 7 *Buildings for which data sets were obtained*

Building Reference	Building Type	Location	SANS204 Climate Zone	HVAC Type
Building 1	Hospital	Bloemfontein	Zone 1-Cold Interior	Centralized HVAC
Building 2	Hospital	Johannesburg	Zone 1-Cold Interior	Centralized HVAC
Building 3	Office Building	Johannesburg	Zone 1-Cold Interior	Centralized HVAC
Building 4	Office Building	Johannesburg	Zone 1-Cold Interior	Centralized HVAC
Building 5	Bank (Office Building)	Pretoria	Zone 2-Temperate interior	Centralized HVAC

The thermal inertia of the buildings is not applicable in this study as thermal inertia have not been encountered in any of the literature. Thermal inertia is mostly obtained through calibrated simulations whilst this study focusses on the whole-building methodology. Building sizes and occupancy levels were unavailable as the data used for this study was available from commercial M&V projects.

3.2 DATA PRE-PROCESSING

In order to ensure good data integrity and reliability it is necessary to pre-process the data before introducing the data set to each model. The first requirement is to get the electricity data to correspond to the temperature data. In other words a data set has to comprise of an electricity reading and temperature reading the occurred at the specific time of day.

Metering systems often fail to record for certain time periods and therefore data entries are often missing. This problem is mitigated by creating a chronological timestamp series and then using lookup functions to place the correct entry next to the correct timestamp. Missing data can then easily be identified when timestamps have no entries next to it. This is even more identifiable when plotting the data to create a trend line of the data versus a time line. The trend line will drop to zero and "flat line" for a certain time period during which the meter or logging system malfunctioned.

Certain outliers are also seen when looking at the trend lines. A commercial building will have a certain pattern of energy use and from these patterns a base load and peak load is also easily identifiable. When data points are far above or below the average peak load or base load respectively, it can be regarded as outliers (erroneous data entries or out-of-normal operations).

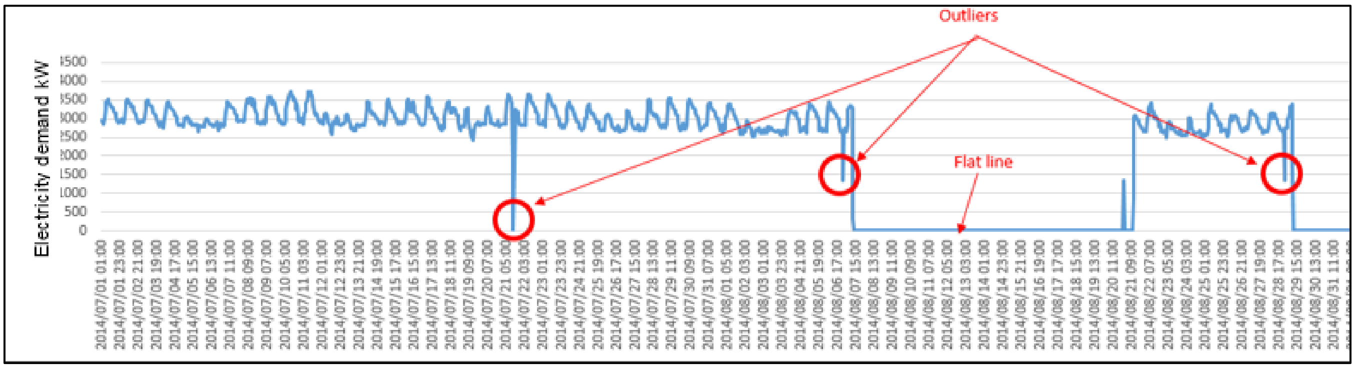

Figure 11 Trend line showing erroneous data entries

The bad data point will be discarded before developing any linear regressions to ensure the data is a true reflection of building normal operations.

3.3 TYPICAL REGRESSION MODELLING

Typical regression modelling is first used to create a relationship between electricity consumption and ambient temperature. This relationship can be either linear or polynomial based on a buildings energy use for heating and cooling. In other words, the electricity consumption is expressed as a function of ambient temperature. This regression model is most commonly used in M&V although it seldom provides accurate results.

This typical regression model is not applied to all data sets as it has been proven in literature to deliver poor results. Instead it is only applied to two data set to confirm the poor results in comparison with other better performing models.

3.4 TIME CATEGORIZING

In the third phase, time is categorized as it is necessary to evaluate the relationship between electricity and ambient temperatures during various periods throughout the day and throughout the year. The various time periods through the year consist of monthly values time intervals as well as seasonal time intervals. In other words, a data set was created in which both energy use and temperature for a single hour of the day for a single day type was listed for a complete year's data. This same data set is then divided into four seasonal categories to create a data set which consist of the energy and temperature data for a single hour of the day for a single day type for each season. Lastly the data is again divided into twelve monthly categories to create a data set which consist of the energy and temperature data for a single hour of the day for a single day type for each month.

This is sensible because a better relationship will be obtained when evaluating time periods with little variation in temperatures for that specific hour. An example of this would be to evaluate a specific hour of the day for summer days only as all temperatures in this time category will be relatively high with little variation. In this study four time categories have been used to develop multiple regression models, these are mentioned in Table 8 below:

Table 8 *Time categorizing for multiple regression models.*

Reference Categorization Level	**Number of Regressions**	**An Individual Regression includes**	**Each Data Point in a Regression Represents**
1	1	All data	Hourly energy use
2	24 x 7	Data for a single hour of the day, for a single day type	Hourly energy use
3	24 x 7 x 4	Data for a single hour of the day, for a single day type, for a single season	Hourly energy use
4	24 x 7 x 12	Data for a single hour of the day, for a single day type, for a single month	Hourly energy use

3.5 MULTIPLE REGRESSION MODELS

In the fourth phase a linear regression is done for each time category as seen in Table 8. The function/regression model for each time category was then used to model the energy use in relation to the accompanying temperatures. Thus the same temperatures from which the models were developed were then used as input for each function to model the energy usage. This modelled energy usage is basically a simulated model of the actual values during baseline period. For savings determination the second year's temperatures is used to adjust the baseline to simulate what the energy use would have been in relation to the accompanying temperatures. The second year's actual measurements can be compared to this adjusted baseline to determine savings.

For the purpose of this study, it is however required to evaluate buildings that have not done anything to achieve savings during the second year as it is required to determine the applicability of the same baseline model for a completely new set of data. It is very difficult to determine whether all conditions of a building remained the same over the period of a year as maintenance which is a routinely activity often affects some technologies and the behaviour of occupants can also differ from one year to the next.

3.6 MODEL EVALUATION

In order to determine how accurately each model predicts the energy usage, it has to be evaluated against the actual data from which it was developed. The goodness-of-fit is evaluated visually (qualitatively) and statistical measures are used to quantify the errors of the models. These measures include the (CV)RMSE, the R^2 value of the actual values vs. modelled values and the precision interval at 80% confidence level. The statistical metrics of these models will then be compared to determine which models perform better.

It was also necessary to adjust each model with new temperatures (test data) to evaluate the applicability of each model after temperature changes. This is a standard practice in M&V and the whole purpose of a baseline – it should be adjustable in the year of evaluation with the relevant energy governing factor to predict what the energy use would have been in the absence of EE measures and under the same conditions. From the data obtained it was only certain that two of the buildings did not implement any EE measures and therefore only the data sets for these two buildings could be used to evaluate how accurate the baseline predicts the energy use of a second year.

Table 9 below summarizes the models developed from each data set:

Table 9 *Summary of the models developed per data set*

Building Reference	Baseline Models Developed (Year 1)	Baseline Models Verified (Year 2)
Building 1	Typical Regression	DTT (24 x 7 x 1)

Building Reference	Baseline Models Developed (Year 1)	Baseline Models Verified (Year 2)
	DTT (24 x 7 x 1) DTT (24 x 7 x 4) DTT (24 x 7 x 12)	DTT (24 x 7 x 4) DTT (24 x 7 x 12)
Building 2	Typical Regression DTT (24 x 7 x 1) DTT (24 x 7 x 4) DTT (24 x 7 x 12)	DTT (24 x 7 x 1) DTT (24 x 7 x 4) DTT (24 x 7 x 12)
Building 3	DTT (24 x 7 x 1) DTT (24 x 7 x 4) DTT (24 x 7 x 12)	None
Building 4	DTT (24 x 7 x 1) DTT (24 x 7 x 4) DTT (24 x 7 x 12)	None
Building 5	DTT (24 x 7 x 1) DTT (24 x 7 x 4) DTT (24 x 7 x 12)	None

After applying this methodology to the data sets of the five buildings, the results for each statistical metric can be compared to find the best performing baseline models. The results are provided in the section that follows.

CHAPTER 4: RESULTS

The results presented in this study focus on comparative model assessment. The metrics which are mentioned in section 3.6 of the methodology are used to understand the error of the various regression modelling techniques, that is, the R^2, CV-RMSE, median absolute relative total error and the relative precision at 80% confidence level.

4.1 COMPARATIVE MODEL ASSESSMENT FOR YEAR 1

The following tables show the results obtained from applying the data of year 1.

Table 10 *R^2 results from simulation with year 1 temperatures*

R^2 of each model per building (%)

Baseline Model	Building 1	Building 2	Building 3	Building 4	Building 5	Average
Typical regression	4.5	20.9	-	-	-	12.7
DTT 24 x 7 x 1	72.9	87.5	77.4	83.7	70.4	78.3
DTT 24 x 7 x 4	86.1	88.7	84.3	91.1	80.4	86.1
DTT 24 x 7 x 12	94.4	93.7	90.3	97	87.2	92.5

Table 11 *CV-RMSE for each model per building (year1)*

CV-RMSE for each model per building in year 1 (%)

Baseline Model	Building 1	Building 2	Building 3	Building 4	Building 5	Average
Typical regression	19.4	20.9	-	-	-	20.2
DTT 24 x 7 x 1	9.9	8.1	10.2	8.0	5.0	8.2
DTT 24 x 7 x 4	7.1	7.5	8.5	5.9	4.1	6.6
DTT 24 x 7 x 12	4.5	5.8	6.7	3.4	3.3	4.7

Table 12 *med(absRTE) for each model per building (year 1)*

Median(absRTE) for each model per building in year 1 (%)

Baseline Model	Building 1	Building 2	Building 3	Building 4	Building 5	Average
Typical regression	25.7	18.4	-	-	-	22.0
DTT 24 x 7 x 1	7.1	4.9	5.6	5.0	3.1	5.2
DTT 24 x 7 x 4	4.3	3.9	4.8	2.5	2.4	3.6
DTT 24 x 7 x 12	2.4	2.6	3.3	1.1	1.7	2.2

Table 13 Relative precision for each model per building (year 1)

Precision at 80% confidence level for each model per building (%)						
Baseline Model	**Building 1**	**Building 2**	**Building 3**	**Building 4**	**Building 5**	**Average**
Typical regression	25.1	25.4	-	-	-	25.2
DTT 24 x 7 x 1	13.0	10.9	13.5	10.7	6.4	10.9
DTT 24 x 7 x 4	9.3	10.0	11.3	7.8	5.3	8.7
DTT 24 x 7 x 12	6.0	7.8	8.9	4.6	4.2	6.3

In the case of a single regression model, the accuracy is firstly assessed by examining the Coefficient of Determination, which is a measure of the extent to which variations in the dependent variable from the mean are explained. In this study the major focus was on multiple regressions of energy use versus the ambient temperature (independent variable). The modelled energy use of each regression model was then summated to form the total modelled energy use and this modelled/predicted energy use was then compared to the actual energy use.

The results in Table 10 show how the models performed when simulating the hourly energy use of each building. With this evaluation metric it could already be seen that typical regression modelling produces a very poor model when developing hourly baseline models and therefore only the data sets of 2 buildings were evaluated to prove the expected outcome. Overall the DTT 24x7x12 model performed the best with an R^2 value that ranged from 87.2 -97%. The R^2 of the DTT 24x7x4 model ranged from 80.4 -91.1% and the DTT 24x7x1 ranged from 70.4-87.5%. There is no universal standard for minimum acceptable R^2 values although the IPMVP suggests a 0.75 minimum value. This value is considered a reasonable indicator of a good causal relationship amongst the energy use and energy driver [11].

The CV-RMSE for each model on hourly energy quantity are summarized in Table 11 above. The CV-RMSE quantifies the size of the error but does so relative to the mean of the data. Overall the DTT 24x7x12 model performed the best with a CV-RMSE value that ranged from 3.29- 6.67%. The CV-RMSE of the DTT 24x7x4 model ranged from 4.06- 8.48% and the DTT 24x7x1 ranged from 4.98- 10.2%.

The median of the absolute relative total error (med(absRTE)) for each model on hourly energy quantity are summarized in Table 12 above. The med(absRTE) is evaluated to understand the error in the prediction of the total energy use over the baseline modelling period. This error is very similar to the *mean absolute percent error* but is less sensitive to the outlying values because the median can better quantify the central tendency. Overall the DTT 24x7x12 model performed the best once again with a med(absRTE) value that ranged from 1.13- 3.25%. The med(absRTE) of the DTT 24x7x4 model ranged from 2.4- 4.81% and the DTT 24x7x1 ranged from 3.14- 7.1%.

Precision is the measure of the absolute or relative range within which the true value is expected to occur with some specified level of confidence. Confidence level refers to probability that the quoted range contains the estimated parameter [21].

The relative precision for each model on hourly energy quantity are summarized in Table 13 above. Using the temperatures of year 1 for prediction, the DTT 24x7x12 model performed the best with a relative precision value that ranged from 4.22- 8.89%. The relative precision of the DTT 24x7x4 model ranged from 5.25- 11.27% and the DTT 24x7x1 ranged from 6.42- 12.99%.

From the results it is also evident that building 3 performed the worst for all metrics, followed by building 1 and 2. This can be due to the fact that building 3 consists of multiple joined buildings, which means that some buildings are not subject to the same condition, such as irradiation and will therefore not always correlate well with ambient temperature. Buildings 1 and 2 are hospitals which often make use of energy intensive equipment and at certain instances energy use will not necessarily correlate well with temperature. It might therefore be worthwhile investigate the inclusion of other factors, such as the number of procedures performed, when developing baseline models for hospitals.

For a further verification of each model, the energy use of building 1 and 2 was again simulated using the following year's ambient temperatures. This data set is regarded as test data where temperature values were inserted into each equation as derived from the multiple regressions using data from year 1. The predicted energy use values were then once again compared to the new energy values of year 2. These results can be seen in the tables below.

4.2 COMPARATIVE MODEL ASSESSMENT FOR YEAR 2

The following tables show the results obtained from applying the data of year 2. This part was done to verify the models by using new temperature data as input for the regression models.

Table 14 *R^2 results from simulation with year 2 temperatures*

R^2 of each model per building (%)

Baseline Model	Building 1	Building 2	Building 3	Building 4	Building 5	Average
Typical regression	-	-	-	-	-	
DTT 24 x 7 x 1	71.1	76.2	-	-	-	73.6
DTT 24 x 7 x 4	80.1	76.1	-	-	-	78.1
DTT 24 x 7 x 12	59.8	70.9	-	-	-	65.3

Table 15 CV-RMSE for each model per building (year 2)

CV-RMSE for each model per building in year 2 (%)						
Baseline Model	**Building 1**	**Building 2**	**Building 3**	**Building 4**	**Building 5**	**Average**
Typical regression	-	-	-	-	-	
DTT 24 x 7 x 1	10.3	13.3	-	-	-	11.8
DTT 24 x 7 x 4	8.2	12.9	-	-	-	10.5
DTT 24 x 7 x 12	14.3	14.7	-	-	-	14.5

Table 16 med(absRTE) for each model per building (year 2)

Median(absRTE) for each model per building in year 2 (%)						
Baseline Model	**Building 1**	**Building 2**	**Building 3**	**Building 4**	**Building 5**	**Average**
Typical regression	-	-	-	-	-	-
DTT 24 x 7 x 1	7.6	8.1	-	-	-	7.9
DTT 24 x 7 x 4	5.0	8.3	-	-	-	6.7
DTT 24 x 7 x 12	5.4	8.5	-	-	-	6.9

Table 17 Relative precision of each model per building (year 2)

Precision at 80% confidence level for each model per building (%)						
Baseline Model	**Building 1**	**Building 2**	**Building 3**	**Building 4**	**Building 5**	**Average**
Typical regression	-	-	-	-	-	-
DTT 24 x 7 x 1	13.8	18.4	-	-	-	16.1
DTT 24 x 7 x 4	10.9	18.0	-	-	-	14.4
DTT 24 x 7 x 12	19.5	18.7	-	-	-	19.1

This second simulation was only done on the data sets of buildings 1 and 2 as these are the buildings that did not undergo changes with regards to energy efficiency improvement. It is therefore assumed that the energy use would remain relatively the same as in year 1.

In this second simulation the best performer with regards to R^2 was the DTT 24x7x4 model with the DTT 24x7x12 performing the worst. These results can be seen in Table 14 The best performer with regards to CV-RMSE was the DTT 24x7x4 model with the DTT 24x7x12 performing the worst. These results can be seen in Table 15. In terms of med(absRTE) error the best performer was the DTT 24x7x4 model with the DTT 24x7x1 performing the worst as can be seen in

Table 16. Lastly, in terms of relative precision at 80% confidence, the best performer was the DTT 24x7x4 model with the DTT 24x7x12 performing the worst as show in Table 17.

From all the results it is evident that the DTT 24x7x4 performs the best when using a complete set of new temperatures to predict what the energy use would have been. This is a good indication that this model is the most accurate while still remaining stable and verifies the DTT 24x7x4 model as the best applicable model.

The performance of these models can also be compared visually by plotting the errors from one model versus the errors from another. With one model's errors on the x-axis and another model's error on the y-axis, any points that fall directly on the 45-degree line indicate that both models performed similar. Points above the 45-degree line indicate that the model on the y-axis had larger errors while points below the 45-degree line indicate that the model on the x-axis had larger errors.

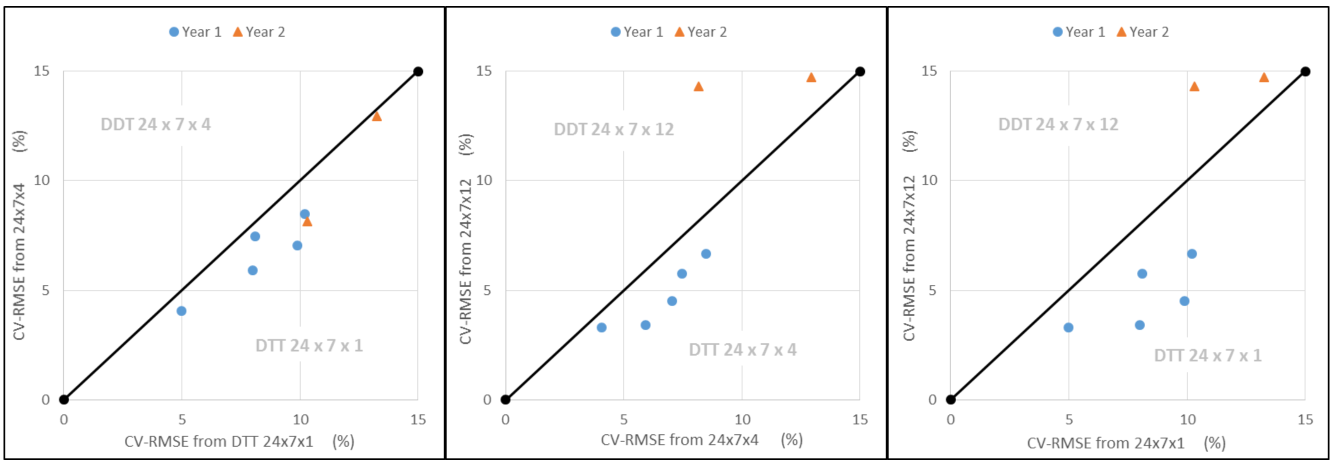

Figure 12 *CV-RMSE of the various DTT models against each other for each building.*

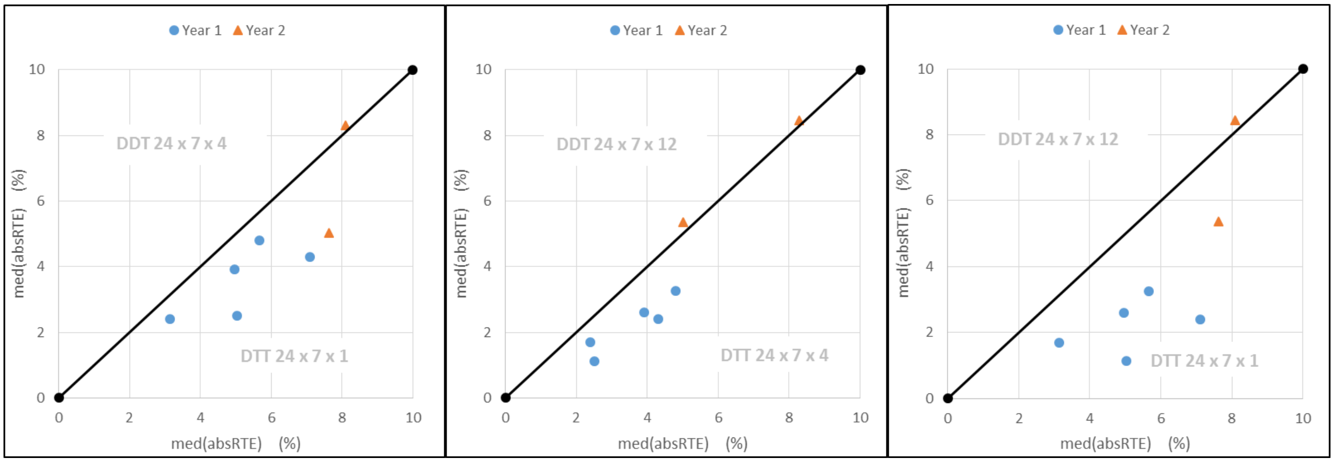

Figure 13 *abs(RTE) of the various DTT models against each other for each building.*

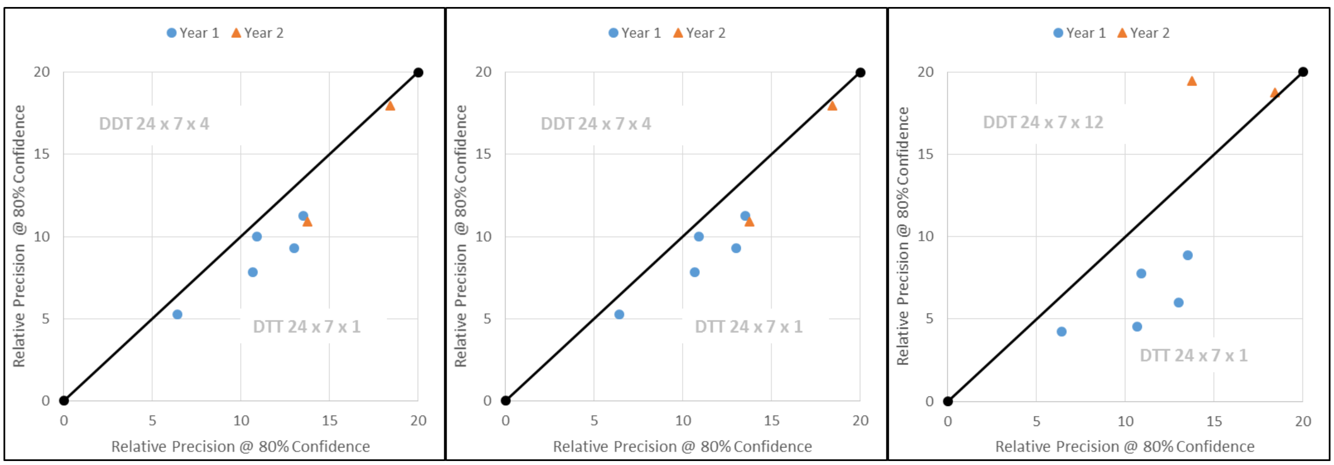

Figure 14 *Relative precision of the various DTT models against each other for each building*

4.3 GOODNESS-OF-FIT

In section 4.1 and 4.2 the different baseline models were evaluated using a quantitative approach to verify these models in terms of criteria as set out in the M&V standard and guidelines. To validate these models, it is necessary to look at these results in qualitative manner. In this approach the actual energy use of each buildings was evaluated against the modelled energy use in a visual manner. This means an evaluation is done on the peaks, valleys and cyclic patterns when comparing the modelled values to the actual values.

The R^2 of each model provides a visual illustration to evaluate the goodness-of-fit as a scatter plot of the actuals versus the modelled values show how each modelled data point deviates from the actual data points. An example of the scatter plots for year 1 and year 2 of the DTT 24x7x12 is shown in Figure 15 below. The full set of scatter plots are provided in Appendix A.

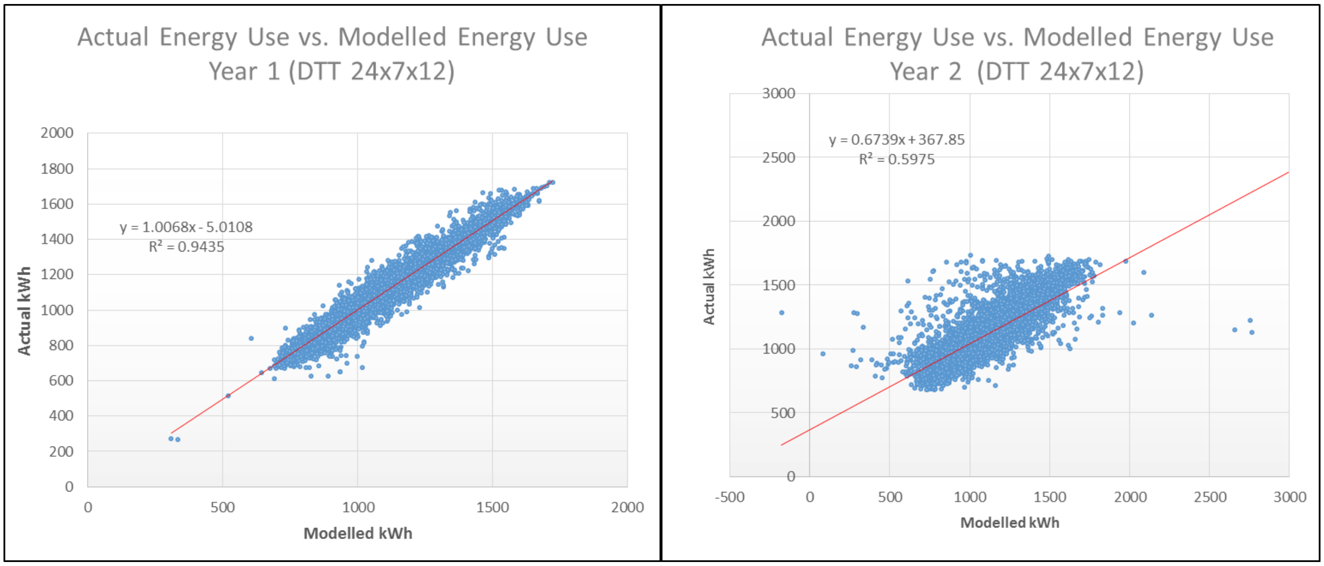

Figure 15 *Scatter plot of actual energy vs. modelled energy for year 1 and year 2 of building 1*

By evaluating the R^2 results it is seen that the DTT 24x7x12 model performs the best for all the buildings when using the same set of data (year 1) that the model was developed from. However this model does not hold up as well when using the data from year 2. Instead the DTT 24x7x4 model performs the best when using the following year's temperatures as input to the model.

The poorer performance of DTT 24x7x12 during year 2 is very likely caused due to too few data points. This model creates a regression model for each hour of each day type per month. This means that each regression model consist of only 4-5 data points as each month typically consists of only 4-5 of each day type. These few data points might not be very representative of the relationship between energy use and temperature. On the other hand the DTT 24x7x4 model creates a regression model for each hour of each day type per season. A season typically consists of 12-13 of each day type. This amount of data points will better represent the relationship between energy use and temperature.

An example of the hourly regressions for the DTT 24x7x4 model is shown Figure 16. Here it is seen that the regression model for this specific hour shows a sensible correlation between energy use and temperature. This correlation consists of 13 data points that accounts for a temperature range from -4°C to 12.5°C.

It is very likely that this temperature range would occur in the following year for this exact time during winter months. From this regression it is more likely to produce an accurate and stable energy use model due to the wider range of temperatures and the amount of data points.

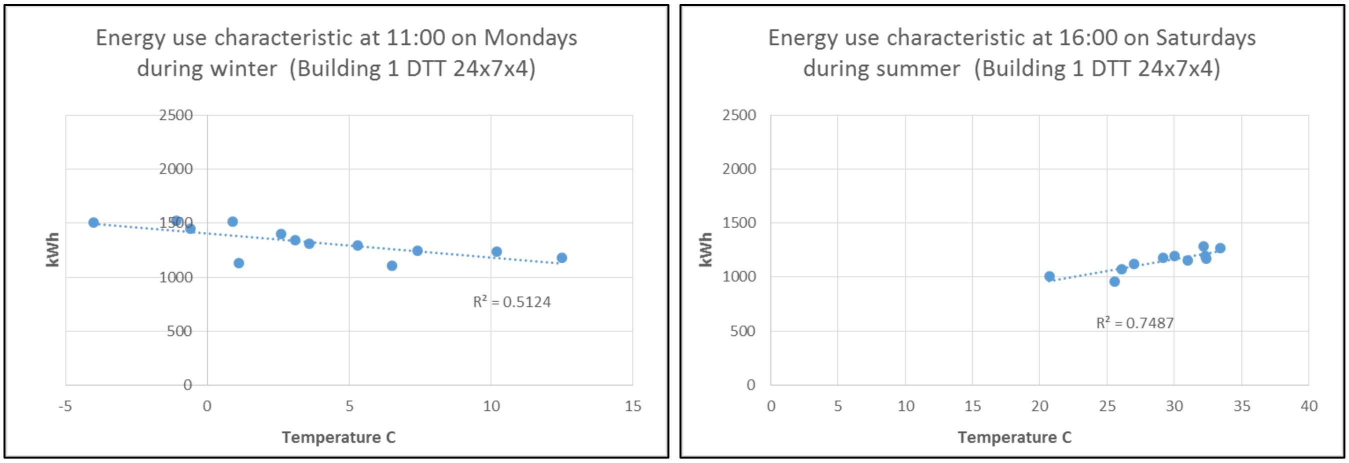

Figure 16 *Hourly regression of DTT 24x7x4*

As the R^2 results have shown, the DTT 24x7x12 model produces the most accurate models when simulating the energy use with the same temperatures that the model was created from. Figure 17 below shows a very good correlation but only for a temperature range of 10 to 21°C. It is very likely that the following year's temperature occurrences are outside this smaller range and that the model does not fully represent what the energy use would have been outside of that range. This can cause the model to predict a complete inaccurate value. Some of these outlying values can be seen in the scatter plot of year 2 in Figure 15.

It is also worth noting that when looking at energy use in very small intervals, in this case hourly intervals, energy use will often not be related to temperature as out-of-normal operations and equipment use can occur, causing a poor correlation between energy use and ambient temperature as can be seen on the right hand side scatter plot of Figure 17. If more points are available for the regression, these out-of-normal operations would clearly stand out as outliers.

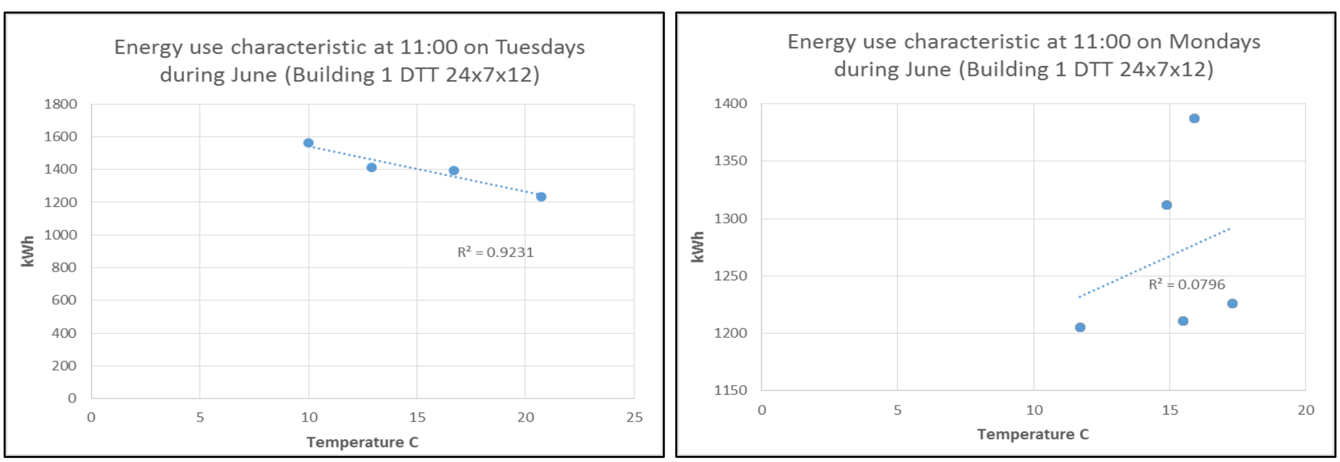

Figure 17 *Hourly regression of DTT 24x7x12*

If these multiple regressions are developed from more data points, as in the case of the DTT24x7x4 model, a truer relationship between energy use and temperature can be found even though some outliers may occur.

When developing multiple linear regressions of even more data points as in the case of the DTT 24x7x1 model, where the energy use of a specific hour for a specific day type across an entire year (52 data points) is evaluated, the model also remains more stable but delivers less accurate results than the DTT 24x7x4. This is due to the fact that the energy use across an entire year's temperature range is not represented well by a linear function but rather a 2nd order polynomial function. Figure 18 below shows an example of the poor linear correlation of the DTT 24x7x1 model on an hourly level.

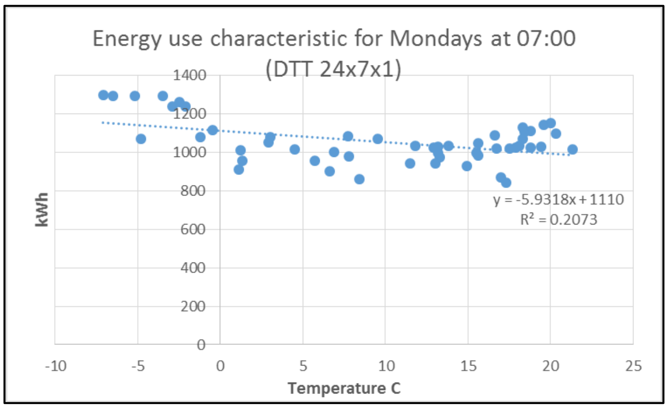

Figure 18 Hourly regression of DTT 24x7x1

Another method for evaluating the goodness-of-fit is to view the energy demand profiles of both the actual and modelled values. Samples of building 1's profiles for each model are given in Figure 19 - Figure 21 below for 1 Apr -17 Apr in year 1.

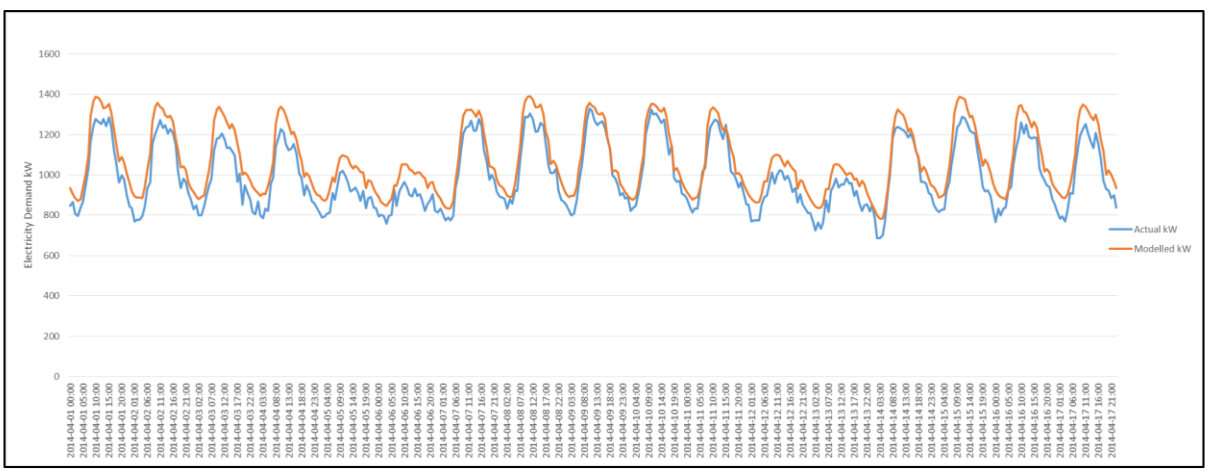

Figure 19 Actual kW vs. Modelled kW for DTT 24x7x1 (Building 1, Year 1)

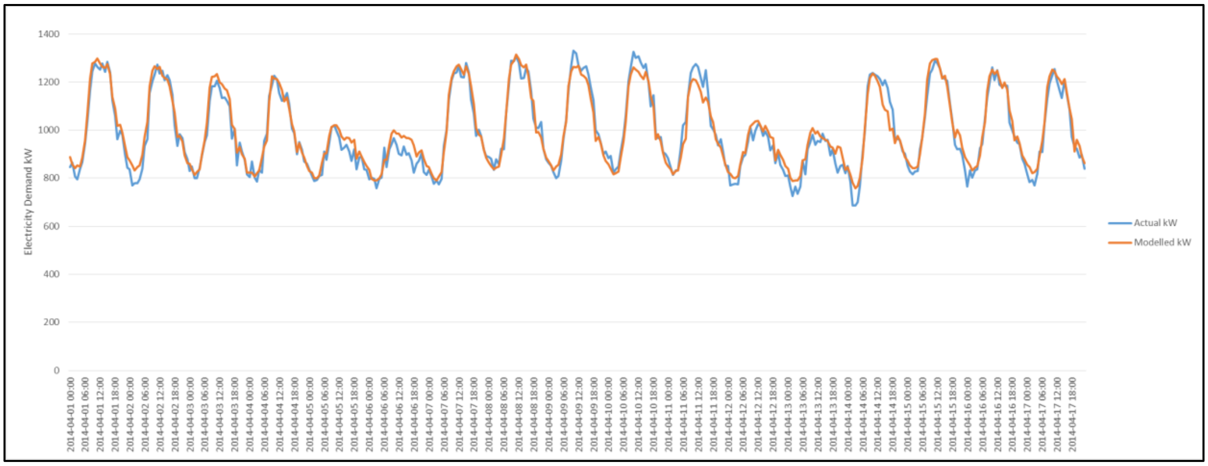

Figure 20 *Actual kW vs. Modelled kW for DTT 24x7x4 (Building 1, Year 1)*

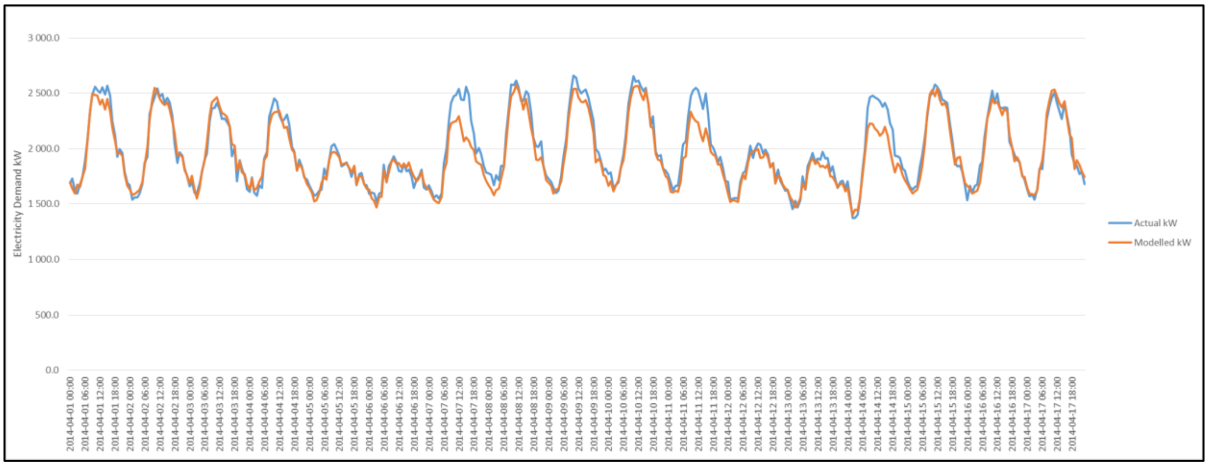

Figure 21 *Actual kW vs. Modelled kW for DTT 24x7x12 (Building 1, Year 1)*

In Figure 19, which is a sample from the DTT 24x7x1 model it is seen that the model follows the cyclic pattern very well, however the model does not accurately meet each peak and valley of the actual profile. This is due to the fact that this model was created from multiple linear regressions developed from an entire year's range of temperatures, which in actual fact can be better represented by a polynomial function as already mentioned in section 4.3. Therefore this model does not variate well as a linear regression over this range of temperatures becomes a rather horizontal line and does not account for very high or low temperatures.

In Figure 20, which is a sample from the DTT 24x7x4 model it is seen that the model also follows the cyclic pattern very well and it meets each peak and valley very well. This model also remains stable throughout the modelling time period.

The DTT 24x7x12 model which is seen in Figure 21 follows the actual profile even closer than the previous two models, but this model is somewhat unstable at certain points because it was developed from a very small range of temperatures and the regression might not be representative of the relationship between energy use and temperature as mentioned in section 4.3.

To further validate these models, the goodness-of–fit was again evaluated when modelling the energy use of year 2. The samples of the actual energy use and modelled energy use profiles of building 1 are given in figure below for 10-26 July 2015. The full set of sample profiles for each building are provided in Appendix B.

In Figure 22- Figure 24 it is seen that the DTT 24x7x1 and 24x7x4 models behave in the same manner during year 2 as in year 1. The DTT 24x7x12 model behaves rather erratically however as the impact of the poor correlation of some of its hourly regressions become very evident now.

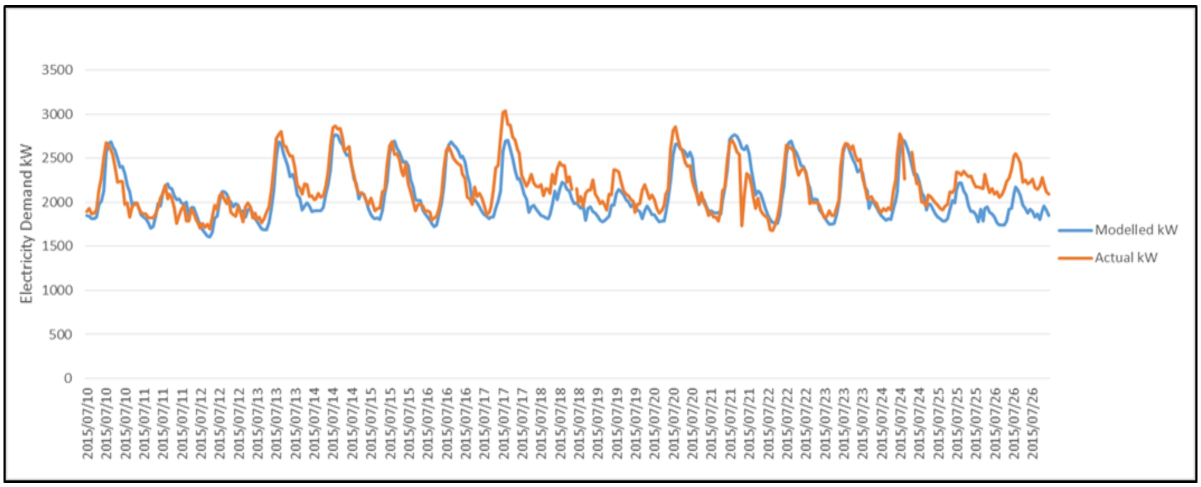

Figure 22 *Actual kW vs. Modelled kW for DTT 24x7x1 (Building 1, Year 2)*

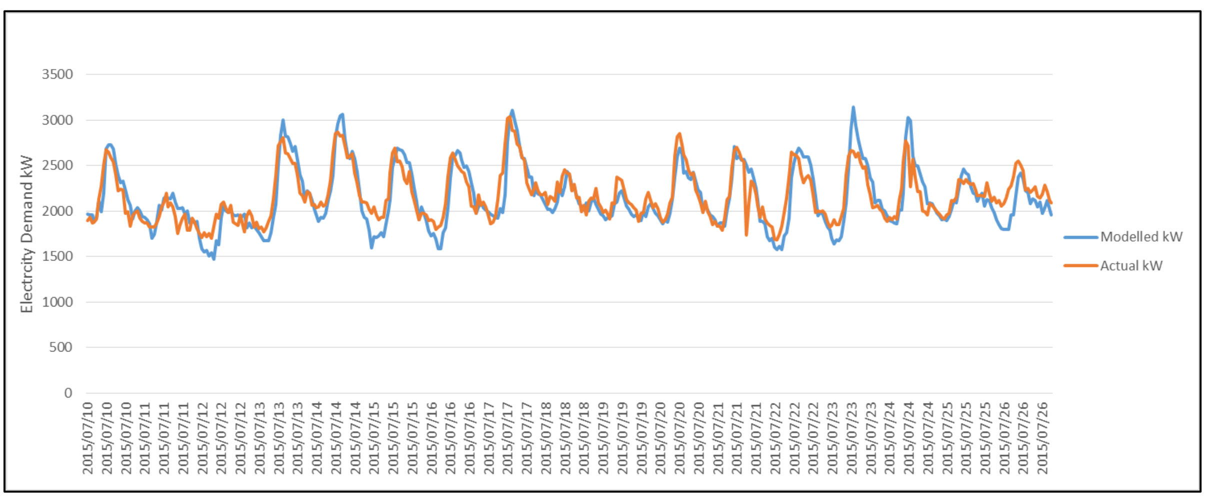

Figure 23 *Actual kW vs. Modelled kW for DTT 24x7x4 (Building 1, Year 2)*

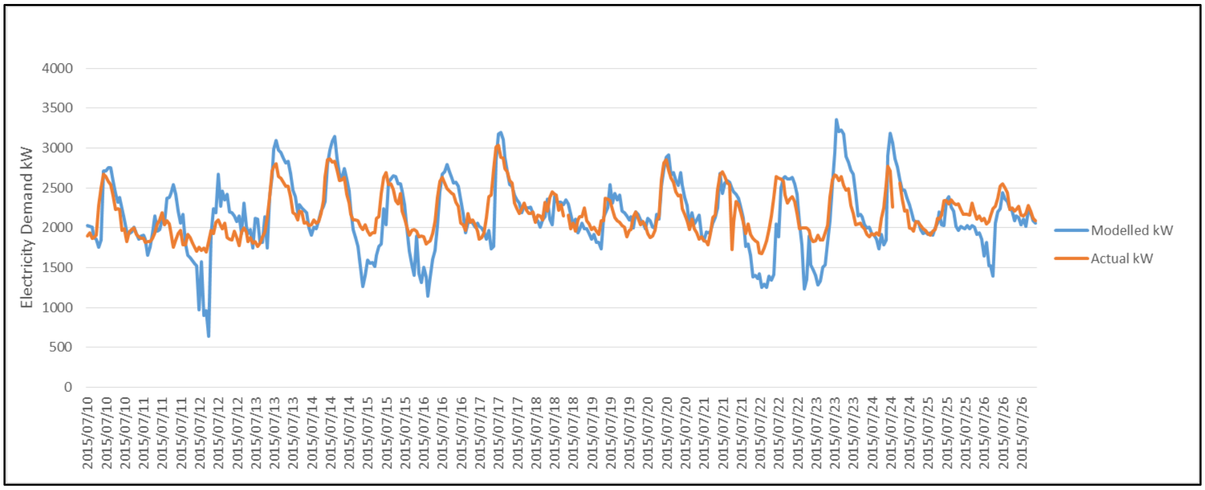

__Figure 24__ Actual kW vs. Modelled kW for DTT 24x7x12 (Building 1, Year 2)

4.4 RESULTS DISCUSSION OF AVERAGE VALUES

The results of this comparative model assessment are summarized and given in Table 18 and Table 19 below.

__Table 18__ Range of statistical metric values for year 1

Year 1

	Range of errors (5 buildings)			
Baseline Model	**R^2**	**CV-RMSE**	**Med(absRTE)**	**Relative precision**
Typical regression	4.5-20.92%	19.41-20.92%	18.38-25.68%	25.07-25.39%
DTT 24 x 7 x 1	70.4-87.5%	4.98-10.2%	3.14-7.1%	6.42-13.5%
DTT 24 x 7 x 4	80.4-91.1%	4.06-8.48%	2.4-4.81%	5.25-11.27%
DTT 24 x 7 x 12	87.2-97%	3.29-6.67%	1.13-3.25%	4.22-8.89%

__Table 19__ Range of statistical metric values for year 2

Year 2

	Range of errors (2 buildings)			
Baseline Model	**R^2**	**CV-RMSE**	**Med(absRTE)**	**Relative precision**
Typical regression				
DTT 24 x 7 x 1	71.1-76.2%	10.31-13.25%	7.62-8.09%	13.77-18.43%
DTT 24 x 7 x 4	76.1-80.1%	8.15-12.94%	5.02-8.29%	10.91-17.96%
DTT 24 x 7 x 12	59.8-70.9%	14.32-14.74%	5.36-8.45%	18.74-19.45%

From the results it is clear that the typical regression method does not perform well at all when modelling hourly energy use. The introduction of the time factor, which requires multiple regression models deliver significantly more accurate results. By modelling the energy use with the temperatures of year1, which is the same temperatures that was initially used to develop the model, the DTT 24x7x12 performs the best with the DTT 24x7x4 being the second best performer. The DTT 24x7x1 model also performs relatively well but does not respond well to temperature changes as we've seen in section 4.3.

When modelling with temperatures of year 2, the DTT 24 x7x12 performs the worst due to the instability caused by the hourly regressions that consisted of only 4- 5 data points with poor correlation. The DTT 24x7x4 model however consists of 12- 13 data points that capture the relationship between energy use and a range of temperatures that occur for each specific hour in each season. The DTT 24x7x4 model is therefore the overall best performer as it remains stable with low errors even when using a second set of data model the energy use.

For M&V projects is very important to determine what the actual savings are for energy efficiency measures. The next section gives an indication of what the energy savings are with the most accurate models.

4.5 BASELINE APPLICATION

In sections 4.1 and 4.2 it was mentioned that data from only building 1 and 2 was used to validate the models with a second set of data as it is known that these buildings have not implemented any projects that affects energy efficiency. Therefore the energy consumption should have stayed relatively the same as in the first year and could be used to determine model applicability.

From the remaining three buildings it is known that building 3 and building 5 implemented energy efficiency measures and it is therefore sensible to apply the best performing baseline models to determine the savings achieved and the applicability of the chosen models.

The savings are determined by subtracting the total actual consumption from the modelled consumption (adjusted baseline):

$$kWh_{savings} = kWh_{adjusted\ BL} - kWh_{actual} \qquad \text{Equation 11}$$

By applying both the DTT 24x7x4 and the DTT24x7x12 models to the data from buildings 3 and 5, the following savings results were obtained (please note the uncertainty refers to relative precision):

Table 20 Savings achieved in building 3 by applying DTT 24x7x4

Adjusted Baseline kWh (6 months)	1378147
Actual kWh	1273426
Savings kWh	104721.3
Uncertainty @ 80 % Confidence	11.27%
Certain savings kWh	92919.22

Table 21 Savings achieved in building 3 by applying DTT 24x7x12

Adjusted Baseline kWh (6 months)	1387568
Actual kWh	1273426
Savings kWh	114142.3
Uncertainty @ 80 % Confidence	8.89%
Certain savings kWh	103995

Here we can see that the DTT 24x7x12 model delivers more savings which can be expected should the model remain stable. However we know from the results in the previous section that the DTT 24x7x12 model often reacts erratically and this can be seen in the case of building 5.

Table 22 Savings achieved in building 5 by applying DTT 24x7x4

Adjusted Baseline kWh	12533758
Actual kWh	11911211
Savings kWh	622546.9
Uncertainty @ 80 % Confidence	5.25%
Certain savings kWh	589863.2

Table 23 Savings achieved in building 5 by applying DTT24x7x12

Adjusted Baseline kWh	12491276
Actual kWh	11911211
Savings kWh	580065.1
Uncertainty @ 80 % Confidence	4.22%
Certain savings kWh	555586.4

In this case we clearly see that, even though the DTT 24x7x12 has a lower uncertainty, it delivers less savings as the erratic behaviour of the model causes the adjusted baseline to be lower at various instances. This once again confirms that the DTT 24x7x4 model is the most applicable model for use in measurement & verification of whole buildings.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

The aim of this study was to find or develop whole-building energy baselines with low uncertainty for the use in measurement and verification of commercial- and hospital buildings. Low uncertainty is required to maximize whole-building energy savings as uncertainty is generally deducted from savings achieved. It was also required to find baseline models that can be used to track performance on small time intervals.

From the literature it was found that the major consumer in commercial buildings is the HVAC system. This indicated that ambient temperature is the most relevant energy governing factor, however time-of-day also play a vital role as occupancy differs throughout the day and different modes of HVAC operation may occur at different times of the day.

Energy use baselines require that a model can be adjusted during a performance tracking period to model what the energy use would have been for the temperature range that occur during the year of performance evaluation. Regression modelling is the most common method used to develop baselines for energy use. This creates a relation between the energy use and the energy governing factor. A regression model therefore expresses energy use as a function of the ambient temperature. This single regression of energy use versus ambient temperature delivers very poor results as it does not consider the time factor.

The DTT model was the best performer (lowest uncertainty) found in literature. This model addresses both the influences of temperature and the time factor by introducing multiple linear regressions for specific time categories. By analysing the daily load profiles in building energy use, it is possible to identify the different load profiles at different times of the day, which is a good indication of the number of multiple regressions required. Analysing a complete year's profile can become rather intensive especially if the load profiles aren't constant throughout. For this reason it is more effective to create regression models for the smallest possible time categories, in this case hourly.

In literature the DTT model categorized load profiles for each hour of each day type to produce 24 x 7 regression models. In this study the load profiles were categorised even further in the search for more accurate models. The load profiles were categorised for each hour of each day type per season (24 x 7 x 4 regression models) and also for each hour of each day type per month (24 x7 x12) as different energy use is expected for cooling and heating requirements each month or each season.

The performance of the 3 DTT model variations were evaluated against sets of data from 5 buildings spanning a diversity of building types, climate zones and sizes. The first important finding from the results is that accuracy greatly increases by introducing the time factor. Better correlations were found when regressing in time categories during which only a specific heating or cooling energy is required.

From the results it is evident that the DTT 24x7x12 model performed the best when using the temperature data from which the regression model was developed, as input into each energy use function. This model performed poorer however when using a complete new set of input data which indicates some model instability. By using training data, the uncertainty (relative precision) ranged between 4.22-8.89% and by using the test data, the uncertainty worsened to 18.74-19.45%

Both the DTT 24x7x1 and the DTT 24x7x4 model remained stable when using a second set of data although the DTT 24x7x1 model fails to properly adjust to temperature variations due to the limitations of a linear regression that was applied to a polynomial correlation. In this study the main focus was placed on the development of multiple linear regressions as this can easily be determined with available software and keep the modelling techniques simple. It is however recommend that further work in this field include modelling with multiple polynomial regressions. This of course leaves room for future work to reduce the amount of time categories and instead apply more complex regression models on identifiable cyclical patterns of energy use.

In M&V practice the uncertainty of a baseline always refers to the uncertainty of a model with regards to the actual values from which that model was created, and therefore the uncertainties as determined from year 1's data will typically be used to define baseline uncertainty and reduce the savings achieved. Therefore by knowing which models remain stable with the use of new data it is concluded that the best performing model was in fact the DTT24x7x4 model as it performed 2nd best in year one, but performed best during year two whilst being able to adjust well to temperature changes and still remain a stable model. By using training data, the uncertainty of the DTT 24x7x4 model ranged between 5.25-11.27% and by using the test data, the uncertainty only worsened to 10.91-17.96%.

From the results obtained, a final recommendation to M&V professionals would be to test a model with test data from a following year as soon as some data is available. If it seems that the model is unstable when using the test data, it would be wise to use a different approach or model for baseline purposes and prevent the discovery of an unstable model during later stages of energy efficiency projects. Should no testing data be available, it is recommended to stratify data from the baseline period, develop a baseline with the remaining data and use the stratified set as test data.

REFERENCES

[1] SANS 50010:2011, *Measurement and Verification of Energy Savings,* SABS, 2011.

[2] Eskom, “Integrated Report,” Eskom Holdings Limited, Johannesburg, 2011.

[3] U.S. Department of Energy, “Energy Efficiency Trends in Residential and Commercial Buildings,” McGraw Hill Construction, 2010.

[4] NorthWrite, “NorthWrite Energy Modeling,” 2015. [Online]. Available: http://www.northwrite.com/propoganda/White%20Papers/NorthWrite%20EnergyModelingWeb.pdf. [Accessed 21 08 2015].

[5] J. L. Mathieu, P. N. Price, S. Kiliccote and M. A. Piette, “Quantifying Changes in Building Electricity Use, with Application to Demand Response,” Berkeley National Laboratory, California, 2011.

[6] SANEDI, “South African National Energy Development Institute: SANEDI, 12L Energy Efficiency Tax Incentive,” 11 03 2014. [Online]. Available: http://www.sanedi.org.za/wp-content/uploads/2014/03/sanedi%20roadshow%20durban/12l_ee_tax_faq.pdf. [Accessed 28 05 2015].

[7] R. National Treasury, *Income Tax Act, 1962 (Act No. 58 of 1962),* National Treasury, Republic of South Africa, 1962.

[8] Acts Online, “Income Tax Act, 1962 (Act 58 of 1962),” 2013. [Online]. Available: http://www.acts.co.za/income-tax-act-1962/index.html. [Accessed 29 05 2015].

[9] SANEDI, “SANEDI Background,” 2013. [Online]. Available: http://www.sanedi.org.za/background/. [Accessed 28 May 2015].

[10] N. Nene, “Budget Speech,” in *National Treasury Budget Speech*, Cape Town, 2015.

[11] Efficiency Valuation Organisation, *International Performance Measurement and Verification Protocol,* EVO, 2014.

[12] GSEP, *Guidance on data quality,* Clean Energy Ministerial, 2014.

[13] J. Granderson and P. N. Price, “Development and Application of a Statistical Methodology to Evaluate the Predictive Accuracy of Building Energy Baseline Models,” Berkeley Lab, Berkely , 2013.

[14] D. Bing, “Baseline Model Development for Commercial Buildings in the Tropics,” National University of Singapore, 2005, 2005.

[15] Research Into Action, Inc.;Quantum Energy Services & Technologies, Inc.; Stetz Consulting, LLC; Kolderup Consulting; Warren Energy Engineering LLC; Left Fork Energy Inc.; Schiller Consulting Inc., “Verification by Energy Modeling Protocol,” Bonneville Power Administration, 2012.

[16] J. K. Kissock, J. S. Haberl and D. E. Claridge, “ASHRAE RP-1050 Inverse Modeling Toolkit,” ASHRAE, 2003.

[17] J. K. Kissock, J. S. Haberl and D. E. Claridge, “Inverse Modelling Toolkit: Numerical Algorithms,” ASHRAE, 2003.

[18] ASHRAE, *ASHRAE Guideline 14-2002, Measurement and Verification of Energy Demand Savings. American Society of Heating Refrigeration and Air Conditioning Engineers,* ISSN 1049-894X, 2002.

[19] A. T. Reddy, N. F. Saman, D. E. Claridge, J. S. Haberl, D. W. Turner and A. T. Chalifoux, “Baselining Methodology for Facility-Level Monthly Energy Use - Part 1: Theoretical Aspects,” ASHRAE Transactions v. 103, Pt.2, 1997.

[20] L. Webster and J. Bradford, “M&V Guidelines: Measurement and Verification for Federal Energy Projects Version 3,” U.S. Department of Energy, 2008.

[21] Efficiency Valuation Organisation, *International Performance Measurement and Verification Protocol,* Toronto, Canada: EVO, 2012.

[22] Department of Energy, “12L Regulations in relation to the national energy efficiency strategy,” in *SANEDI 12L EE Tax Incentive Workshops*, Unknown, 2005.

[23] South African Government, “Government Gazette:,” *Government Gazette: Regulation Gazette No. 10080,* vol. 582, no. 37136, pp. 8-9, 2013.

[24] BIPM, *International vocalbulary of metrology - Basic and general concepts and associated terms,* JCGM 2008, 2008.

[25] SANS, *Code of practice for electricity metering,* Pretoria: SABS Standards Devision, 2009.

[26] L. Grobler, “Energy Savings - Are you on the road to nowhere?,” *Simply Green,* pp. 2-4, April 2011.

Appendix A

This appendix contains the R^2 results of each model applied to the various buildings' data sets. The R^2 of each model also provides a visual illustration to evaluate the goodness-of-fit as a scatter plot of the actuals versus the modelled values show how each modelled data point deviates from the actual data points. The full set of scatter plots are provided below. The first section shows the result of each model's R^2 by using the actual data (year 1) and the second section shows the results when using a second year's data.

A.1 Year 1

A.1.1 Building 1

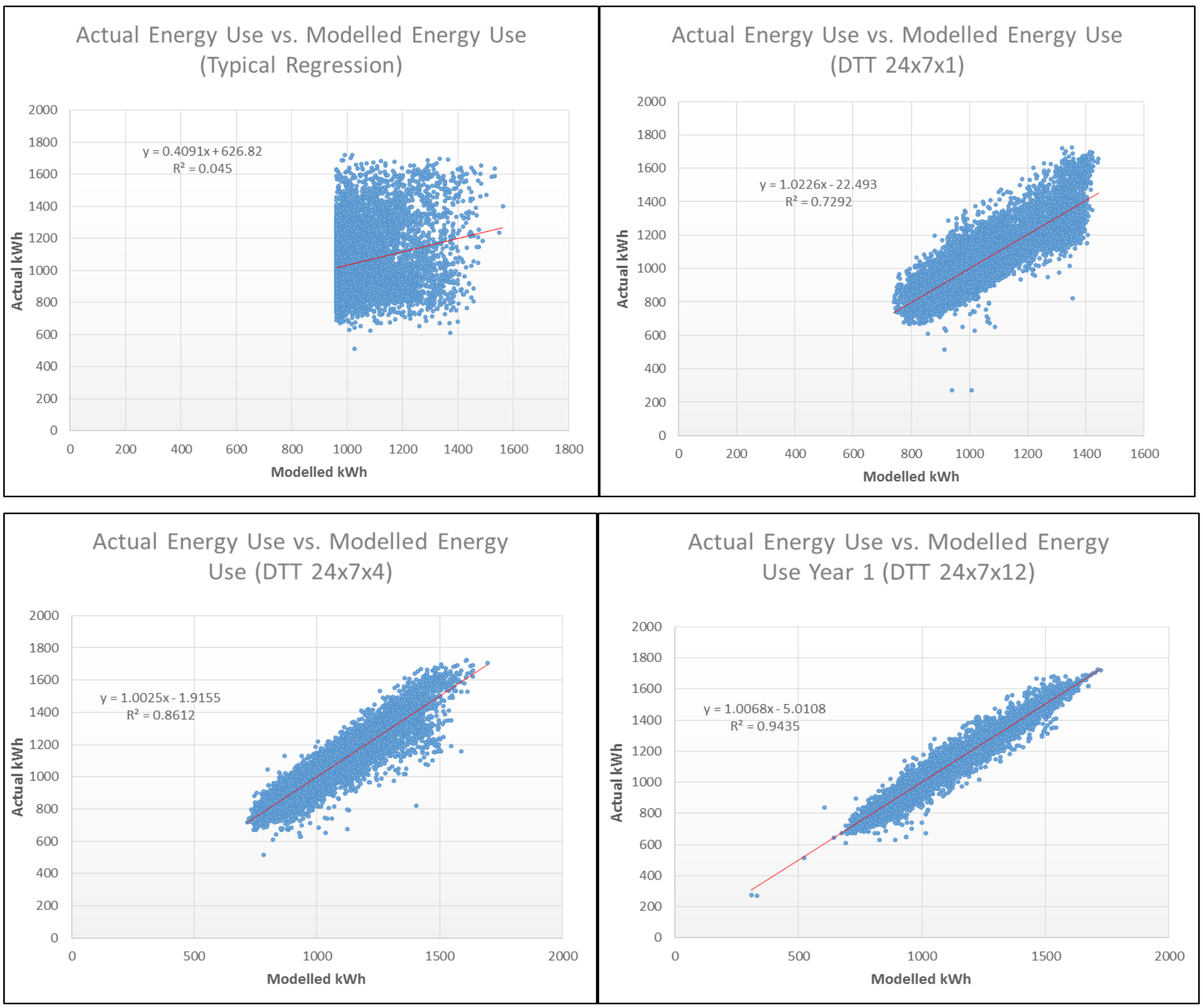

Figure 25 *R^2 of the various models applied to the data set of building 1 for year 1*

A.1.2 Building 2

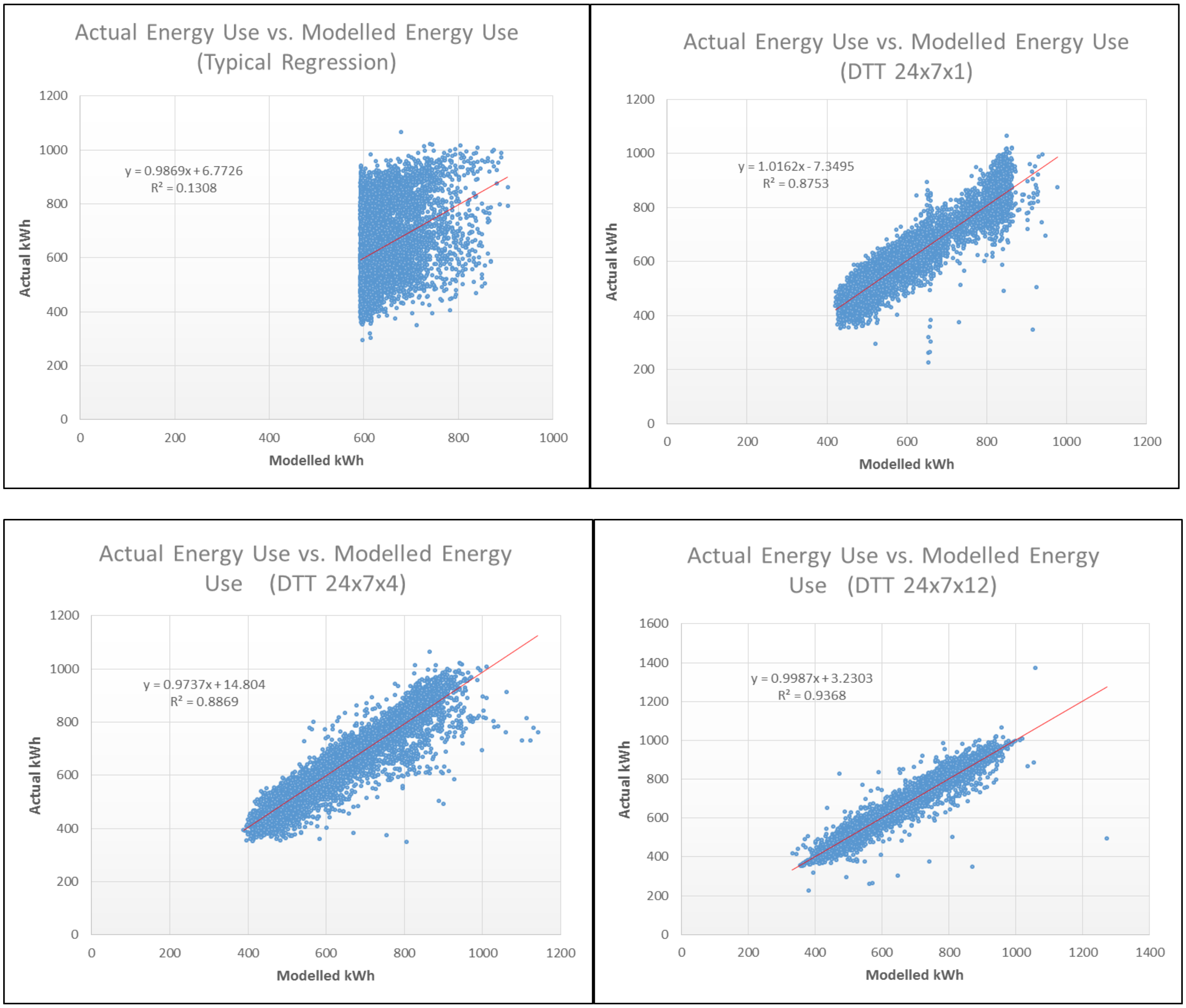

Figure 26 *R^2 of the various models applied to the data set of building 2 for year 1*

A.1.3 Building 3

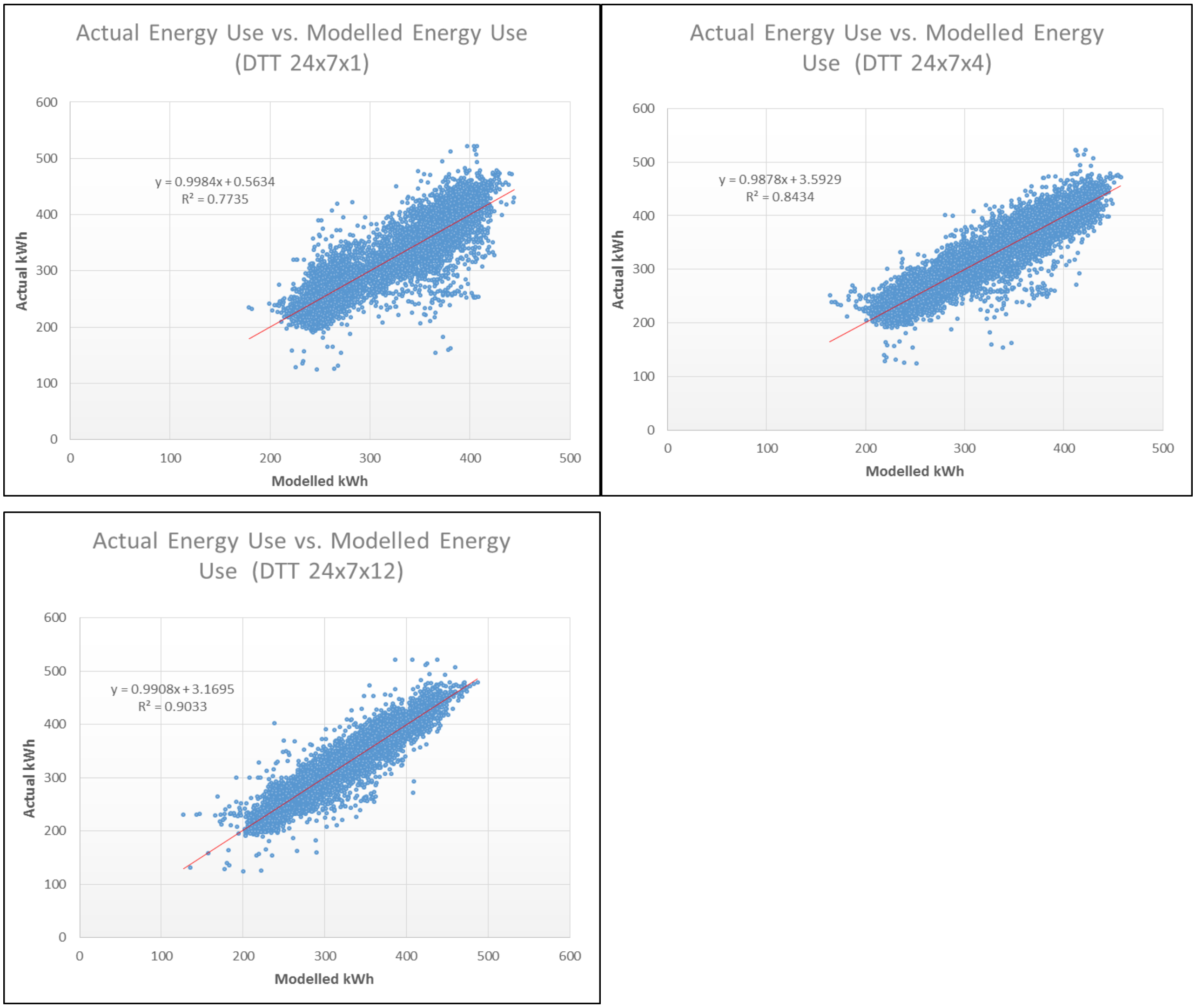

Figure 27 *R^2 of the various models applied to the data set of building 3 for year 1*

A.1.4 Building 4

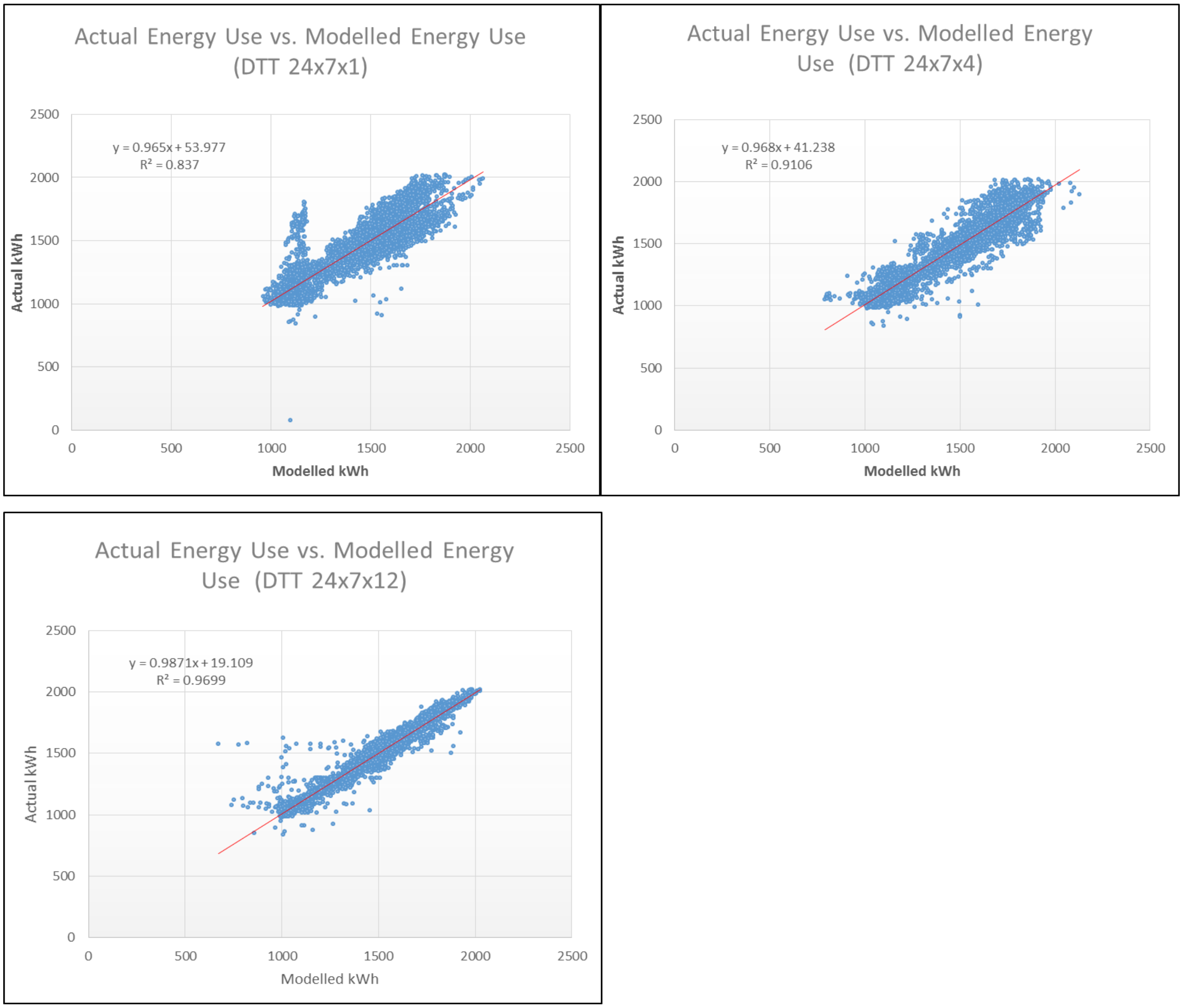

***Figure 28** R^2 of the various models applied to the data set of building 4 for year 1*

A.1.5 Building 5

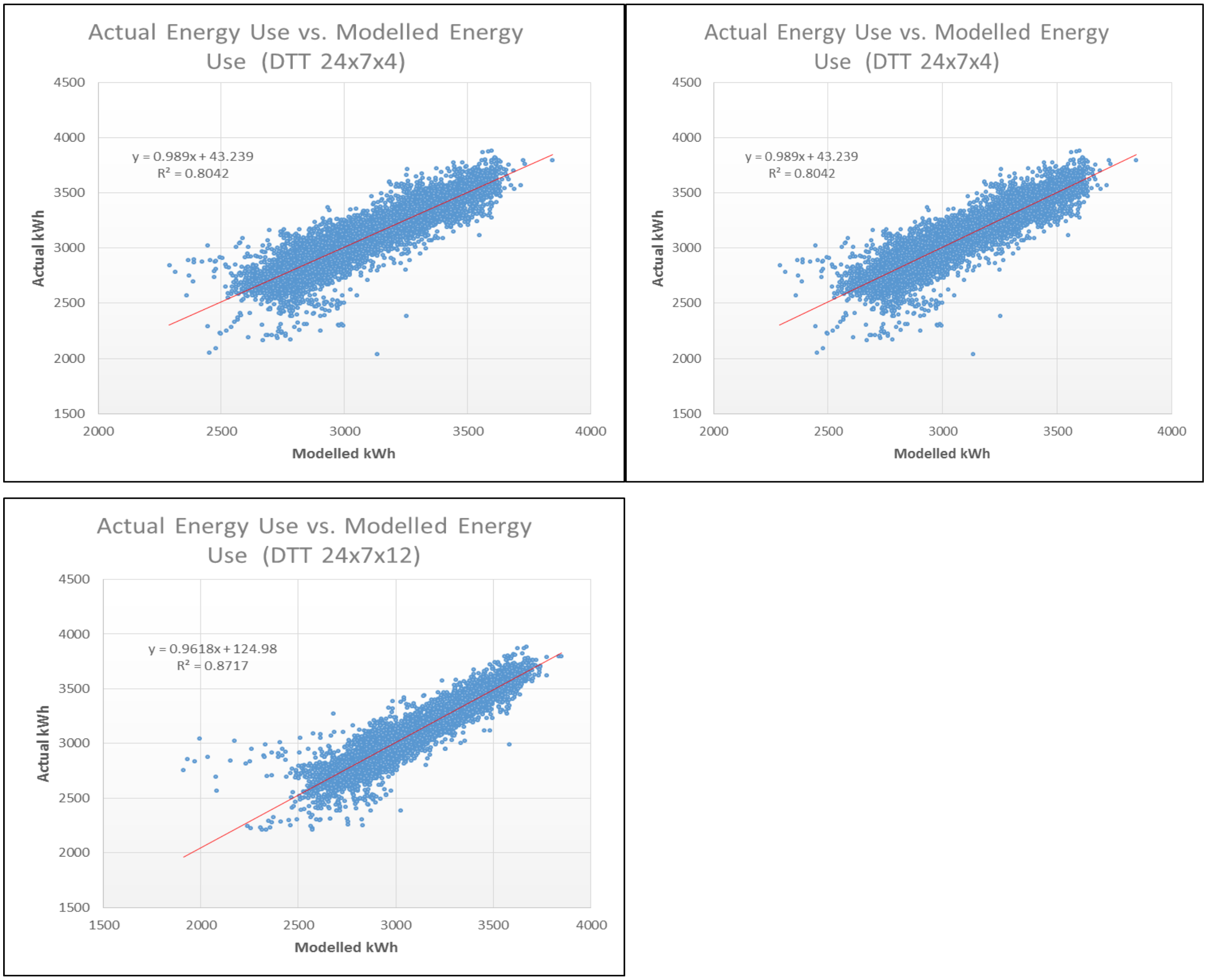

Figure 29 *R^2 of the various models applied to the data set of building 5 for year 1*

A.2 Year 2

A.2.1 Building 1

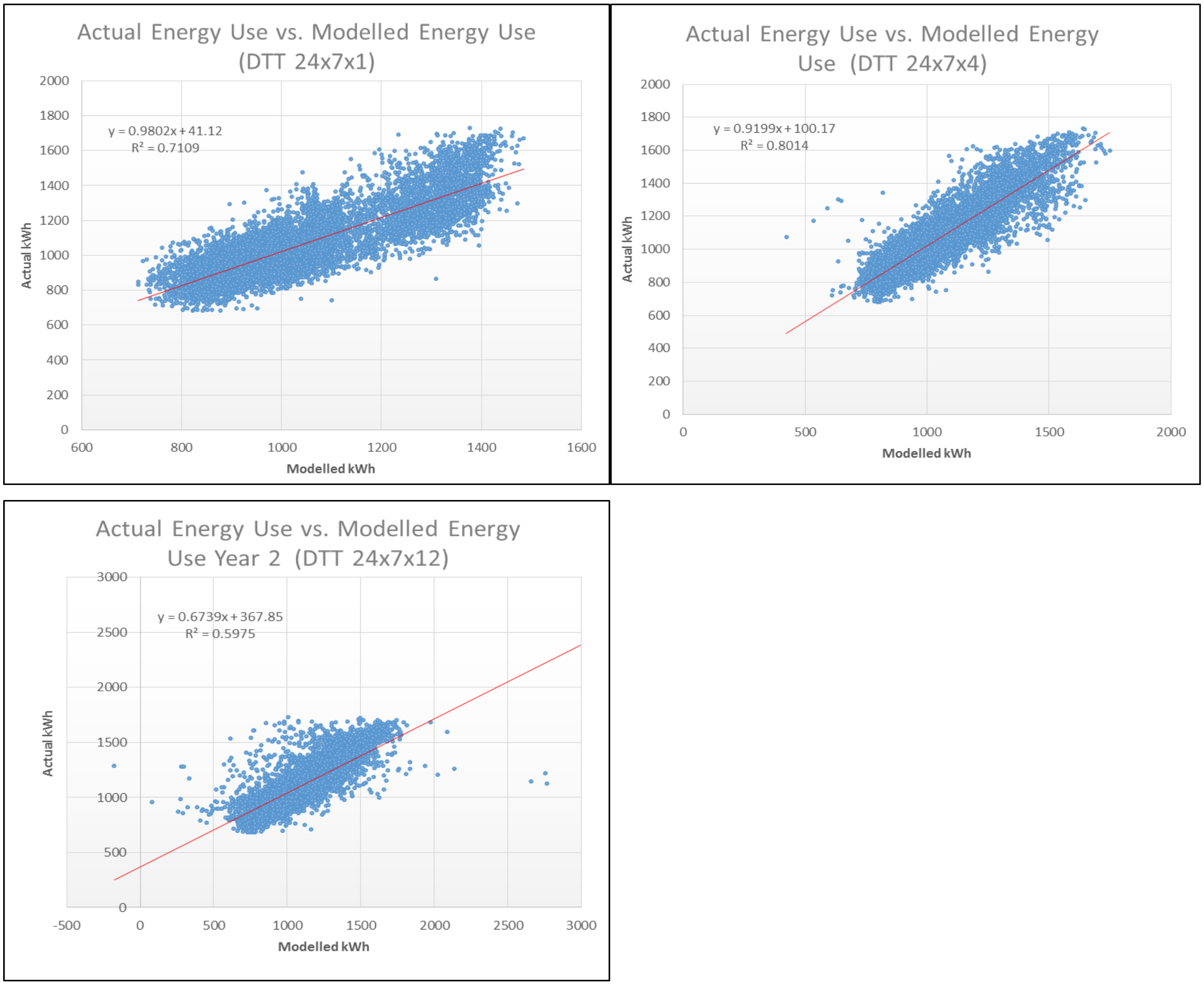

Figure 30 *R^2 of the various models applied to the data set of building 1 for year 2*

A.2.2 Building 2

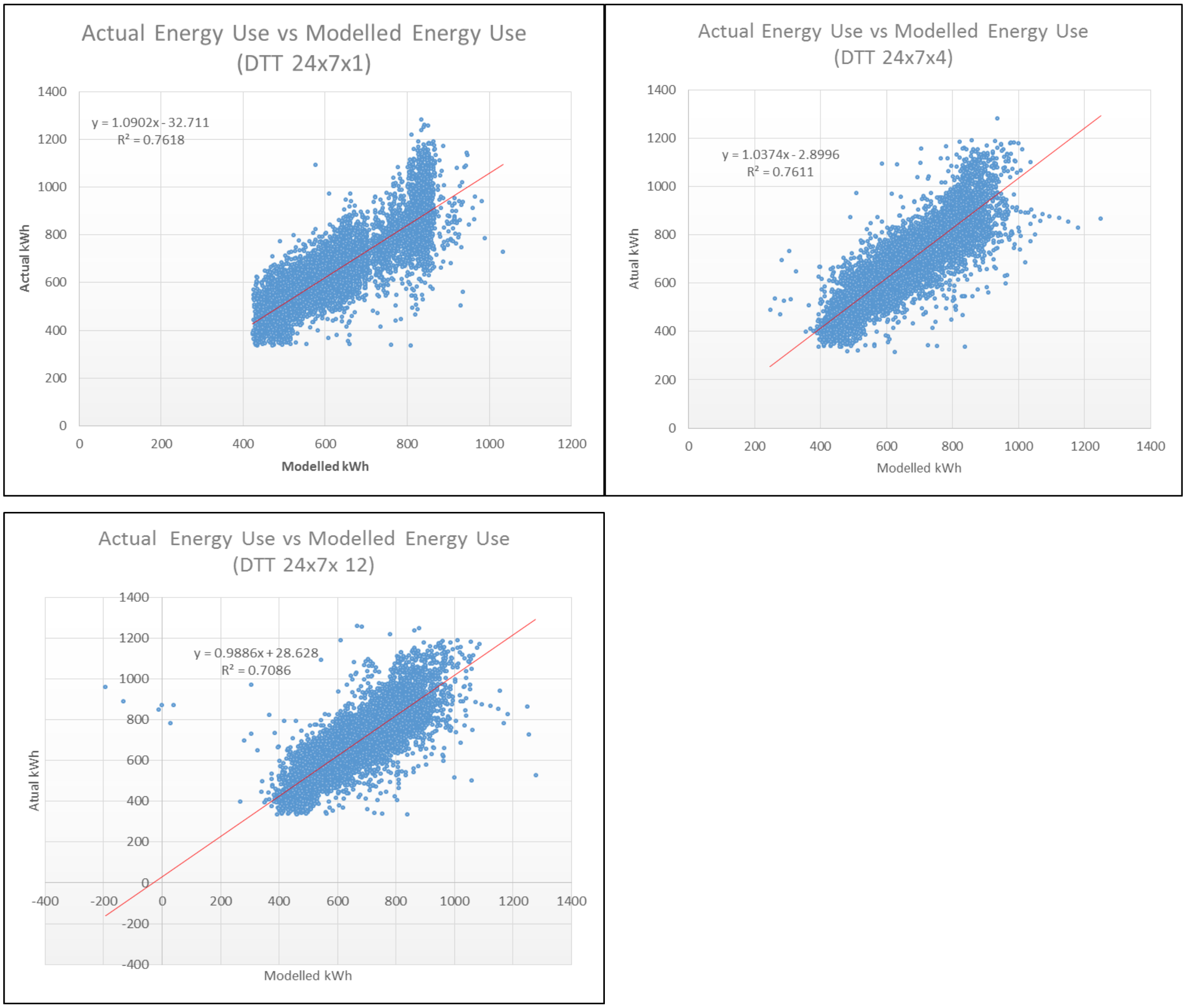

Figure 31 *R^2 of the various models applied to the data set of building 2 for year 2*

In the scatter plots it is seen that the R^2 value improves when introducing smaller time categories that include a certain range of temperatures. When the time category is too large, in the case of the DTT 24x7x1 model, it includes temperature which require both heating and cooling and cannot be accurately represented by piecewise linear regressions.

When the time categories are too small however, each regression model consist of only a few data points which can often be a very small sample that contains no correlation between energy and ambient temperatures. This leads to a very poor fit, especially when a new set of input data is used to model what the energy use would have been. This is very evident when looking at the outliers of the scatter plots of year 2.

Appendix B

This appendix contains samples of each model's fit to the actual data. Here it can once again be seen that the typical regression fits very poorly to hourly data. As we introduce more regressions for selected time categories in which certain linear correlations are found (different slopes for heating and cooling energy at hourly intervals of each day type), we can clearly see major improvement.

The fit improves as the time categories get smaller, but due to the poor correlation that may exist in small categories consisting of few data points, the fit is often poor when applying a new set of data to the model.

The first section contains the fit of each model to each building's actual data. In all cases for year 1 the fit improves as the time categories get smaller. Thus the DTT 24x7x12 model provides the best fit.

The second section contains the fit of each model to each building's data of a second year. In year 2 it is seen that the DTT 24x7x12 model often produces outliers and is therefore regarded as unstable.

The DTT 24x7x4 model remains stable and adjusts well to temperature changes and is therefore considered the favourable model.

B.1 Year 1

B.1.1 Building 1

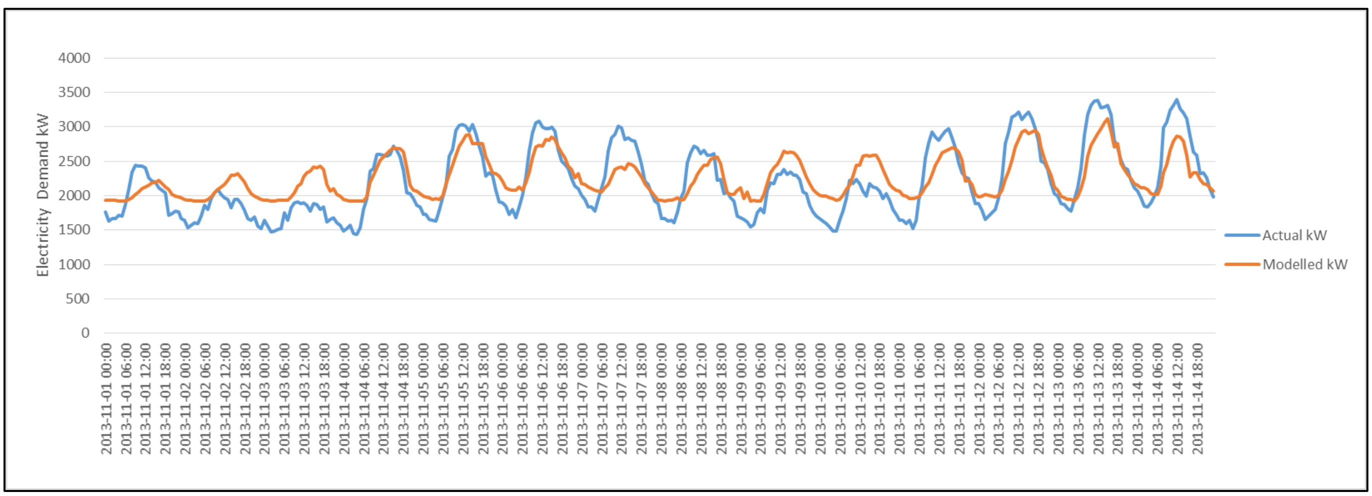

Figure 32 *Goodness-of-fit of a typical regression (one regression) model for building 1 during year 1*

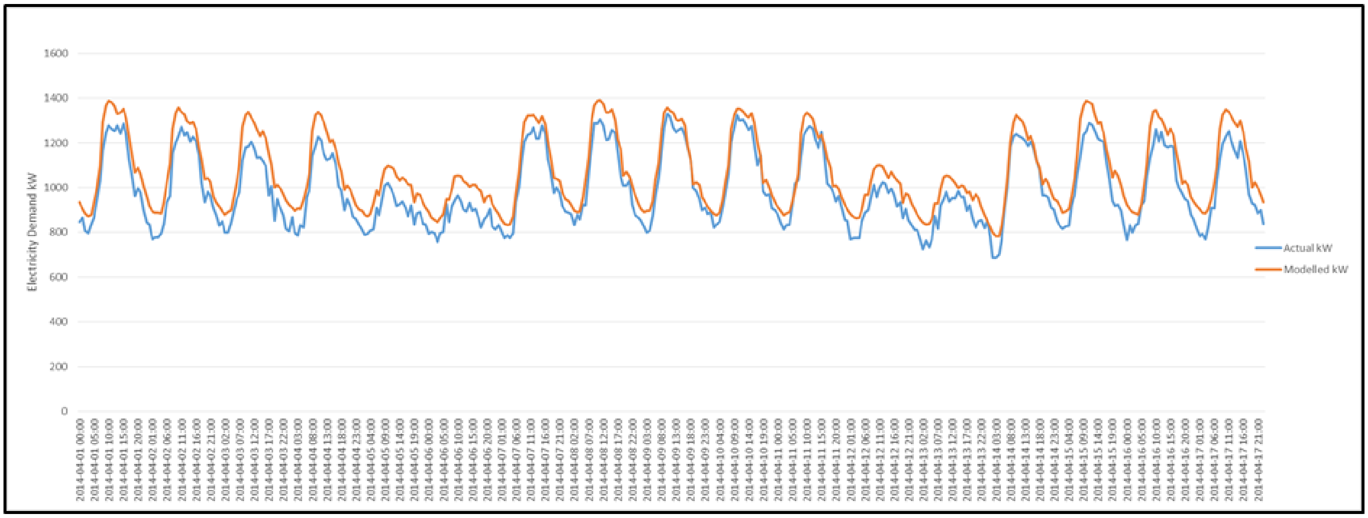

Figure 33 *Goodness-of-fit of the DTT 24x7x1 model for building 1 during year 1*

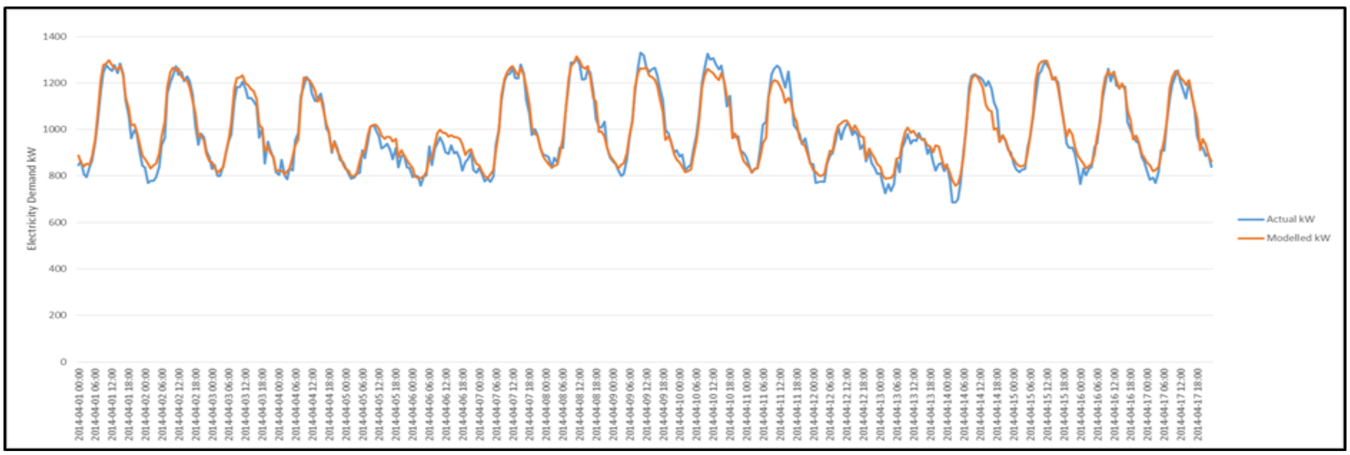

Figure 34 *Goodness-of-fit of the DTT 24x7x4 model for building 1 during year 1*

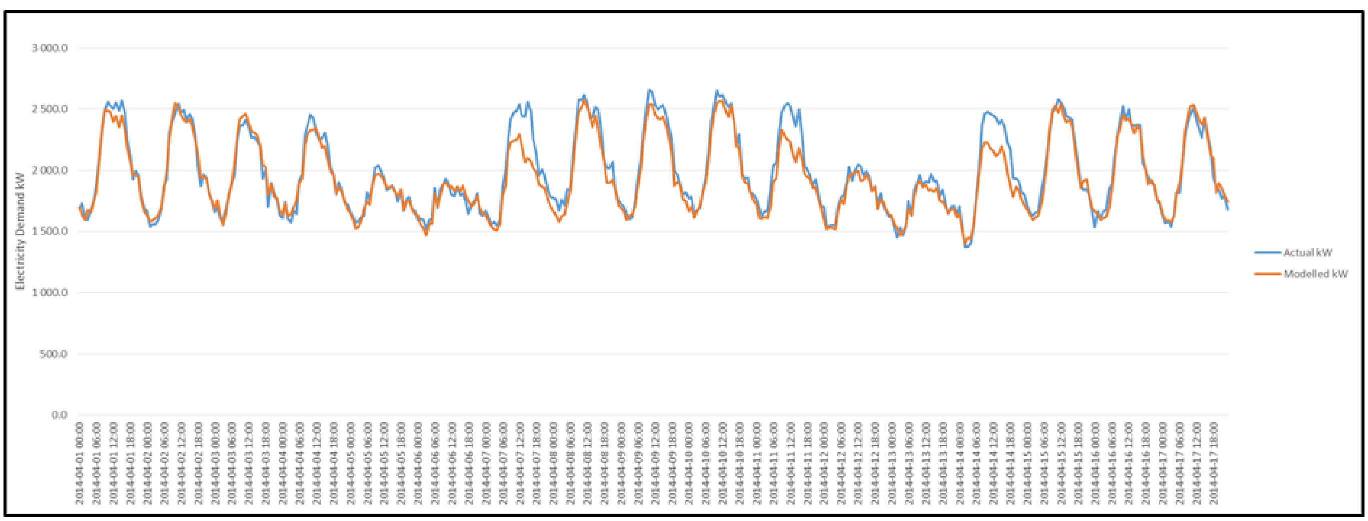

Figure 35 *Goodness-of-fit of the DTT 24x7x12 model for building 1 during year 1*

B.1.2 Building 2

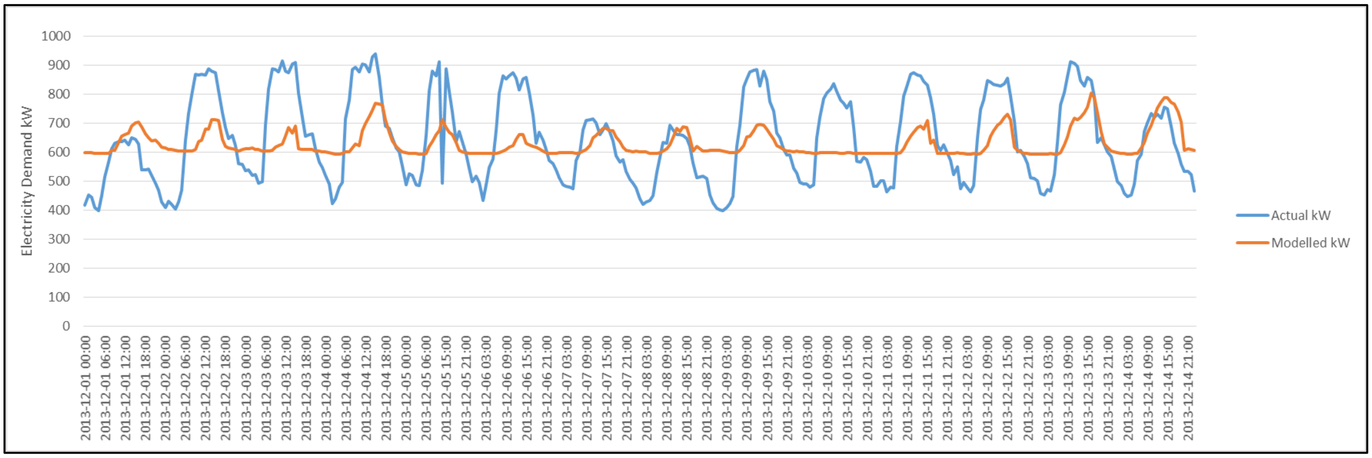

Figure 36 *Goodness-of-fit of a typical regression (one regression) model for building 2 during year 1*

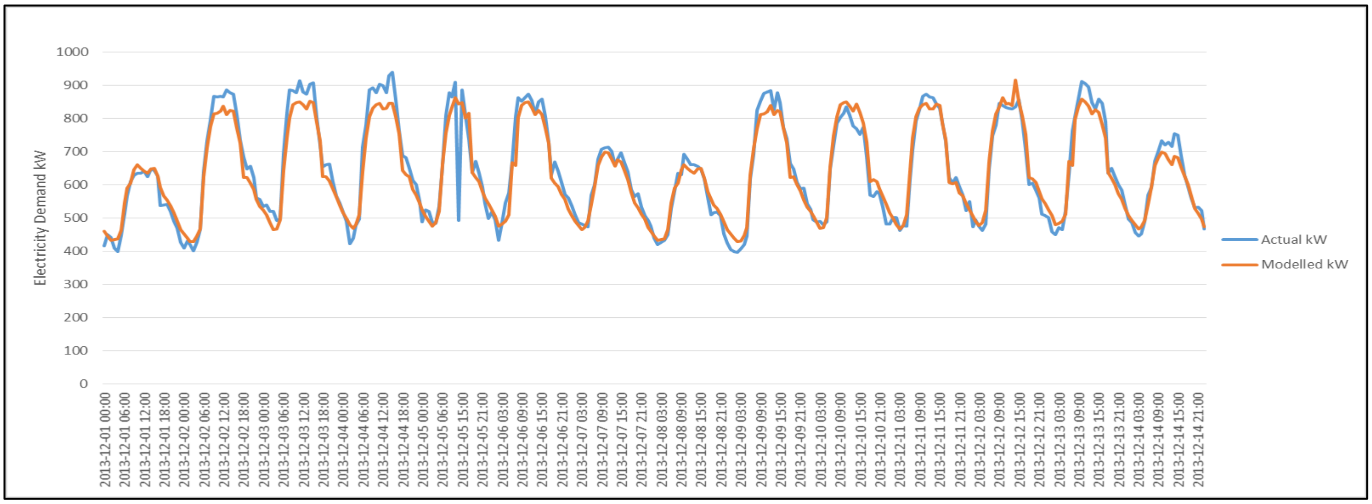

Figure 37 *Goodness-of-fit of the DTT 24x7x1 model for building 2 during year 1*

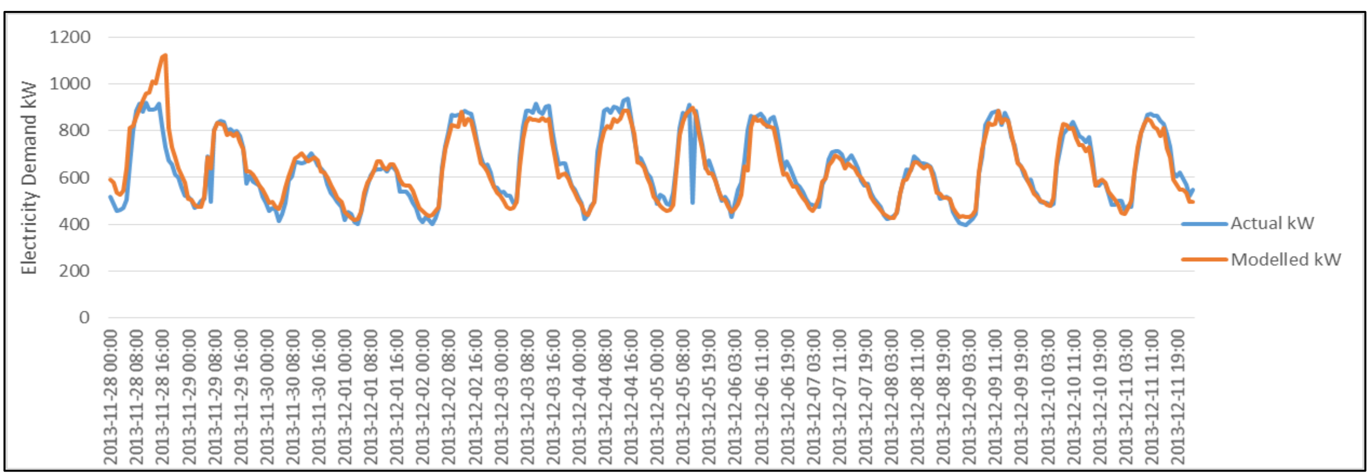

Figure 38 *Goodness-of-fit of the DTT 24x7x4 model for building 2 during year 1*

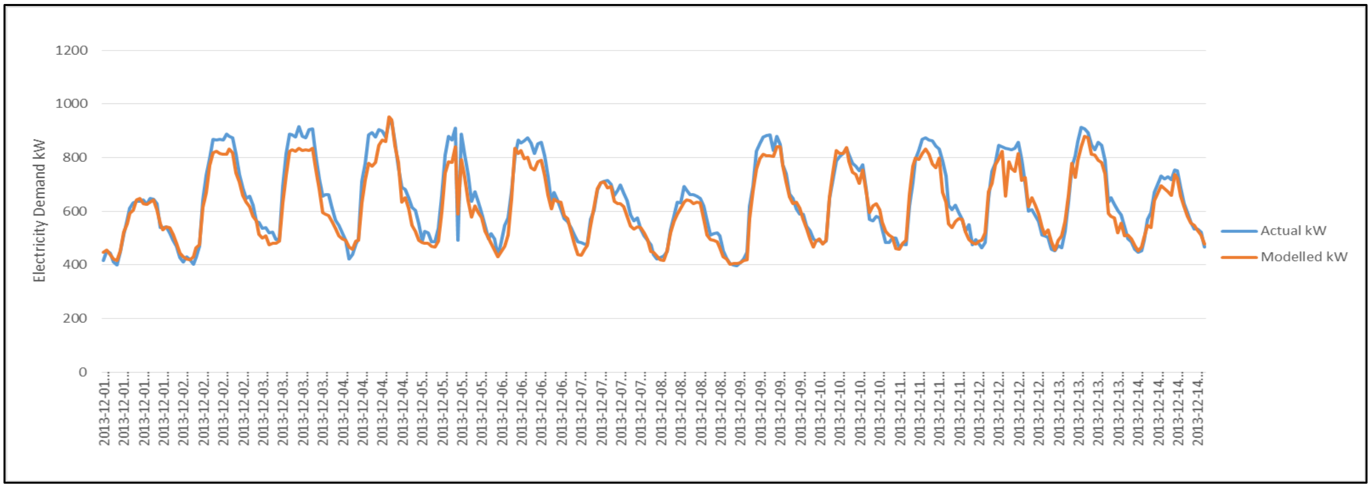

Figure 39 *Goodness-of-fit of the DTT 24x7x12 model for building 2 during year 1*

B.1.3 Building 3

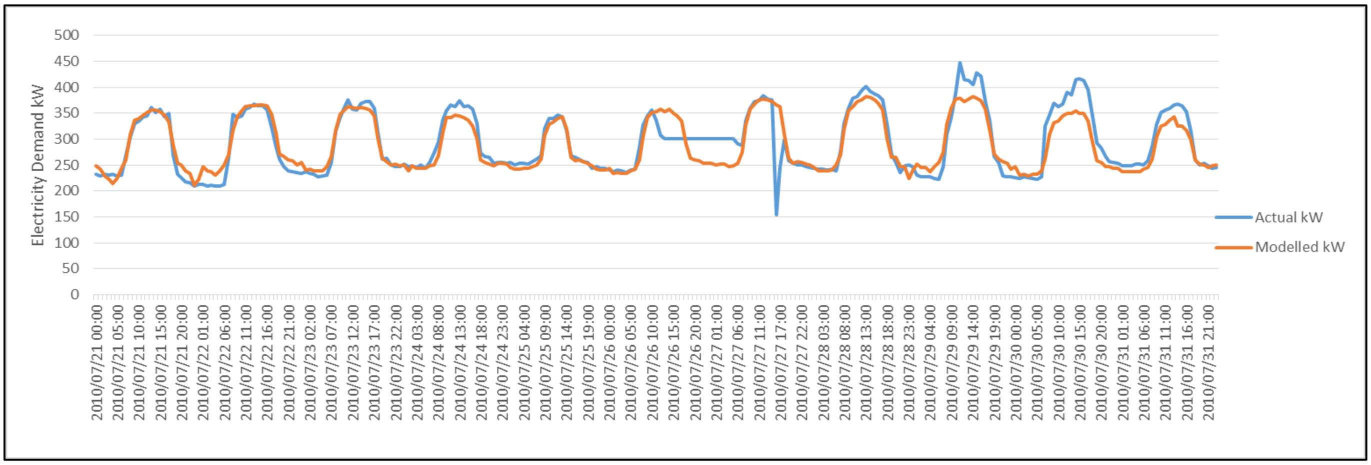

Figure 40 *Goodness-of-fit of the DTT 24x7x1model for building 3 during year 1*

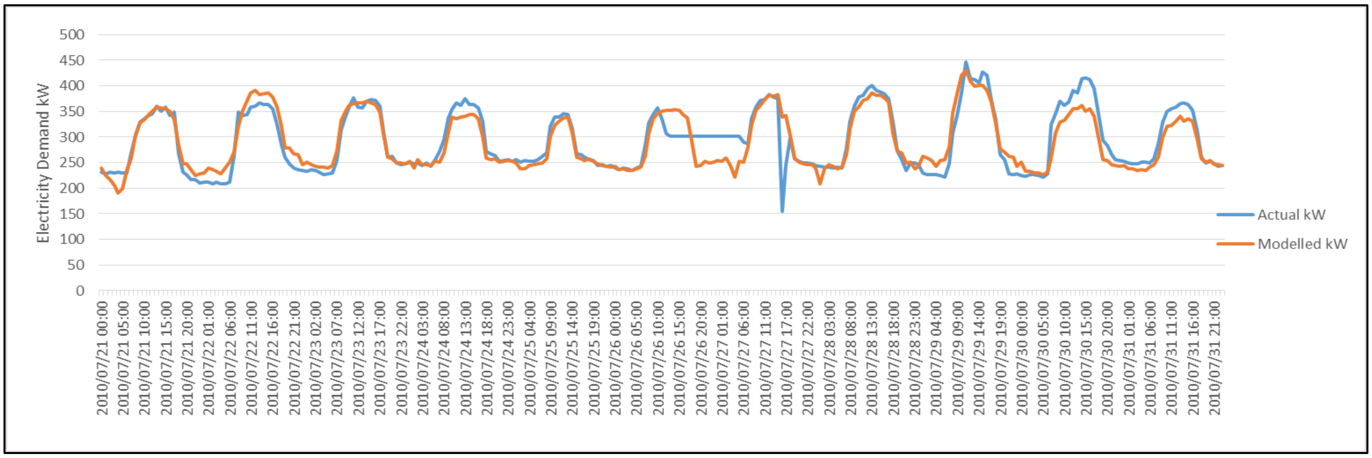

Figure 41 *Goodness-of-fit of the DTT 24x7x4 model for building 3 during year 1*

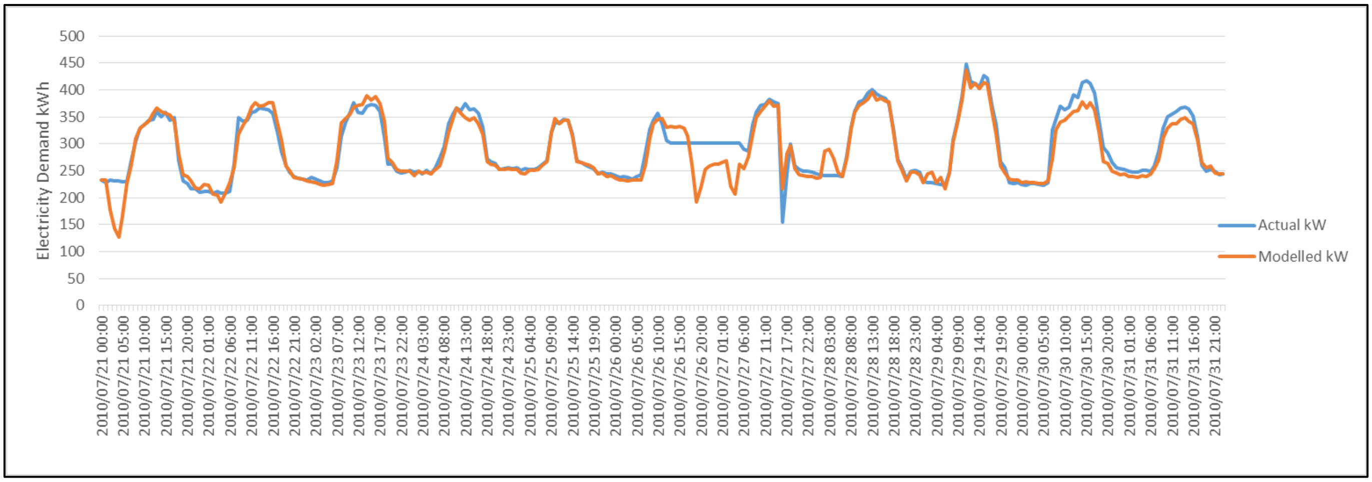

Figure 42 Goodness-of-fit of the DTT 24x7x12 model for building 3 during year 1

B.1.4 Building 4

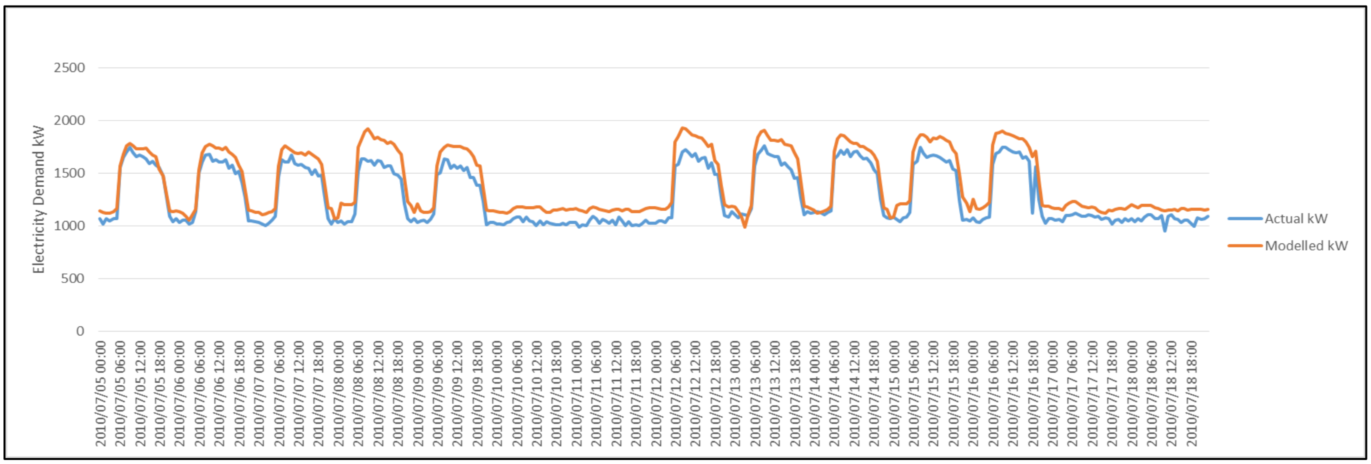

Figure 43 Goodness-of-fit of the DTT 24x7x1 model for building 4 during year 1

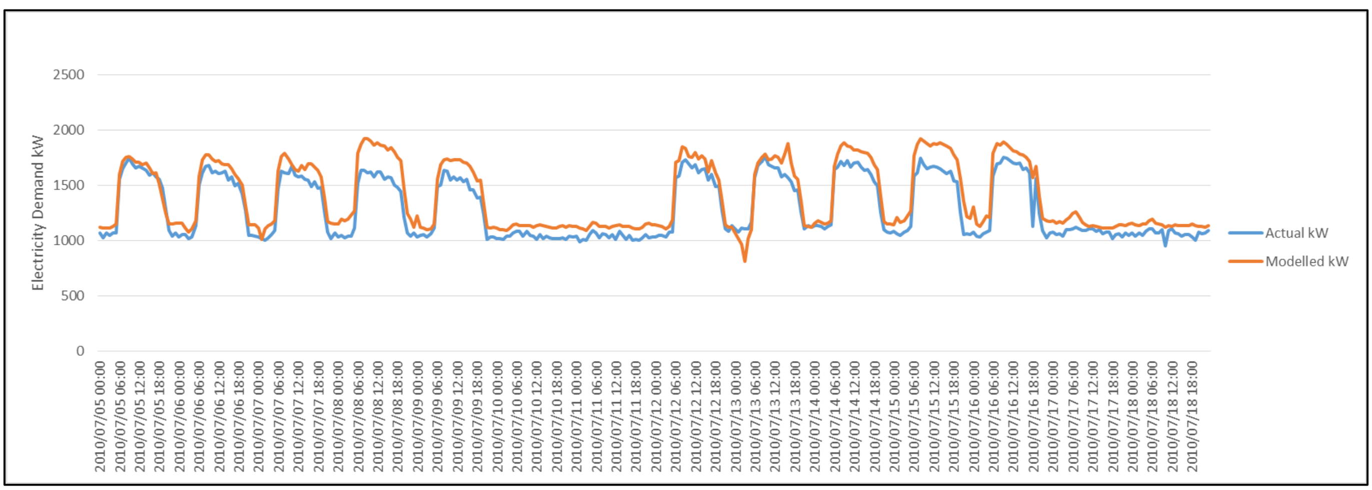

Figure 44 Goodness-of-fit of the DTT 24x7x4 model for building 4 during year 1

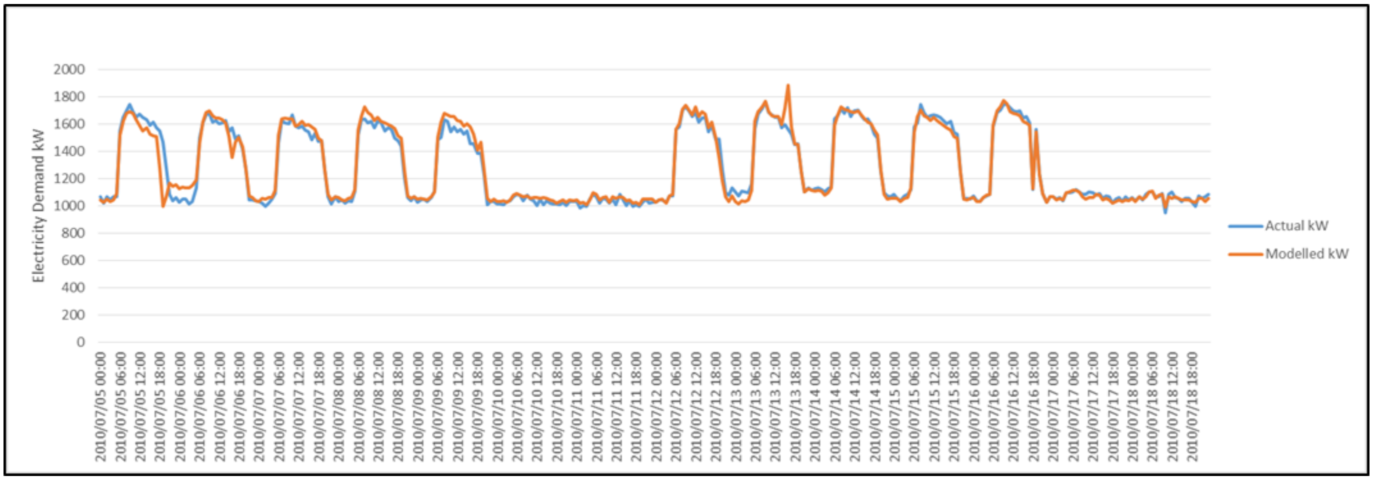

Figure 45 *Goodness-of-fit of the DTT 24x7x12 model for building 4 during year 1*

B.1.5 Building 5

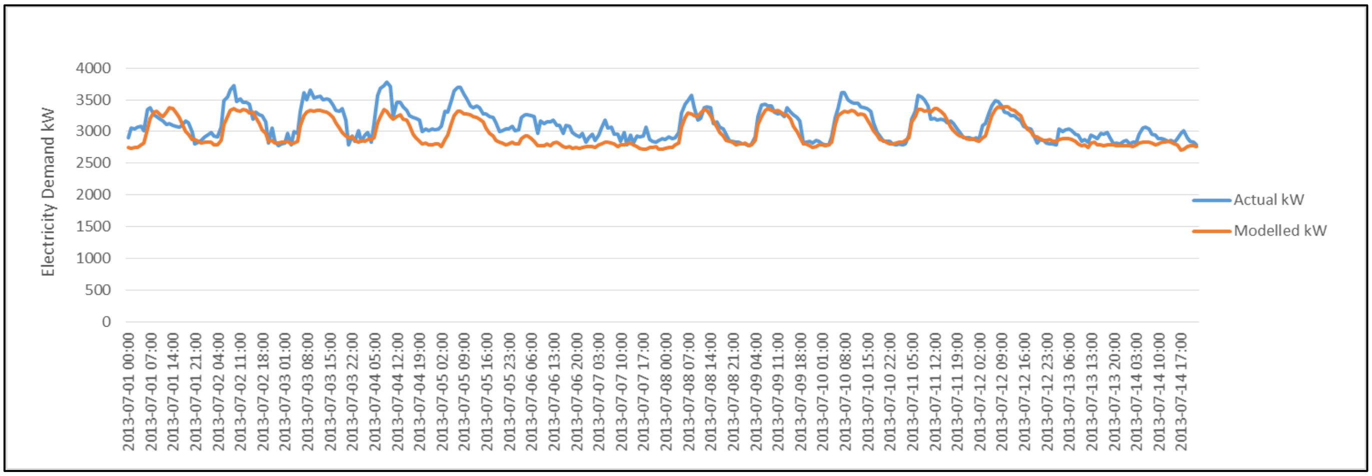

Figure 46 *Goodness-of-fit of the DTT 24x7x1 model for building 5 during year 1*

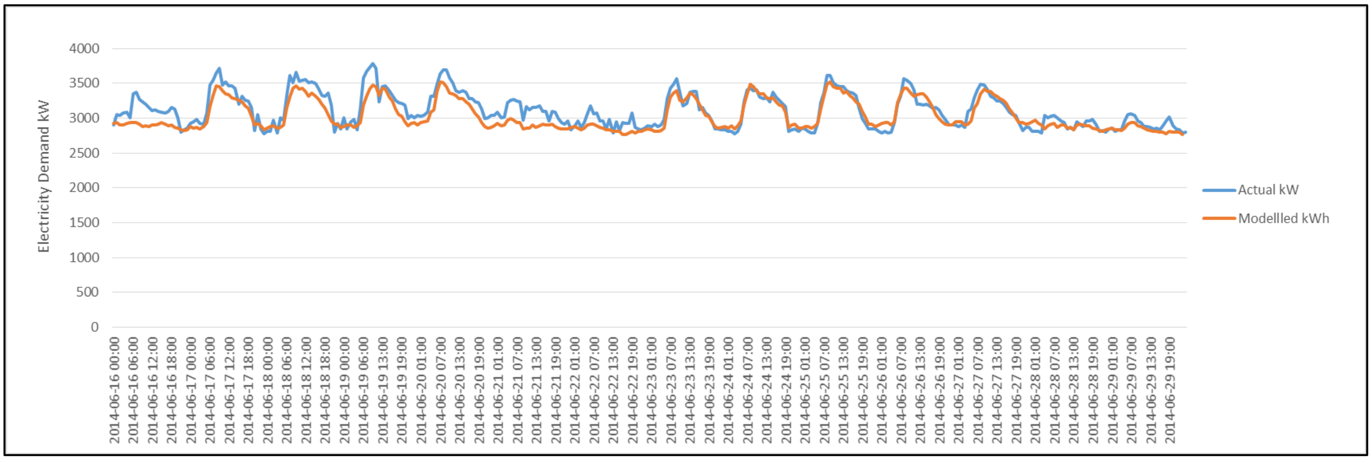

Figure 47 *Goodness-of-fit of the DTT 24x7x4 model for building 5 during year 1*

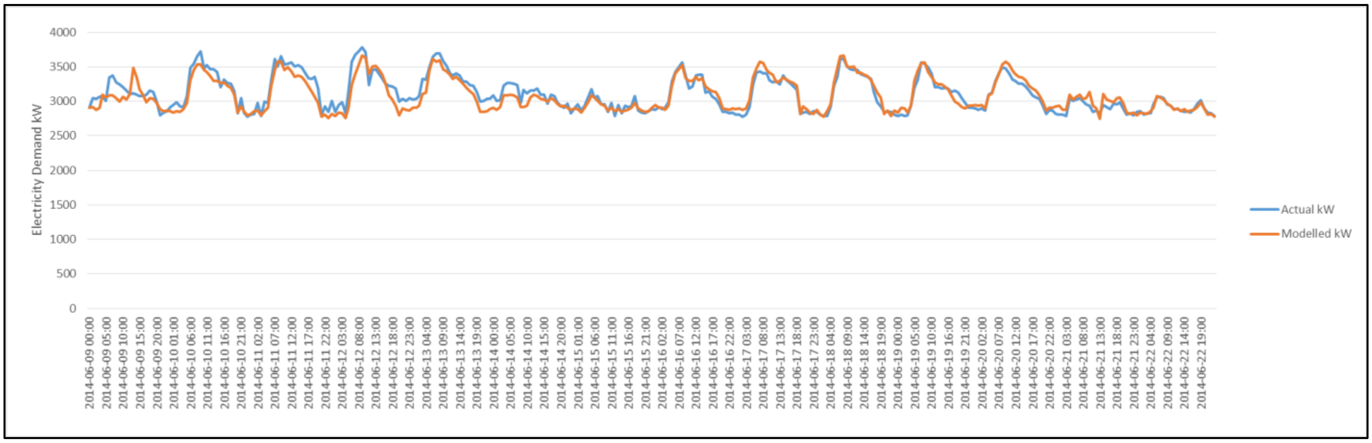

Figure 48 *Goodness-of-fit of the DTT 24x7x12 model for building 5 during year 1*

B.2 Year 2

B.2.1 Building 1

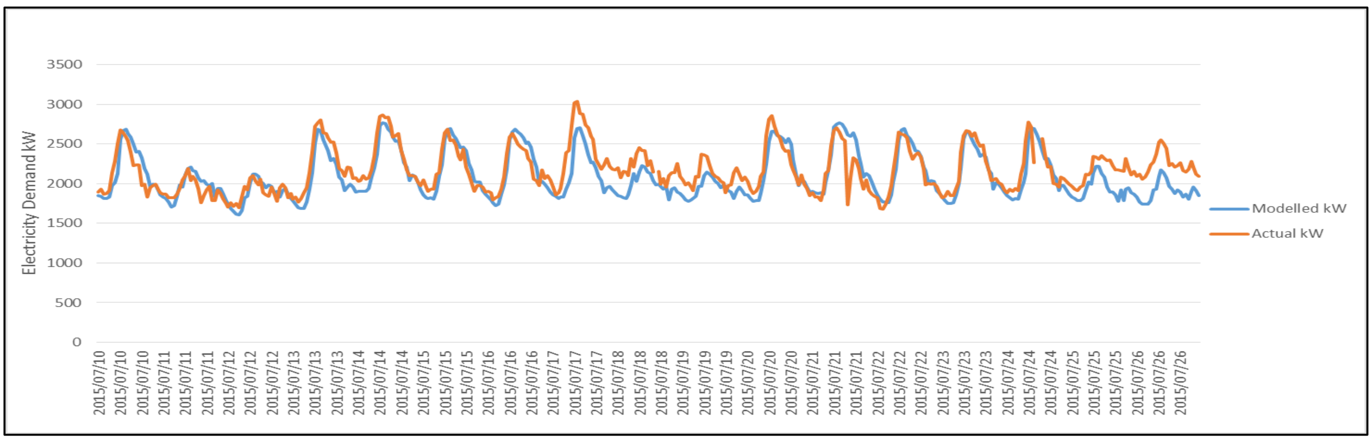

Figure 49 *Goodness-of-fit of the DTT 24x7x1 model for building 1 during year 2*

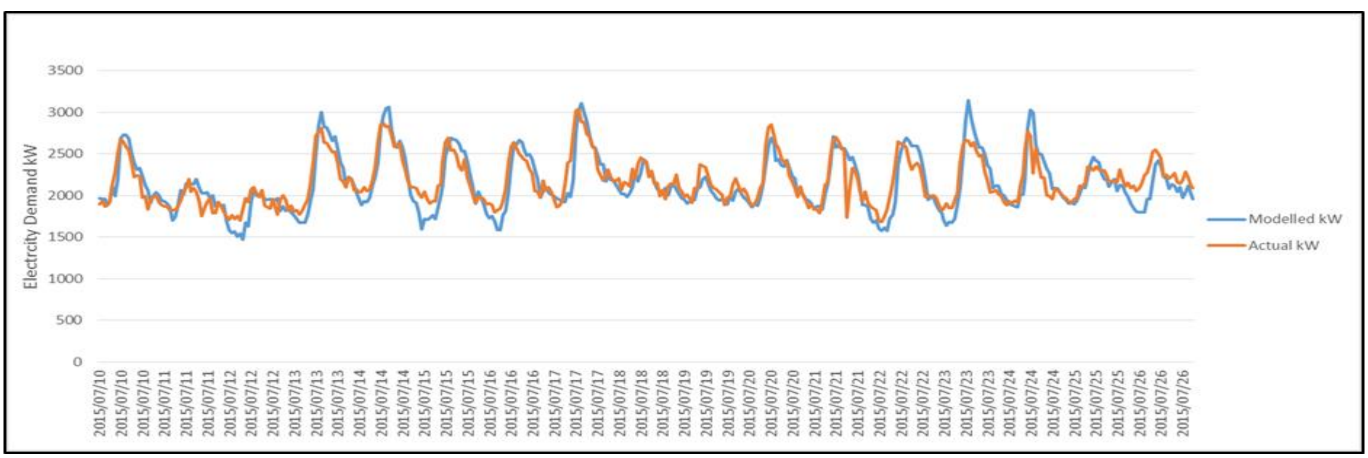

Figure 50 *Goodness-of-fit of the DTT 24x7x4 model for building 1 during year 2*

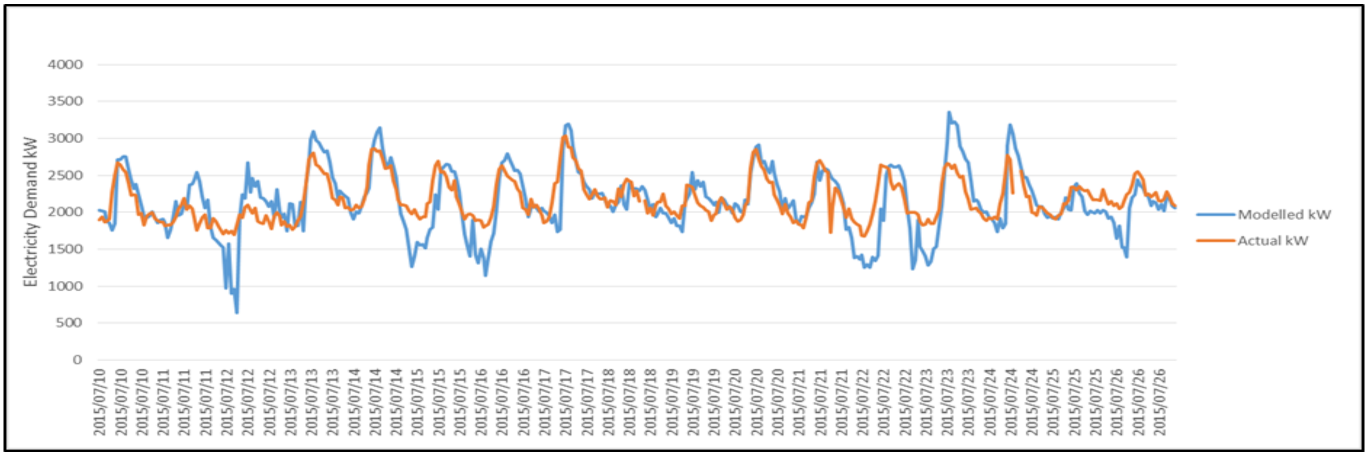

Figure 51 *Goodness-of-fit of the DTT 24x7x12 model for building 1 during year 2*

B.2.1 Building 2

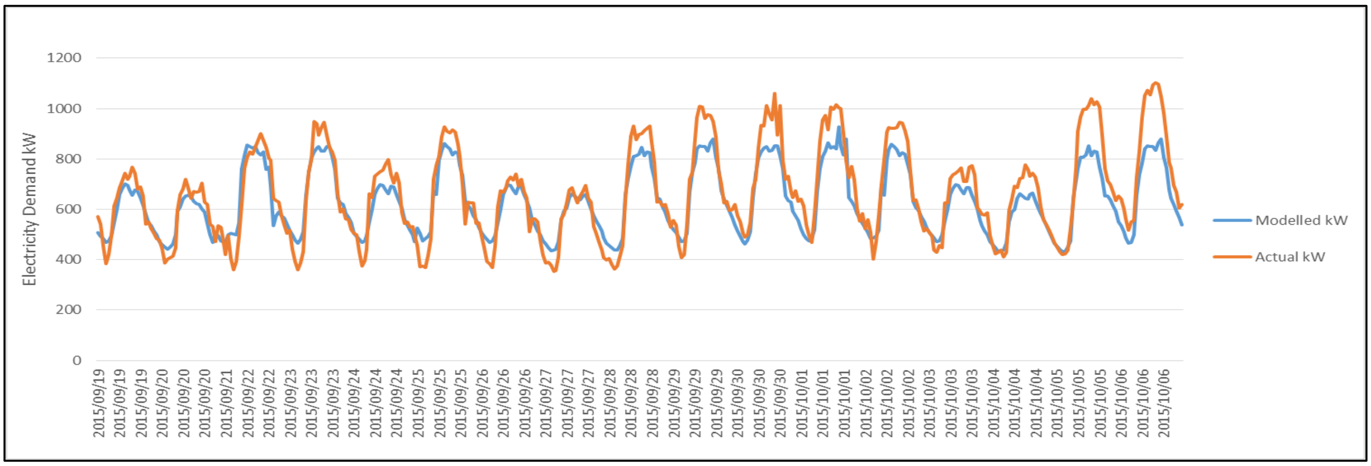

Figure 52 *Goodness-of-fit of the DTT 24x7x1 model for building 2 during year 2*

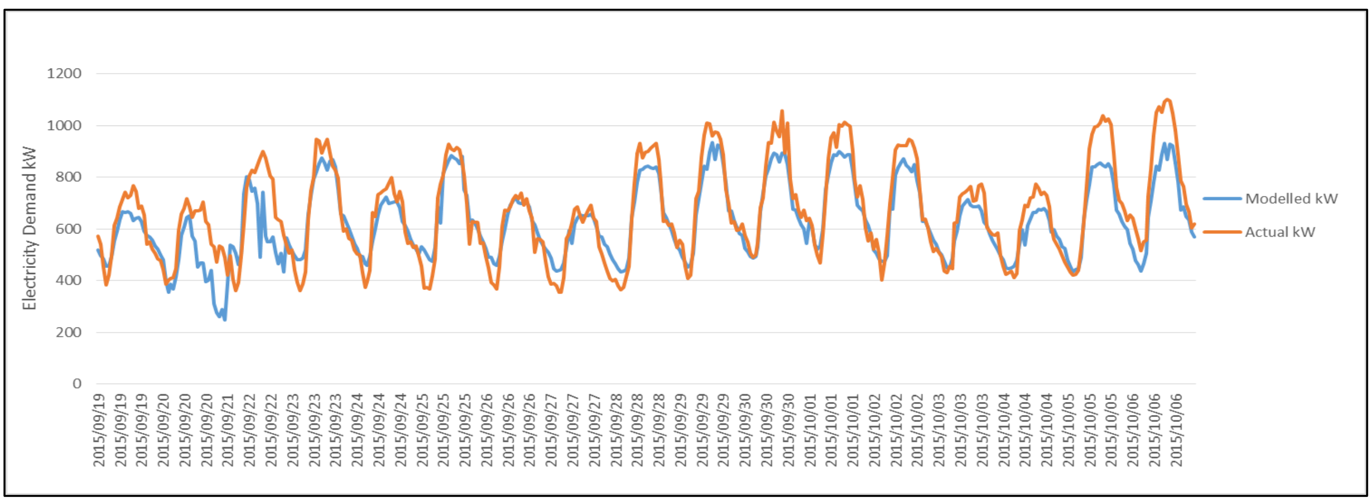

Figure 53 *Goodness-of-fit of the DTT 24x7x4 model for building 2 during year 2*

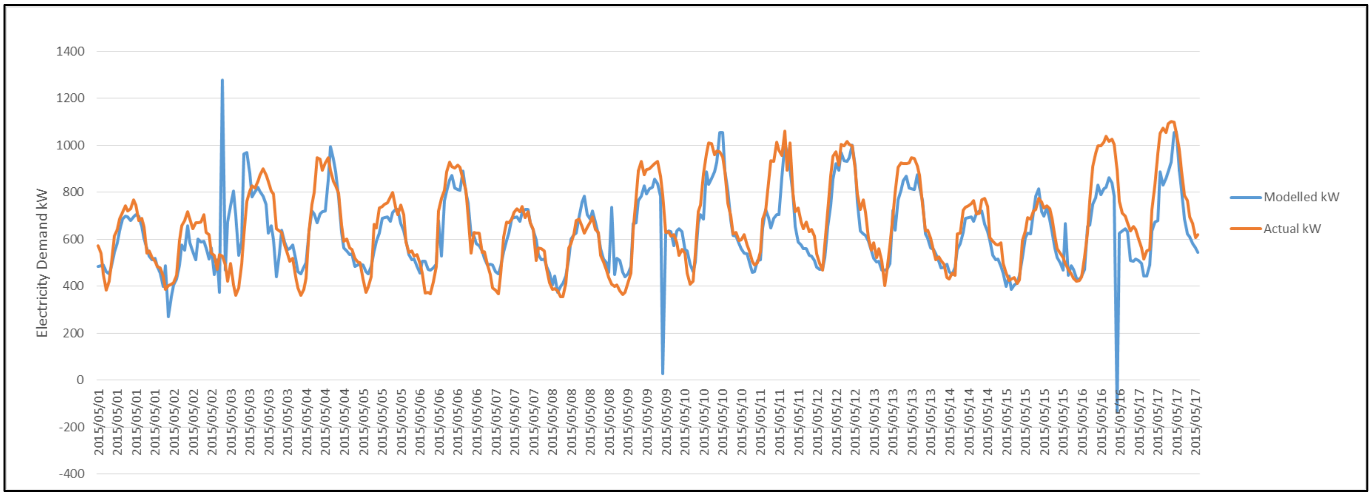

Figure 54 *Goodness-of-fit of the DTT 24x7x12 model for building 2 during year 2*