

An Exploration of Building Energy Model (BEM) Calibration in New Zealand

By Ethan Duff

A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements
for the degree of Master of Building Science

Victoria University of Wellington

2016

"All models are wrong, but some are useful"

George E.P. Box

Preface

This thesis was submitted in fulfilment of the requirements of the degree for Master of Building Science to the School of Architecture, Victoria University of Wellington.

Author

Ethan Duff

School of Architecture

Victoria University of Wellington

Email: ethanduff1@gmail.com

Supervisor

Dr. Nigel Isaacs

Senior Lecturer

School of Architecture

Victoria University of Wellington

Email: nigel.isaacs@vuw.ac.nz

Confidentiality Statement

Due to ethical considerations the author and participants of the study have agreed that all identifiable information is kept confidential.

To maintain confidentiality, pseudonyms are used throughout this paper to replace all identifiable information, such as 'participant', 'company', or 'project'.

The research reported here was carried out under VUW Ethics Approval: 22390 "Usability & Utility: An exploration of simple building energy model calibration" dated 5 October 2015

See section 4.5.6 for additional information regarding confidentiality and ethical considerations.

Acknowledgements

I would like to thank my supervisor Nigel Isaacs for his persistence, patience and support throughout this thesis. The always positive attitude and the 'chin-up' approach to supervision has helped me immensely through this year.

I must also pass on my thanks on to Mike Donn for offering his time and ear in the formative stages of this thesis.

Thanks to BRANZ and Victoria University of Wellington for supporting this study.

To my friends, Jamie, Elly, Yasmin, Michael, Brian and Shaan who were always there to laugh, discuss, think and debate with: I am sincerely thankful for the friendliness and camaraderie in the Building Science school, without which this would not have been nearly as fun.

Thank you to the participants involved in this research. Your help and assistance have helped me grow. Your expertise and willingness to help was invaluable.

To flatmates past and present who have had to put up with my long nights and constant distraction – thank you. To my family who have pretended to understand and listen through my ramblings and nattering, thank-you.

And most importantly to Sylvia. You are my world. I have never felt love, charity, or encouragement so much as when I am with you. You have carried me through thick and thin, an angel at my side.

Abstract

This study explored the use of Building Energy Modelling (BEM) and BEM calibration techniques for existing buildings as currently employed in New Zealand Industry.

Research on the use of BEM for existing building energy efficiency retrofits has increased dramatically over the past few decades. However, this use of BEM has been criticised for inaccurate and unbelievable results. These are often the result of not closely matching the building being modelled due to uncertainties around model inputs and modeller assumptions. As a result, researchers have responded by developing techniques to 'calibrate' models by comparing the simulated building with the actual building energy use thus providing quality assurance.

However, many of these techniques are difficult, esoteric, convoluted or impractical for industry professionals. This research explored if a simple calibration technique developed at Victoria University of Wellington by Dr. Shaan Cory would meet the needs of industry practitioners. The technique was turned into a usable tool and student trialled to prepare it for industry assessment. Four BEM experts were then interviewed in a series of individual interviews and workshops trialling the use of the technique.

The research concluded that the use of BEM is limited in New Zealand due to a perceived Industry value gap – building owners are not aware of the benefits of modelling whole-building retrofits. This leads to reduced uptake of calibration techniques from industry resulting in a credibility gap, where the modeller themselves may not be confident of their own BEMs. This is due, in part, to a lack of industry quality assurance guidelines, usable calibration tools, and certainty around model inputs. The adoption of the streamlined Cory method would be of significant benefit to practitioners. However, it was identified that it did not solve all issues relating to uncertainty estimation.

List of Acronyms and Abbreviations

BEM	Building Energy Model
BMS	Building Management System
BRANZ	Building research Association
COP	Coefficient of Performance
CV(RMSE)	Coefficient of Variation (Root Mean Squared Error)
ECM	Energy Conservation Measure
EECA	Energy Efficiency Conservation Authority
EMANZ	Energy Management Association of New Zealand
GHG	Green-House Gas
HVAC	Heating, Ventilating and Air Conditioning
IPMVP	International Protocol for Measurement and Verification Performance
M&V	Measurement and Verification
MBE	Mean Bias Error
NABERS	National Australian Built Environment Rating System
NABERSNZ	National Australian Built Environment Rating System New Zealand
NZGBC	New Zealand Green Building Council
O&M	Operation and Manual
PID	Proportional–Integral–Derivative controller
SUS	System Usability Scale
UIM	Utility Inspection Method

Contents

PREFACE	II
CONFIDENTIALITY STATEMENT	III
ACKNOWLEDGEMENTS	IV
ABSTRACT	V
LIST OF ACRONYMS AND ABBREVIATIONS	VI
LIST OF FIGURES	XI
LIST OF TABLES	XII
EQUATIONS	XII
<u>1 INTRODUCTION</u>	<u>1</u>
1.1 THE STATE OF INDUSTRY	2
1.2 DIFFICULTIES IN BEM VALIDATION	3
1.2.1 INDUSTRY/RESEARCH DISCONNECT	5
1.3 RESEARCH FOCUS	6
1.3.1 PROBLEM STATEMENT	6
1.3.2 RESEARCH QUESTIONS	7
1.3.3 RESEARCH METHOD	7
<u>2 LITERATURE REVIEW</u>	<u>9</u>
2.1 BUILDINGS IN NEW ZEALAND	9
2.2 NABERS ENERGY EFFICIENCY ASSESSMENT METHODS	10
2.2.1 NEW ZEALAND METHODS OF ENERGY EFFICIENCY ASSESSMENT	12
2.3 BUILDING ENERGY MODELLING (BEM)	14
2.3.1 BEM TOOLS	15
2.3.2 ADVANTAGES OF BEM	15
2.3.3 DISADVANTAGES OF BEM	15

2.4 BEM CALIBRATION	22
2.4.1 STATISTICAL INDICES FOR CALIBRATION	23
2.4.2 CALIBRATION METHODS	26
2.5 CORY CALIBRATION METHOD	32
2.5.1 EVIDENCE BASED APPROACH	32
2.5.2 ENERGY SIGNATURE METHOD	33
2.5.3 CALIBRATION PROCEDURE	36
2.5.4 ADVANTAGES OF THE CORY MODEL	38
2.5.5 DISADVANTAGES OF THE CORY MODEL	38
2.5.6 USING THE CURRENT CORY PROCESS	39
2.6 REDDY METHOD	39
2.6.1 REDDY CALIBRATION PROCEDURE	40
2.6.2 ADVANTAGES OF THE REDDY METHOD	42
2.6.3 DISADVANTAGES OF THE REDDY METHOD	42
2.6.4 APIDAE LABS CALIBRATOR	43
2.6.5 DIFFERENCES IN APPROACH	44
2.7 REVIEW OF BEM INDUSTRY GUIDELINES	44
2.8 SUMMARY	47
 3 RESEARCH METHOD PART 1 STUDENT WORKSHOPS	 49
 3.1 MEASURING USABILITY	 49
3.2 CLASS OVERVIEW	50
3.2.1 AN ALTERNATIVE PROPOSAL	52
3.2.2 RESEARCH METHOD	52
3.3 RESULTS	56
3.4 IMPROVEMENTS	58
3.4.1 STUDENT WORKSHOPS	58
3.4.2 PRINCIPLES OF INTERFACE DESIGN	61
3.5 COMPLETED TOOL	64
 4 RESEARCH METHOD PART 2 INDUSTRY WORKSHOPS	 69
 4.1.1 THE USE OF QUALITATIVE RESEARCH IN BUILDING SCIENCE	 69
4.2 UNDERSTANDING THE RESEARCH QUESTIONS	70

4.3	INDUSTRY WORKSHOPS	71
4.3.1	DATA COLLECTION METHOD	71
4.4	SAMPLE SIZE REQUIREMENTS	73
4.4.1	PROBABILITY AND NON-PROBABILITY SAMPLING.	73
4.4.2	SAMPLE DESIGN – DEFINING THE EXPERT INTERVIEW	76
4.4.3	SAMPLE DESIGN – IDENTIFYING THE POPULATION	77
4.4.4	SAMPLE SELECTION AND LIMITATIONS	78
4.5	THE WORKSHOP AND INTERVIEW PROCESS	78
4.5.2	INTERVIEW DESIGN	81
4.5.3	INTERVIEW INTERACTION	83
4.5.4	WORKSHOP INTERACTION	84
4.5.5	DATA RECORDING AND TRANSCRIPTION	86
4.5.6	ETHICAL CONSIDERATIONS	87
4.5.7	RESEARCH LIMITATIONS	87
4.6	INTERVIEW DATA ANALYSIS	87
4.6.1	THEMATIC ANALYSIS	88
4.7	SUMMARY	90
5	FINDINGS	91
5.1	THE ANALYSIS PROCESS	91
5.2	PRE-WORKSHOP INTERVIEWS	96
5.2.1	FINDINGS USING THEMATIC ANALYSIS	96
5.3	INDUSTRY CONSTRAINTS	97
5.3.1	CLIENT EXPECTATIONS	97
5.3.2	AVAILABLE TOOL CHEST	100
5.3.3	LACK OF BUILDING EXPERIENCE AMONGST MODELLERS	102
5.4	MODELLING CONSTRAINTS	103
5.4.1	AVAILABLE TIME AND TOOL DESIGN	103
5.4.2	REDUCED UPSKILLING AND DISCONNECT FROM LITERATURE	104
5.4.3	ACCESS TO THE RIGHT BUILDING INFORMATION	105
5.4.4	ABILITY TO CAPTURE BUILDING CONTROL	106
5.5	SUMMARY – CREDIBILITY GAP	107
5.6	POST-WORKSHOP INTERVIEWS	108
5.6.1	OVERALL VIEW OF THE CORY METHOD	109

5.6.2	POSITIVE ASPECTS OF THE CORY METHOD	110
5.6.3	NEGATIVE ASPECTS OF THE CORY METHOD	111
5.6.4	CONTRASTING MANUAL AND AUTOMATED APPROACHES	113
5.6.5	SUMMARY – COMPARING THE TWO TOOLS	116
6	DISCUSSION	117
6.1	RESEARCH QUESTION ONE	117
6.1.1	PREFERRED USE OF STATIC EMPIRICAL MODELS TO DYNAMIC MODELS	118
6.1.2	DEALING WITH UNCERTAINTY	119
6.1.3	EXPERIENCE VS INEXPERIENCE	119
6.1.4	BALANCING RISK AND REWARD	120
6.2	RESEARCH QUESTION TWO	121
6.2.1	AUTOMATED VS MANUAL APPROACHES	121
6.3	RESEARCH QUESTION THREE	122
6.3.1	FUTURE CONSIDERATIONS FOR CALIBRATION TOOLS	122
6.3.2	REPLICATION OF BUILDING CONTROL	124
6.4	VALIDITY OF RESULTS AND DISCUSSION OF THE RESEARCH METHOD	126
7	CONCLUSIONS	129
7.1	A CREDIBILITY GAP EXISTS FROM THE PERSPECTIVE OF THE MODELLER	129
7.2	UNCERTAINTY IS A KEY FACTOR	130
7.3	THE NEED FOR A SEMI-AUTOMATED PROCESS	130
7.4	OVERCOMING BARRIERS TO BEM CALIBRATION UPTAKE	131
7.5	FUTURE RESEARCH	132
8	REFERENCES	135
9	APPENDICES	148
9.1	APPENDIX A - INTERVIEW GUIDE	148
9.2	APPENDIX B – ETHIC COMMITTEE DOCUMENTS	151
9.2.1	STUDENT PARTICIPANT INFORMATION AND CONSENT SHEETS	151
9.2.2	INDUSTRY PARTICIPANT INFORMATION AND CONSENT SHEETS	154

List of Figures

FIGURE 1 BENCHMARK FACTORS AND STAR SCALE FOR NABERS SOURCE: (OFFICE OF ENVIRONMENT AND HERITAGE, 2011)	11
FIGURE 2 IMPORTANCE OF PARAMETERS FOR PREDICTING TOTAL ELECTRICITY USE (BERKELEY, 2013)	20
FIGURE 3 IMPORTANCE OF PARAMETERS FOR PREDICTING TOTAL GAS USE (BERKELEY, 2013)	21
FIGURE 4 MEASURED VS ACTUAL ELECTRICITY CONSUMPTION EXAMPLE FROM (WALTZ, 2000)	27
FIGURE 5 EXAMPLE CALIBRATION SIGNATURE (CORY, 2015)	34
FIGURE 6 EXAMPLE CHARACTERISTIC SIGNATURES (CORY, 2015)	36
FIGURE 7 EXAMPLE IDENTIFICATION OF AN INCORRECT INPUT PARAMETER (CORY, 2015)	36
FIGURE 8 CALIBRATION PROCEDURE FROM (CORY, 2015)	38
FIGURE 9 APIDAE CALIBRATION TOOL RESULTS DISPLAY AND PLOT DESCRIPTION	44
FIGURE 10 STUDENT WORKSHOP TIMELINE	52
FIGURE 11 STUDENT WORKSHOP OBSERVATION DIARY EXCERPT	53
FIGURE 12 HISTOGRAM OF INDIVIDUAL SUS SCORES (N=2324) (BANGOR ET AL., 2008)	55
FIGURE 13 HISTOGRAM OF STUDY MEANS OF SUS SCORES (N=206) (BANGOR ET AL., 2008)	55
FIGURE 14 A COMPARISON OF MEAN SUS SCORES BY QUARTILE (BANGOR ET AL., 2008)	56
FIGURE 15 STUDENT WORKSHOPS: DISTRIBUTION OF TIME	58
FIGURE 16 BEES OPEN PLAN TEMPLATE MODEL	61
FIGURE 17 DASHBOARD INPUTS AND BUTTON	61
FIGURE 18 DASHBOARD SIGNPOSTING OF PROCESS STEPS AND ACTIONABLE BUTTONS	62
FIGURE 19 DASHBOARD NOTIFICATION DISPLAY ON TASK COMPLETION	62
FIGURE 20 STREAMLINED CORY PROCESS TOOL – DASHBOARD ONE	64
FIGURE 21 DASHBOARD ONE – INTERFACE FEATURES	65
FIGURE 22 STREAMLINED CORY PROCESS TOOL - DASHBOARD TWO	66
FIGURE 23 DASHBOARD TWO – INTERFACE FEATURES	67
FIGURE 24 EXEMPLAR BEM DEVELOPMENT	81
FIGURE 25 INDUSTRY WORKSHOP TIMELINE	81
FIGURE 26 THEMATIC ANALYSIS APPROACH	89
FIGURE 27 THEMATIC NETWORKS - INDUSTRY AND MODELLING CONSTRIANTS	95
FIGURE 28 SUS SCORES AND ACCEPTABILITY RANGES FOR THE STREAMLINED CORY TOOL	111

List of Tables

TABLE 1 FINANCIAL IMPACT OF GREEN PREMIUMS IN NABERS RATED BUILDINGS (ADJUSTED FOR FLOOR AREA) SOURCE : (NEWELL ET AL., 2011)	12
TABLE 2 CALIBRATION LIMITS (ASHRAE, 2002)	26
TABLE 3 SUMMARY OF CALIBRATION METHODS BY (COAKLEY ET AL., 2014)	30
TABLE 4 ISSUES RELATING TO CALIBRATION TECHNIQUES. ADAPTED FROM (COAKLEY ET AL., 2014)	31
TABLE 5 INTERVIEW GUIDE CATEGORIES	82
TABLE 6 THEMATIC ANALYSIS PROCEDURE (BRAUN & CLARKE, 2006)	90
TABLE 7 EXAMPLE OF INITIAL CODING	92
TABLE 8 EXAMPLE CODE TABLE FROM THE PRE-WORKSHOP INTERVIEW	92
TABLE 9 EXAMPLE THEMATIC ANALYSIS NETWORK MAP	93
TABLE 10 BASIC THEMES	94

Equations

EQUATION 1 MEAN BIAS ERROR FOR MEASURED ENERGY CONSUMPTION (NEXANT INC., 2008)	24
EQUATION 2 ROOT MEAN SQUARE ERROR FOR MEASURED ENERGY CONSUMPTION (NEXANT INC., 2008)	25
EQUATION 3 COEFFICIENT OF VARIATION RMSE FOR MEASURED ENERGY CONSUMPTION (NEXANT INC., 2008)	25
EQUATION 4 CALIBRATION SIGNATURE	34
EQUATION 5 CHARACTERISTIC SIGNATURE	35

1 Introduction

Buildings in most developed countries are responsible for around 40% of all energy use and 35-40% of energy-related CO₂ emissions (IPCC, 2007). If improvements in the energy related emissions of greenhouse gases are to meet the expectations of the Paris Agreement (United Nations, 2015), the existing built environment must be sufficiently improved to reduce the impact of CO₂ emissions.

This gap, between actual energy use and optimal energy use, is often regarded as the 'energy efficiency gap' (Jaffe & Stavins, 1994). If pursued, this could translate into valuable energy and cost savings for building owners and tenants as well as decreases in GHG emissions. Often these savings are worthwhile and can have small payback periods (Popescu, Bienert, Schützenhofer, & Boazu, 2012), however they are often addressed in isolation. The New Zealand Energy Efficiency Conservation Authority (EECA) estimates that up to 20% of the annual building energy use can be saved through little or no cost changes, with up to 50% reduction possible given substantial investment (EECA, 2015a, 2015b).

Energy Conservation Measures (ECMs) are the means by which energy is saved in buildings, such as improving the thermal envelope of a building or introducing energy efficient lamps. The design ECMs are often suggested and modelled as singular improvements to a building rather than as a combined set, making progress slow and tedious. Olgyay & Seruto (2010) write that design teams in the United States typically do not optimise whole-building retrofits, but rather look to make simple savings, leaving larger whole building

savings untouched. They found that the current retrofit rate in the United States is around 2.2% of the building stock with an average level of savings of just 11% per building.

If meaningful change is to be made, ECMs need to be packaged together into a whole building retrofit approach (Olgyay & Seruto, 2010), though it is very difficult to reliably estimate how effective these changes will be. It is even more difficult to say how one ECM might relate to the effectiveness of another due to their possible interaction. E.g. a heating equipment upgrade may improve the energy efficiency by, say, 15%, while a cooling-system upgrade may improve the energy efficiency by 10%. This does not necessarily mean a 25% saving can be guaranteed or even predicted. Shapiro (2011) found that 53% of building energy audits overstate the amount of savings that can be made. Additionally, 80% of audits miss possible improvements that can be made.

So how do we quality assure this process to identify and then make meaningful savings? What tools are available for industry practitioners to help assess the effectiveness of these changes?

1.1 The state of industry

One promising tool that can be used to reliably model and suggest different combinations of ECMs is the Building Energy Model (BEM) (Zhao & Magoulès, 2012). BEMs are a simulation of the real building, an electronic representation used to investigate possible changes to the building (Waltz, 2000). BEMs can be used in early design of new buildings or, as is the focus of this research, in the design of retrofits for existing buildings.

In creating a BEM, the designer can investigate and change the building to be more efficient and comfortable for its occupants. But first the BEM needs to represent the actual building as closely as possible incorporating its form, constructions, energy consuming systems such as the lighting and HVAC (Heating Ventilation and Air Conditioning) and patterns of use. However, this task is very difficult, time consuming and complex (Zhao & Magoulès, 2012).

As evidence to how variable this process can be, (Berkeley, Haves, & Kolderup (2014) observed trained professionals creating a BEM of an existing building.

They investigated the assumptions and the impacts of modeller decisions on the accuracy of the BEMs produced. They found between -11% to +104% difference in predicted electricity consumption and -61% to +1535% difference in predicted gas consumption. Other researchers have found that the availability of information and the influence this has on the modellers' assumptions heavily impact the final outcome of simulation (Alajmi, 2012; Bradley, Kummert, & McDowell, 2004; Guyon, 1997).

A survey of industry in the United States by Tupper, Franconi, Chan, Hodgins et al. (2011) looked at the state of BEM in industry, and major issues within the market. Four main issues were found relating to the use of BEM in industry. Firstly, a credibility gap exists where the quality of BEMs and the expectations surrounding what they can produce are limited. Secondly, the difficulty and time need to create BEMs leaves limited time for the designer to think critically about the project as they have to spend more time creating the model than is desirable. Thirdly, there is a need for more experienced and skilled practitioners. And lastly the low expectations of the market lower the value of BEM and consequently the demand for BEM use.

While it is difficult to pin down exactly the root cause for these issues, researchers have found that the level of quality assurance techniques (BEM validation procedures) and BEM software design are two areas of major concern internationally (Tupper, Franconi, Chan, Fluhrer, et al., 2011; Tupper, Franconi, Chan, Hodgins, et al., 2011). As yet, no such study has been undertaken in New Zealand. While this research does not fill this gap, it does provide some additional context for future research.

1.2 Difficulties in BEM validation

BEM validation is also known as calibration i.e. the process of "tuning" or calibrating the various inputs of the model so that measured energy consumption matches closely with the building being modelled (T. A. Reddy, Maor, & Panjapornpon, 2007a). However, this does not mean that matching annual energy end-uses through a combination of randomly changed model inputs will meet the definition of calibration. Calibration requires a systematic, evidence-based technique to uncover the inputs most likely to be incorrect and

correcting them so as to produce a model whose energy flows or 'behaviour' matches the building being assessed (Coakley, 2014; Cory, 2015; Raftery, 2011). Simply tweaking the inputs in an ad-hoc manner is not enough.

Tupper et al., (2011) suggested that model validation is often inaccurate in industry due to: a lack of standardised methods; She states that there is a risk of BEMs being over-specified (i.e. there are many inputs increasing the number of assumptions likely to go into the model thus increasing the chances of mistakes and inaccuracy). She also states that the difficulty in accessing the building's historical energy data from which to assess the level of calibration; and finally the lack of integrated tools and methods to assist in calibration. The consequence of this is that industry professionals resorting to inexact 'fudging' and 'tuning' of the model to get the desired result (Coakley, 2014). This is akin to turning input dials randomly until the 'desired' (i.e. matched) effect is produced. Reporting of calibration methods are often vague or simply not present, resulting in the BEM being unrepresentative of the building being modelled.

However, there has been an increase in the number of systematic evidence-based techniques to calibrate energy models in recent years. Three different literature reviews, see (Coakley, Raftery, & Keane, 2014; Fabrizio & Monetti, 2015; A. T. Reddy, 2006), have summarised this work. A list of over 20 different approaches and their limitations is shown in Table 3.

Despite this work, many approaches are unlikely to be adopted by industry in the immediate future without further refinement. Some methods involve the use of computationally intense statistical and sensitivity analyses, optimisation algorithms, and use hundreds if not thousands of models making their adoption unlikely due to time and budget constraints making their adoption unlikely due to time and budget constraints (Coakley, 2014; T. A. Reddy et al., 2007a). Others involve methods such as Bayesian probability sampling and Monte Carlo simulation (see section 2.4.2) likely go beyond the knowledge (and desire) of the average practitioner (Fabrizio & Monetti, 2015). The consequence of these findings are that there are some significant credibility issues in the results from some BEMs.

1.2.1 Industry/research disconnect

Despite the recent advancements, it would appear that very few, if any, calibration techniques are making their way from academia and research to industry. A review of 4 industry modelling guidelines and protocols (see page 44) revealed that common practices have changed little since Tupper et al. (2011). How then can the practitioner calibrate their BEMs if the tools are not available to them? Moreover, are industry professionals even likely to add this additional step to the already time constrained process of creating and using a BEM?

There is little research connecting development of BEM calibration tools to the needs of industry. Consequently, there are few methods available to industry beyond the ad-hoc process so lamented in the literature (A. T. Reddy, 2006). Personal correspondence with industry organisations in New Zealand such as the New Zealand Green Building Council (NZGBC), the Energy Management Association of New Zealand (EMANZ), the Energy Efficiency Conservation Authority (EECA), and the Building Research Association of New Zealand (BRANZ) has suggested that in the few cases where BEM is used to investigate ECM opportunities, it is highly unlikely that systematic calibration processes are undertaken. Correspondence with professionals in Australia and the United States was consistent with this finding.

However, that is not to say that there are not methods that could be applied to industry. A method developed at Victoria University of Wellington by Dr Shaan Cory (2015) was supported by the Building Energy End-use Study (BEES) (Amitrano et al., 2014) to calibrate a statistically representative cross-section of 48 non-domestic buildings. Cory's method borrows from two previous calibration methods. The first, by Raftery (2011) uses an evidenced based methodology to provide transparency to the calibration process. It uses detailed building information to generate detailed BEM which is largely reflective of the building being assessed. The second method, an energy signature method by Liu et al., (1998) was further developed by Bensouda (2004) and is used to identify and correct incorrect input parameters based on the energy signature method (see section 2.5.2). This method, though relatively simple, is unrefined and unintuitive (i.e. difficult to understand and use for the

uninitiated) but could be easily transformed and streamlined into a quick, understandable and usable tool for industry use.

1.3 Research focus

This research is based on assessing the suitability and usability of the calibration method developed by Cory from the user's perspective. The Cory method is a fast and computationally efficient method for calibrating energy models. It uses a number of simple techniques to calibrate energy models which may make it a suitable method for energy assessors in practice. It can be completed by a single user on a single computer in a short amount of time and does not necessarily require a large number of simulation runs. It also uses building physics concepts familiar to the practitioner meaning the process is less esoteric than other approaches. This may make it a more likely candidate for energy assessors to use in practice than other methods.

This research has identified that there is a limited body of knowledge in BEM calibration particularly from the perspective of the industry professional. As such, this thesis is primarily exploratory, looking to open areas for discussion and further research. This thesis can be considered a pilot or investigative study, rather than a confirmatory study.

1.3.1 Problem Statement

A lack of BEM calibration guidelines, procedures and tools is a major contributing factor to the BEM credibility gap (Tupper, Franconi, Chan, Fluhrer, et al., 2011). While there are some calibration methodologies have been researched to fill this void, few have been successful in reaching industry in any meaningful way despite often being effective and novel solutions (Fabrizio & Monetti, 2015).

Problem statement:

Few researched methods for calibration are suitable for immediate use in industry. As such, the lack of a quick and usable Building Energy Modelling (BEM) calibration tool for use in industry is contributing to the BEM credibility gap.

The lack of a usable calibration tools is inhibiting the use of BEM in industry

Understanding how practitioners work, their needs, and thoughts on the use of BEM and on the adoption of calibration techniques is critical to improving the links between industry and researchers.

1.3.2 Research Questions

1. *Is a lack of a fast and usable calibration process inhibiting the use of BEM in industry?*

This question looks to establish how industry currently operates. This will serve as context for the second research question, which looks to respond to the problem statement by evaluating the suitability of the Cory method for uptake in industry:

2. *Would the introduction of the Cory method improve the outcomes of energy modelling in industry?*

The third question asks which features of the calibration tool are useful to industry and whether future development is required

3. *Which aspects and features of calibration tools require development in order to increase industry uptake of BEM quality assurance*

1.3.3 Research method

The research method is split into two distinct phases. The first phase evaluates the current Cory method through a week long usability workshop with two post-graduate Building Science research students. It then makes improvements to the method, streamlining it into a more understandable and usable tool ready for the second research phase.

The second research phase looks to answer the research questions through a series of individual interviews and workshops with expert industry participants. This phase explores how industry currently conducts their BEM work for existing buildings, what, if any, systematic calibration procedures they use, and how suitable the streamlined Cory method would be to fulfilling their needs.

2 Literature Review

A number of different factors influence the ability of practitioners to use BEM effectively in practice. Generally, this can be broken down into three key areas:

- Firstly, the influence of the market and desire of building owners to invest in projects requiring the use of BEM. The size of the market can play a large role in the methods practitioners will invest in, such as BEM.
- Secondly, the availability of data to the modeller, which has a major influence on how well they can model the building. Often it is the case that building data is unavailable which leads to the third factor;
- The techniques researched in the literature are not accessible to the practitioner thus reducing their ability to correct mistakes in a systematic way.

2.1 Buildings in New Zealand

The Building Energy End-use Study (BEES), conducted by BRANZ, is the largest and most comprehensive energy survey of the New Zealand commercial building stock to date. The BEES research started in 2007 and ran for 6 years, gathering information on energy and water use through surveys and monitoring of non-residential buildings, specifically commercial and retail buildings. This includes baseline estimates on the number of buildings, total energy use in New Zealand and average energy use intensity, understanding the current attitudes toward resource use in buildings and what areas energy efficiency adoption could occur (Amitrano et al., 2014).

One feature of the BEES data analysis is the ratio and energy efficiency gap between new and existing buildings in New Zealand. It has been estimated that existing buildings outnumber new buildings by 50 to 1 in New Zealand (Hills, 2014) with over 80% of these buildings being built before the new millennium (Amitrano et al., 2014). Furthermore, these older buildings tend to be less energy efficient with newer buildings consuming about 40% less energy than older ones (Cory, 2015). Given the already established global GHG emission problem stated by the IPCC (IPCC, 2007), research suggests that the focus for change in New Zealand should be on existing buildings (Amitrano et al., 2014).

2.2 NABERS energy efficiency assessment methods

In 1999 Australia introduced the National Australian Built Environment Rating System (NABERS) for use on commercial buildings. The scheme was introduced as a commercial building performance rating system with the aim of reducing the impact of buildings on the environment, with an initial focus on greenhouse gas emissions (Office of Environment and Heritage, 2014b). At the time of writing, the scheme has four main rating tools measuring: Energy Performance, Water Use, Waste Management, and Indoor Environmental Quality (IEQ). Of the four tools, the energy efficiency tool is the most widely recognised and used (Office of Environment and Heritage, 2014b). The energy efficiency tool rates buildings from one to six stars based on annual energy use when compared against a benchmark score.

The NABERS energy (performance) tool measures actual energy use of a building over a nominal 12-month period. All energy used in the building is required to be metered for assessment. The energy use is then multiplied by a factor to establish the equivalent CO₂ GHG emissions in kg.CO₂/m²/yr. NABERS uses this index along with a location benchmarking factor to account for variance in local climate to calculate the final rating star scale as shown in Figure 1. A 2.5-star rating denotes average performance.

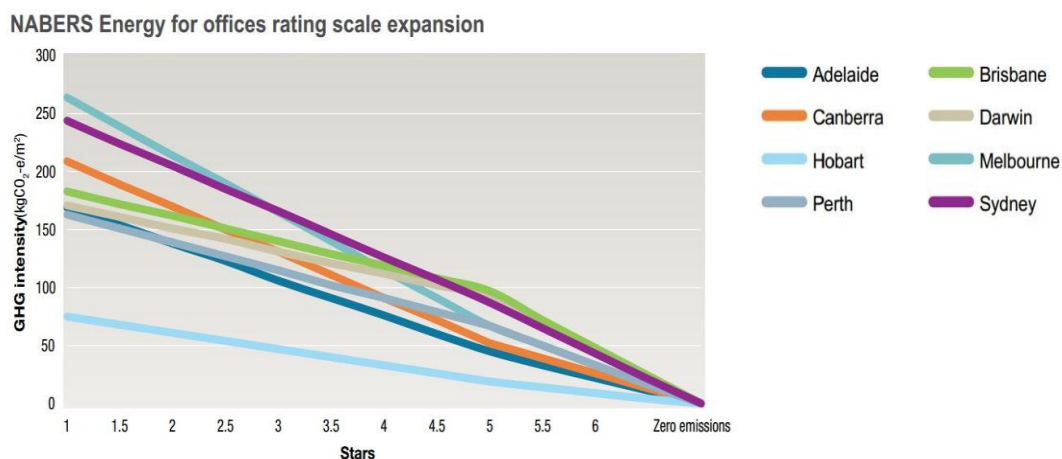


Figure 1 Benchmark factors and star scale for NABERS Source: (Office of Environment and Heritage, 2011)

In Australia, the NABERS energy tool has resulted in large reductions in GHG (averaging 8.5% per building or about 30kWh/m².yr), as well as having a sizable effect on the commercial property market. Owners promoting buildings rated at 2.5 stars or below generally have to give discounted leases, while those rated 3 stars or above generally charge a premium as well as receiving reduced vacancies and longer leases (Investment Property Databank Ltd., 2013; Newell, MacFarlane, & Kok, 2011).

Growing interest in Australia for increased sustainability measures in buildings has allowed the NABERS rating to influence the tenancy market and the overall building value. A report by Newell et al. (2011) describes the impact of different ratings on a buildings rent, value, vacancy, incentives, yield and outgoings as summarised below in Table 1 as a percentage of baseline levels. Incentives are calculated in terms of the number of months at low or no rent based on a 10-year lease. The most significant advantage for building owners comes from 5+ star rated buildings indicating that market demand is for high performance buildings. The lower end buildings (i.e. lower NABERS star ratings) are also less stable than those on the higher end of the scale with shorter leases and increased need to incentivise prospective tenants.

	Star rating			
	2/2.5	3/3.5	4/4.5	5+
Gross rent	0.9%	0.2%	-0.3%	0.6%
Value	1.1%	2.5%	2.6%	8.7%
Vacancy	-4.1%	-4.7%	-6.7%	-5.7%
Incentives	5.8%	2.4%	2.1%	0.2%
Yield	0.0%	-0.04%	-0.05%	-0.15%
Outgoings	5.2%	1.7%	-2.7%	-6.5%

Table 1 Financial impact of green premiums in NABERS rated buildings (adjusted for floor area) Source : (Newell et al., 2011)

Uptake was very slow when NABERS was first introduced in 1999, with voluntary adoption taking place mostly in one-off office ratings (Bannister, 2011). There was no motivation for the market to take notice of the tool as there were no inherent benefits. This changed when the New South Wales state government required that all of its leased buildings to have a minimum 3.5 star rating for existing buildings and a 4.5 star rating for new buildings (Bannister, 2011). The gradual adoption of NABERS by building owners and developers was dependant on the security of long-term lease agreements with government departments (Bannister, 2011).

After 13 years of operation in Australia, NABERS was introduced in New Zealand in June 2013 as NABERS New Zealand (NABERSNZ) (EECA, 2012). The NABERSNZ toolkit provides building owners with a similar energy benchmarking system for commercial building owners to compare and track energy use. The system, though modified to suit the New Zealand context and climate, closely follows its Australian predecessor.

However, there has been limited involvement in New Zealand, with only 15 buildings rated in its first year of operation (Office of Environment and Heritage, 2014a). Despite the already stated benefits, in neither Australia nor New Zealand does the NABERS schemes offer advice on *how* to improve the energy efficiency of a building. This is left to the expertise of the building energy assessors and designers.

2.2.1 New Zealand methods of energy efficiency assessment

Currently if a building owner wants to improve their building they have to rely on the ability and experience of the energy assessor to provide a reliable answer. This may be simple for small incremental changes at low capital cost,

such as targeting the first 20% of possible energy savings (EECA, 2015b) but if a building owner wants to improve their building from, say, 2 to 5 stars it may be at higher cost and may also require multiple changes to the building .

While it is assumed that energy assessors are proficient in building energy use assessment, often, as appears to be the case in New Zealand, they tend to calculate savings using empirical analysis in the form of static steady-state spreadsheets. Personal correspondence with the NZGBC (October 6th, 2015) noted that in the majority of cases spreadsheet tools were used to estimate savings of multiple ECMs by simply adding the calculated values together thus ignoring interactive effects. Further correspondence with EMANZ (October 7th 2015) that suggested that BEM is used by only a few practitioners for existing building energy efficiency analysis helped confirm these suspicions. The problem with this kind of empirical analysis is that it can be heuristic, relying on past experience to gauge ECM savings (Ma, Cooper, Daly, & Ledo, 2012).

For instance, sizing of HVAC plant may be done through the widely used 'degree day' method, which sizes plant based on the Heating (or Cooling) Degree Day (Al-Homoud, 2001), a measure of the required heat energy below which a building needs to start heating. The difference in desired indoor and outdoor temperature is used to determine how many Heating Degree Days and thus how much energy is required to heat the building. This is done over a small period of time usually a few days in the year based on historical temperature records. The historical maximum and minimum temperature is used for these set days (Al-Homoud, 2001; Thom, 1954). This method has been in use for many years, but has a number of limitations, the most obvious of which is its static nature. Temperature and time are continually changing along with the amount of heat gains and losses in the building, therefore considering this process as static, with temperature and time fixed to only a few points across the year is unlikely to produce the most efficient answer.

The most common method for evaluating improvements to building energy performance is through empirical and heuristic auditing which is backed by the energy assessor's experience (Coakley, 2014). Static models have two major drawbacks; firstly, they do not provide flexibility in design, meaning changes tend to be based on what has worked in the past for similar cases rather than

being tailored to the specific building, and secondly, they are not particularly efficient in their design outcomes (Al-Homoud, 2001; Augenbroe, 2002). That is to say the design is not necessarily optimised to reduce GHG emissions.

2.3 Building Energy Modelling (BEM)

The need to reliably predict building energy use has been the focus of research for some time (Ma et al., 2012; Socolow & Sondregger, 1976). However, varying characteristics such as building location, building fabric, structure, operation, occupancy, and the interaction between building systems make contrasting and comparing buildings a very difficult task (Ma et al., 2012; Zhao & Magoulès, 2012). Adding to the complexity is the criteria beyond energy use that require 'satisfying' in order to achieve an optimal solution. Energy, thermal and visual comfort all need to be considered if a successful outcome is to be found (Cory, 2015).

Though the list of techniques to predict building energy use has expanded rapidly with the adoption of computer processing, some techniques have proven to be more fruitful than others. Zhao and Magoulès (2012) summarise the advantages and disadvantages of the various techniques with relevant examples. One technique described is Building Energy Modelling. A Building Energy Model is a computer model made to present or replicate the actual building, mapping its energy flows and operations (Waltz, 2000). They use complex calculations combining heat and energy loads, climate data, and building operation details such as occupancy patterns, to determine a far more precise calculation than manual spreadsheet methods. These models are usually simulated to account for all 8760 hours in the year, with numerous outputs and details being provided in the form of data sheets (Joe A. Clarke, 2001).

Building Energy Models (BEM) are law driven, meaning their calculations are defined by scientific phenomena (e.g. heat transfer) (Coakley et al., 2014). In effect, this means that they imitate how materials, occupants, equipment, weather etc. interact in real life, resulting in an approximation of actual building behaviour. Due to the enormous libraries of law-driven interactions, Building Energy Models can be used to model variations on the existing conditions. This

ability means that once they are able to reflect the current behaviour of a building, they can then be used to investigate various design options to improve the building performance, or to see what effect on performance various options might have. Data-driven models, such as a statistical analysis of historical energy data (Zhao & Magoulès, 2012), do not have this capacity (Coakley et al., 2014).

2.3.1 BEM tools

There are in excess of 400 different tools available for BEM ranging from simple spreadsheets to diverse software. A comprehensive online list is kept by the US Department of Energy (2015). (Crawely, Hand, Kummert, & Griffith (2008) created a list of the 20 most popular tools available in 2008, most of which are complex programmes that work at a very detailed level i.e. able to provide a detailed sub-system level or whole building energy use result. Generally, simulations will give very detailed environmental (e.g. temperature and humidity) and energy performance information. The laws that drive these models are simulated at discrete time steps, often hourly, over the course of a full year using a Typical Mean Year (TMY) weather file. Examples of such tools are Energy Plus, TRNSYS, IES VE and ESP-r (Crawely et al., 2008).

2.3.2 Advantages of BEM

BEM has a wide range of advantages over other methods, such as statistical or mathematical models, the most notable of which is the ability to *"predict system behaviour given previously unobserved conditions"* (Coakley et al., 2014). In other words, they have the ability to take a real building in its current form, and change it to observe alternative scenarios. This means that, *if* a BEM can be made to represent a real building to a trustworthy and reliable state, they can also be used to analyse the effects of proposed changes to that building. This makes simulation a very powerful tool for energy assessors.

2.3.3 Disadvantages of BEM

Uncertainty in BEM is described as being the elements of the model which are unknown or are unable to be accounted for (Coakley et al., 2014). Uncertainty will always exist in a model because it is considered *closed*, with no other 'real world' effects able to be taken into account. This is an intrinsic issue in simulation because the real world is open, meaning external factors (such as

the weather) can change alter how the building is operated. While the building data that goes into a BEM may be measured and accounted for, it is impossible to account for other effects such as human behaviour. Coakley (2014) describes four types of modelling uncertainty (as adapted from (A. T. Reddy, 2006)):

- **Specification uncertainty.** Where incomplete information from building systems, schedules, specifications etc. exist in the model.
- **Modelling uncertainty.** Simplifications made to the model or simulation process. Examples being simplified thermal zoning strategies, or simplifications to the simulation calculations
- **Numerical uncertainty.** Numerical errors in the model (typographical errors such as substituting "0.1" for "0.01"
- **Scenario uncertainty.** External factors such as occupant behaviour.

Scenario uncertainty cannot ever be accounted for perfectly, so it is often regarded as being unattainable. It can be argued that the goal of creating a BEM of an existing building is to get a good understanding of how the building performs rather than to perfectly replicate how the building was used (Ma et al., 2012). Meanwhile numerical uncertainty can be fixed with careful checking and re-checking of input values.

The first two types of uncertainty, specification and modelling uncertainty, are where care needs to be taken. They rely on access to the building information, and the ability of the modeller (and simulation program) to accurately replicate that data within the model.

While simulation can offer a great amount of flexibility and a richness in detail, they are notoriously over parameterised (too many inputs) and under determined (outputs that give a result, but do not help to describe how this result was determined) (Coakley, 2014).

This problem was noted by Hornberger and Spear;

“Most simulation models will be complex, with many parameters, state-variables and non-linear relations. Under the best circumstances, such models have many degrees of freedom and, with judicious fiddling, can be made to produce virtually any desired behaviour, often with both plausible structure and parameter values.”

- (Hornberger & Spear, 1981).

Research shows a major issue with BEM for existing building simulation is the replicability of real conditions and trustworthiness of simulation results (Coakley et al., 2014; T. A. Reddy et al., 2007a).

Specification uncertainty

Access to correct and detailed building information is a critical part of all building energy modelling (Fabrizio & Monetti, 2015). Usually base information is provided from a building energy audit (Alajmi, 2012; Coakley et al., 2014). Auditing is primarily about understanding the building energy flows and identifying opportunities for energy savings. Energy audits are often standardised process and guidelines defined such as ANSI/ASHRAE/IESNA 100-2006, ASNZS 3598:2014 and ISO 50002:2014 (ASHRAE, 2006; International Organization for Standardization, 2014; Standards New Zealand, 2000). Energy Audits aim to “provide essential information to determine how efficiently energy is being consumed, identify energy and cost saving opportunities and highlight potential process and productivity improvement” (Standards New Zealand, 2000). New Zealand auditing is covered by ASNZS 3598.1:2014

Three separate level of energy audit are defined in ASNZS 3598.1:2014, which mirror the ASHRAE 100-2006 process:

- **Type one Audit:** Energy Auditing in its most simplistic form involving a quantitative overview to identify low and no-cost retrofit or improvement opportunities with payback periods of up to two years. These are often used as scoping exercises for level two and three audits in larger and more complex buildings (EMANZ, 2015).

- **Type two Audit:** Type two audits “define a detailed level of audit involving a comprehensive review and analysis of equipment, systems, and operational characteristics of the whole building to enable quantifiable energy savings recommendations” (EMANZ, 2015). They investigate the energy use across the entire building including all systems and subsystems at a detailed level.
- **Type three Audit:** Type three audits detail “specific subsystems, with additional data gathering and measurement to provide a higher level of accuracy. Audits of this type are typically focussed on a process or subsystem level, such as for HVAC, building management systems, compressed air or lighting, rather than a whole site” (EMANZ, 2015). Type three audits tend to focus on single systems within the building however they still require large amounts of data to be collected in order to quantify costs and benefits of various capital investments.

Energy auditing, while not guaranteed to discover all possible areas for improvement, has been highlighted in a number of studies as a critical point in the retrofit process because it reduces specification uncertainty and increases the amount of information available to the modeller (Alajmi, 2012; Ascione, Bianco, De Masi, De’Rossi, & Vanoli, 2015; Ma et al., 2012).

Alajmi (2012) investigated the use of level one and two energy audits on a small building in Kuwait. The retrofit strategies in the level one audit, which outline no or little capital investment, would yield a 6.5% energy saving over a year, while the level two audit found more significant savings of 49.3% could be achieved but with a significant level of investment. These results were verified through the use of a highly detailed BEM.

Ascione et al. (2015) describes five categories of information needed from an energy audit for use in simulation; building form and geometry, characterisation of the envelope and building materials, building system information such as HVAC plant, occupancy patterns and internal loads, and historical energy data. In addition to this, diagnostic surveys were used to identify inefficient equipment and improper control schemes which resulted in increased inefficiencies. The addition of this diagnostic information proved important for the EnergyPlus BEM in their study.

Despite the large savings that can be provided through good energy management and auditing, a large percentage of these are not realised or even identified (Shapiro, 2011). Roth, Westphalen, Feng, Llana, & Quartararo. (2005) found that building faults, such as duct leakage and systems being left on during un-occupied hours, can account for 2 to 11% of building energy use. Though these faults may be difficult to detect when performing a type one or two audit, they can impact on the areas of improvement suggested by the auditor. In turn it is impossible to model this information in a BEM if it is unknown due to it being outside of the scope of the energy audit.

In addition to this, the impact of using 'default' or recommended input values in a BEM can have a large impact on the reported end-energy use. Daly, Cooper, & Ma (2014), found default values for hard to measure inputs found in Australian simulation guidelines and protocols could result in large variations in predicted energy savings. On a single case-study they found that using different assumptions could result in a 2.4 to 10.3-year payback period on simple lighting retrofits.

2.3.3.1 Modelling uncertainty

The impact of the modeller and their assumption have long been noted as important aspects to simulation (J. A. Clarke, Strachan, & Pernot, 1993). While the practitioner can have a rich understanding of the building they are assessing, there are some aspects that may be missed. When information is lacking, educated estimates are often used (Berkeley et al., 2014).

A number of studies have looked at the effect the modeller can have on simulation (Berkeley et al., 2014; Bradley et al., 2004; Guyon, 1997; Ibarra & Reinhart, 2009). Though these examples are not all energy simulation related, all found significant variation in results when giving modellers the level of freedom described by Hornberger and Spear (1981).

Berkeley (2013) studied the impacts of modeller decisions of 12 professional energy modellers. The modellers were asked to model a building in eQuest (a simulation program) using a set of documents given to them detailing the architectural drawings, HVAC plans and schedule, lighting plans and schedule etc.

Berkeley (2013) found the most important parameter values when trying to predict electricity consumption were HVAC, followed by internal loads density, and then by schedules, and these differ when predicting gas use (see Figure 2 and Figure 3). These were based on a series of sensitivity analyses of the results found in the assessment of modeller decisions. The sensitivity analyses Berkeley used looked at the combined impacts of different modelling assumptions made by the 12 professional modellers. The assumptions were then extrapolated into a series of 'decisions' about which assumption was likely to have the most impact on the model results. These decisions were run through the sensitivity test resulting in decision 'votes'. The 'votes' in Figure 2 and Figure 3 represent how these decisions most affect the model.

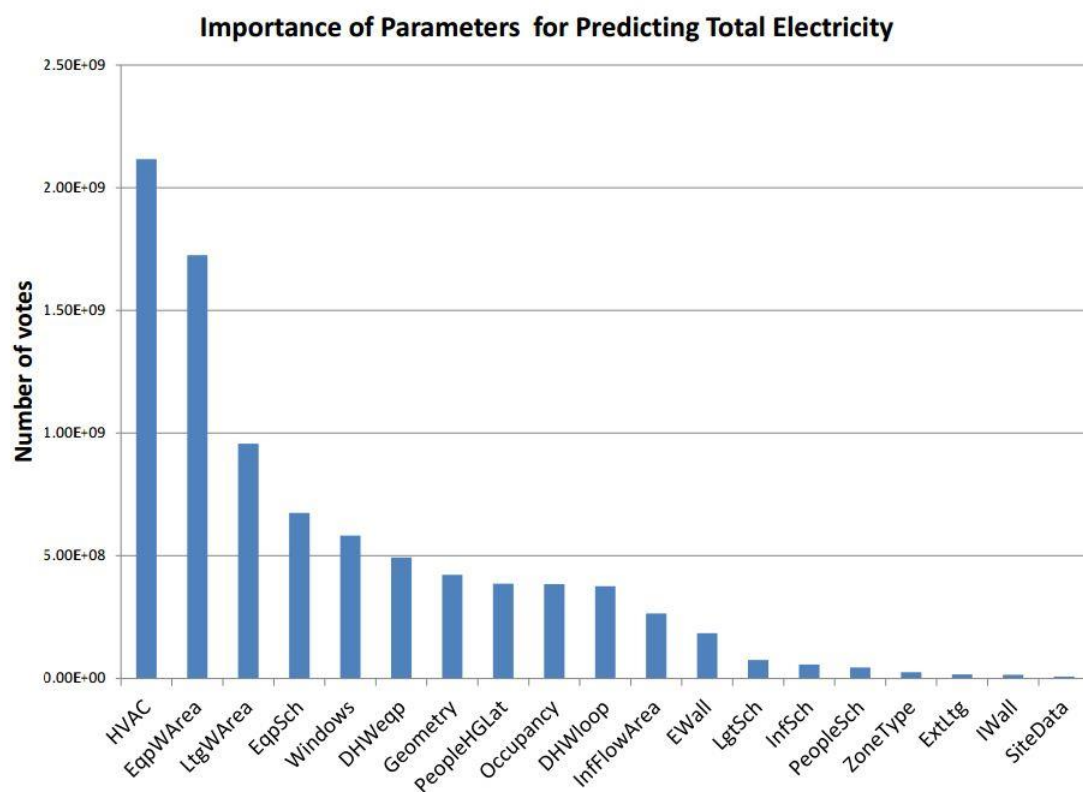


Figure 2 Importance of parameters for predicting total electricity use (Berkeley, 2013)

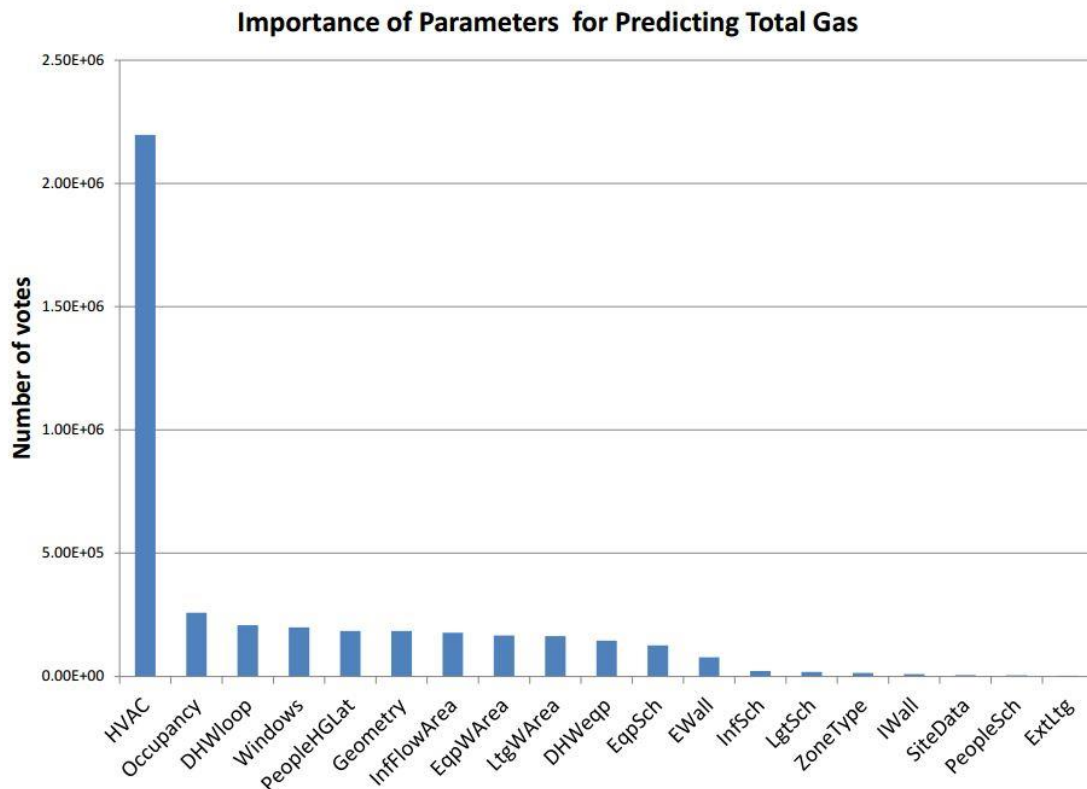


Figure 3 Importance of parameters for predicting total gas use (Berkeley, 2013)

Using the BESTEST procedure, an industry standard simulation program validation methodology, Bradley, Kummert and McDowell. (2004) tested the assumptions of three different expert modellers against the ANSI/ASHRAE Standard 140-2001 (ASHRAE Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs) using the simulation program TRNSYS. They found that expert modellers, with different approaches and simulation backgrounds, interpreted elements of the same prescriptive standard differently though without significant differences in the simulation results. While this case found slight differences in expert modelling practice, the case for novice modellers is not as clear-cut.

Ibarra and Reinhart (2009) tested the ability of two daylighting programs, RADIANCE and ECOTECT, to produce accurate daylight factor predictions based on 69 novice/student modellers. They also tested the impact of common novice mistakes in daylighting models. They tested for modelling mistakes in two different workshops; one held in 2005 where modellers were not given any instruction, and one in 2006 where modellers were given expert modelling tips. The modellers in 2006 had considerable improvements in accuracy over their 2005 counterparts. However, the 2005 student modellers tended to make

dramatic errors and to overlook a number of important parameters when modelling, particularly building geometry accuracy. Even when some parameters were addressed, they continued to get inaccurate results due to the impact of other variables. The authors highlighted the need for even simple modelling guidelines to assist in any simulation workflow.

Guyon (1997) tested 4 categories of modellers (12 modellers in total) on their ability to follow a validation exercise on the simulation program CLIM2000 when modelling energy use in a house. The modellers ranged from users familiar with the program to users with little familiarity, but with additional support. The results showed differences in energy prediction ranging from -41% to +39% representing an absolute value of a 2:1 difference. The modelling processes taken by the 4 different categories of participants resulted in some similar parameters being modelled incorrectly, but to varying degrees within each group.

The role of the modeller is an important factor in simulation. The next question is 'how do you remove these modeller impacts/biases from the process in order to produce a valid model?' A method is required which corrects faults and incorrect inputs independently of the modeller in order to reduce fudging and tinkering. In other words, a systematic process of input parameter correction is required. This has been a problem since the early days of simulation (J. A. Clarke et al., 1993).

2.4 BEM Calibration

The level of uncertainty in BEM is important enough to warrant systematic approaches to correction and this occurs through the process of calibration (see section 1.2).

Using the example of a simple equation to represent a BEM simulation, the problem of calibration can be defined in simple terms:

$$5 \times 6 + 2 = 32$$

as does

$$4 \times 7 + 4 = 32$$

While the result is the same, the inputs (on the left-hand side) are different. Matching looks to ensure that the result of the equation looks as it should on the right-hand side (see section 1.2) whereas calibration aims to ensure the operation (left-hand side) is as similar as possible, thus making both sides of the equation match. Expanded to the context of a building with thousands of input parameters and non-linear relationships the problem is magnified.

While calibration is used to reduce the level of discrepancy between measured energy consumption and simulated energy consumption, Coakley et al. (2014) notes that the issue is not in reducing these discrepancies to match energy consumption levels (the result), but is reflecting the 'behaviour' of the building itself which is the driver for the consumption patterns (the operation). Matching occurs when the energy used in the model matches the energy use in the building. Calibration occurs when the simulated building behaviour matches the real building behaviour.

2.4.1 Statistical Indices for Calibration

Model calibration limits are defined by ASHRAE guideline 14 (2002) and IPMVP (2001). The ASHRAE guide gives time-step tolerance levels outside of which a model cannot be considered calibrated. These limits are set between the real and simulated data for specific times periods e.g. hourly, monthly and annually. The two main metrics to assess this tolerance are the Mean Bias Error (MBE) and the Coefficient-of-Variation of the Root Mean Squared Error CV (RMSE).

2.4.1.1 MBE

Mean Bias error calculates how closely the model meets the actual measured energy consumption, and gives the degree to which the model over or under-simulates this amount. Positive values indicate over-prediction, while negative numbers indicate under-prediction.

$$MBE(\%) = \frac{\sum_{Period} (S - M)_{Interval}}{\sum_{Period} M_{Interval}} \times 100$$

Where:

M is measured kWh or fuel consumption during time interval

S is the simulated kWh or fuel consumption during time interval

Equation 1 Mean Bias Error for Measured Energy Consumption (Nexant Inc., 2008)

However, this can result in a cancellation effect, where one under predicted and one over-predicted month can falsely indicate calibration. To account for this, the CV(RMSE) metric is also used.

2.4.1.2 CV(RMSE)

The CV(RMSE) indicates the overall uncertainty in the model prediction as it only uses positive values to account for all over and under-simulation results. As such, the lower the CV(RMSE), the closer the level of calibration.

*"The CV(RMSE) is a normalized measure of variability between two sets of data. For calibrated simulation purposes, it is obtained by **squaring** the difference between paired data points, **summing** the squared differences over each interval through the period, and then **dividing** by the number of points, which yields the mean squared error. The square root of this quantity yields the root mean squared error (RMSE). The CV(RMSE), is obtained by dividing the RMSE by the mean of the measured data for the period."*

- (Nexant Inc., 2008) (emphasis added)

In simpler terms, the RMSE is found by squaring the summed the remainders of simulated data minus the measured data. This is then divided by the number of intervals e.g. 12 if calibrating annual data on a monthly basis. The square root of the result is taken to produce the Root Mean Squared Error.

$$RMSE_{Period} = \sqrt{\frac{\sum (S - M)_{Interval}^2}{N_{Interval}}}$$

Where:

M is measured kWh or fuel consumption during time interval

S is the simulated kWh or fuel consumption during time interval

N_{Interval} are the number of time intervals in the monitoring period

Equation 2 Root Mean Square Error for Measured Energy Consumption (Nexant Inc., 2008)

To find the Coefficient-of-Variation, the average measured energy consumption over the total period being examined is divided by the RMSE, and multiplied to give the final CV(RMSE) percentage.

$$A_{Period} = \frac{\sum_{Period} M_{Interval}}{N_{Interval}} \quad Cv(RMSE_{Period}) = \frac{RMSE_{Period}}{A_{Period}} \times 100$$

Where:

A_{Period} is the average measured energy consumption over the total examined period

Equation 3 Coefficient of Variation RMSE for Measured Energy Consumption (Nexant Inc., 2008)

2.4.1.3 Calibration limits

Table 2 shows the widely accepted calibration variation values for BEM modelling.

Calibration Type	Metric	Calibration Tolerance
Monthly	MBE_{month}	$\pm 5\%$
	$CV(RMSE)_{\text{month}}$	15%
Hourly	MBE_{month}	$\pm 10\%$
	$CV(RMSE)_{\text{month}}$	30%

Table 2 Calibration limits (ASHRAE, 2002)

2.4.1.4 Calibration issues

BEMs are inherently complex. To create a BEM, the practitioner has to firstly understand the building, and secondly understand how to replicate it in the model. Adding calibration to this makes the task longer and thus more expensive. Furthermore, a lot of trial and error can occur given the number of possible combinations that could incorrectly lead to a plausibly ‘calibrated’ model. To understand this process, the required expertise of the practitioner is likely to be greater than is normally required of them, and beyond what a client is expecting or willing to pay for.

2.4.2 Calibration methods

Calibration of simulation first began in the 1980s’ with work by (Carabott, 1989; Carroll & Hitchcock, 1993; Diamond & Hunn, 1981; Heidell & Taylor, 1985; Waltz, 2000) and used the same crude trial and error efforts described above. This work tended to use similar approaches, verifying simulation output data with empirically measured electricity and natural gas data. This usually consisted of a three step process: firstly, entering as much known building data into the model as possible, secondly, comparing the simulated data with the metered data, and thirdly, using experience and building knowledge to modify uncertain inputs. The process repeats until a satisfactory match was found as seen in Figure 4. However these processes did not use the same statistical indices as described earlier.

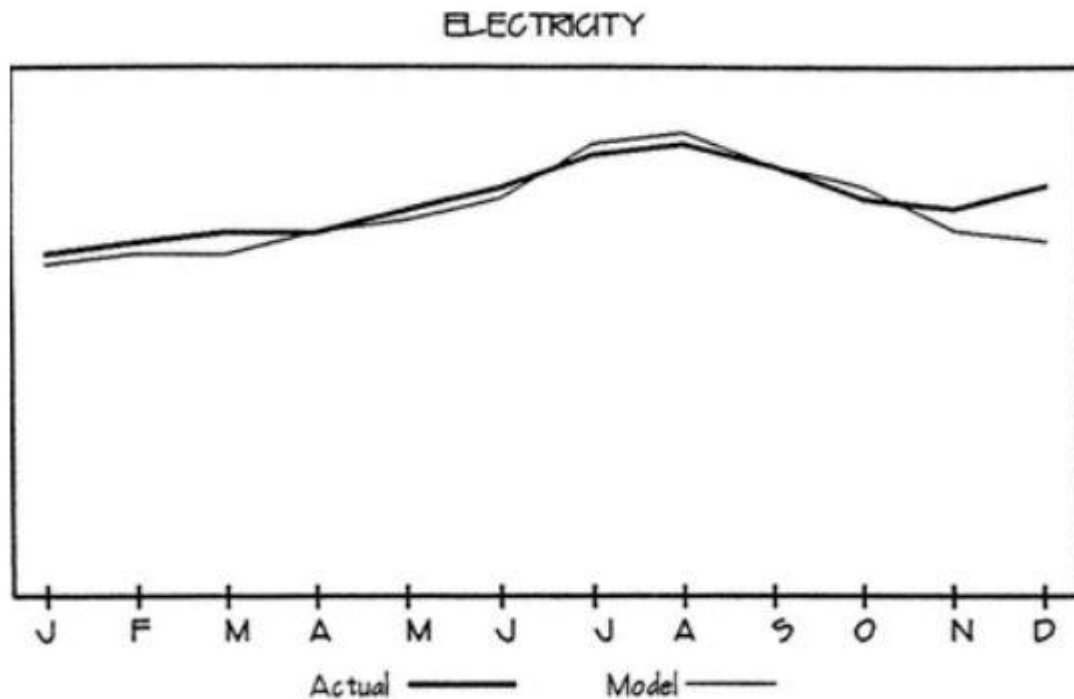


Figure 4 Measured vs actual electricity consumption example from (Waltz, 2000)

Though these methods were crude, given the simplicity of the models and the low resolution of the monthly utility meter data, they were often sufficient. Personal correspondence with the NZGBC (October 6th, 2015) and EMANZ (October 7th 2015) suggests that practitioners in New Zealand and Australia may undertake this level of analysis when looking to match simulated and measured data. This does not do anything to reduce the amount of uncertainty in the model because the process that guides it does not change.

More recent research has developed different methods for calibration, though these methods tend to be resource intensive, esoteric and are not often well understood. A number of literature reviews on the topic area have been conducted in the past few years (Coakley et al., 2014; Fabrizio & Monetti, 2015; A. T. Reddy, 2006).

Coakley (2014), describes calibration methods as being either manual (generally iterative and user driven) or automated (iterative and non-user driven). While there are no hard rules about what can be considered manual or automated, or where the boundary lies between the two, the classification serves well as a general rule of thumb.

- **Manual** approaches rely on iterative changes by the modeller to correct mistakes and reduce uncertainty.
- **Automated** approaches are not driven by the user as such. Rather they involve processes that either speed up the process of calibration to assist the user.

A range of approaches, from manual graphical techniques to automated optimisations and Bayesian probability sampling are suggested as appropriate techniques throughout the literature (Coakley et al., 2014; Diamond & Hunn, 1981; Fabrizio & Monetti, 2015). Coakley et al. (2014) describe a wide range of both manual and automated methods which can be seen in Table 3. An example of both calibration types, manual and automated, is presented below. Table 4 presents issues relating to each calibration process.

One of the first users of optimisation as a means of reducing the gap between simulated and measured data was Carroll & Hitchcock (1993). Using the difference between the simulated and measured data as the cost-function, they tweaked and changed inputs until this difference was minimised. However, they noted the danger in determining which combinations of input values were likely to be correct. This is the problem Hornberger and Spear (1981) noted, where virtually any plausible result could eventuate and can be considered as plausible as another. Carroll and Hitchcock selected the most likely combinations based on experience i.e. the combinations were heuristically defined, meaning that in the hands of a less experienced modeller the chosen combinations could be quite different.

Robertson, Polly, & Collis (2013) reviewed four different calibration methods against synthetic (not-existing) housing scenarios. The four methods compared were mathematical approaches, which were assessed for accuracy against the synthetic house data and then used to see if the calibration methods had an improvement on the accuracy of ECMs applied. While it was found that the use of calibration did improve the accuracy of the ECM predictions, they also found that significant errors can exist between the different calibration methods even when meeting the statistical definition of calibration i.e. the billing data can be matched for the wrong reasons, thus leading to a false positive. This leads to a conundrum of sorts. While the use of automated approaches is helpful and

more efficient than manual ones, they may still come up with a plausible yet incorrect combination of inputs. How then is the industry practitioner to ensure that this does not occur? One option is to use manual verification tests on automated approaches, or in place of automated approaches when convenient.

Characterisation techniques	
Name	Description
Detailed audit	Detailed audits are often conducted prior to building model development in order to gain a better knowledge of the building systems and characteristics (Geometry, HVAC systems, Lighting, Equipment, and Occupancy Schedules)
Expert knowledge / templates / model database	Approaches which utilise <ul style="list-style-type: none"> - Expert knowledge or judgement as a key element of the process. - Prior definition of typical building templates. - Database of typical building parameters and components in order to reduce the requirement for user - Inputs during model development
Intrusive testing	Intensive techniques require some intervention in the operation of the actual building, such as 'Blink Tests' whereby groups of end-use loads (e.g., plugs loads, lighting etc.) are turned on and off in a controlled sequence in order to determine their overall impact on the baseline building load
High-resolution data	Recorded at hourly (or sub-hourly) levels as opposed to utilising daily load profiles or monthly utility bill data
Short-term energy monitoring	Metering equipment is used to record on-site measurements for a short period of time (>2 weeks). This may be used in identifying typical energy end-use profiles and/or base-loads
Advanced Graphical Techniques	
Name	Description
3D-graphical comparison techniques	Three-dimensional graphs are used to aid comparison and/or calibration of measured and simulated data. This technique allows users to visualise large quantities of data, compared to traditional 2-D scatter plots etc. which are overwhelmed when analysing large quantities of data points
Signature analysis method	Signature analysis techniques are a specific type of graphical analysis technique, typically used by HVAC simulation engineers to identify faulty parameters in Air-Handling Unit (AHU) simulation. They may also be used to develop optimised operation and control schedules. Signature analysis methods are commonly used for the calibration of models based on the simplified energy analysis procedure (SEAP)
Statistical displays	This refers to the graphical representation of statistical indices and comparisons for easier interpretation. This includes data comparison techniques such as carpet plots, box whisker mean (BWM) plots and monthly per cent difference time-series graphs

Optimisation techniques	
Name	Description
Bayesian calibration	Bayesian calibration is an alternative statistical approach to model calibration. The approach offers the advantage of naturally accounting for uncertainty in model prediction through the use of prior input distributions
Objective / penalty function	Most mathematical techniques employ some form of optimisation function to reduce the difference between measured and simulated data. An objective function may be used to set a target of minimising, for example, the mean square error between measure and simulated data. Conversely, a penalty function may also be employed to reduce the likelihood to deviating too far from the base-case
Alternative modelling techniques	
Name	Description
Artificial neural network	Neural networks are computational models consisting of an interconnected group of artificial neurons. They are used for modelling complex relationships between inputs and outputs or for finding patterns in data
Primary and secondary term analysis and re-normalisation	Analytical tool for the meaningful estimation of parameters of a complex building from a few data channels over a short period (a few days). An Audit description of the building (capturing nominal building fabric parameters) is used to estimate heat-flows. These heat-flows are then re-normalised to satisfy an energy-balance equation
Meta modelling	The use of computationally efficient analytical surrogate models which emulate the performance prediction of their complex engineering-based counterparts
Simplified energy analysis procedure	The simplified energy analysis procedure refers to the use of simplified engineering models to represent the building. This may be accomplished by dramatically reducing the number of zones or AHU's in the model by grouping them together
Systems identification	This technique refers to the process of constructing models based only on the observed behaviour of the system (outputs) and a set of external variables (inputs), instead of constructing a detailed model based on first principles of well-known physical variables.

Table 3 Summary of calibration methods by (Coakley et al., 2014)

Category	Classification	Approach	Calibration Issues					
			Standards	Expense	Simplification	Inputs	Uncertainty	Identification and automation
Manual	Characterisation techniques	Detailed audit	x			x		
		Expert knowledge/templates/model database			x	x		
		Intrusive testing				x		
		High-res data				x		
		Short-term energy monitoring				x		
	Advanced Graphical Techniques	3D-graphical comparison techniques						x
		Signature analysis method			x			x
		Statistical displays	x					
	Model Simplification Techniques	Base-case modelling				x		
		Model parameter estimation			x	x		
		Parameter reduction			x	x		
		Data disaggregation			x	x		
	Procedural Extensions	Evidence-based model development	x			x	x	
		Sensitivity analysis	x			x	x	
		Uncertainty quantification					x	
Automated	Optimisation techniques	Bayesian calibration		x			x	x
		Objective/penalty function	x					
	Alternative modelling techniques	Artificial neural network		x	x			x
		Primary and secondary term analysis and re-normalisation			x			
		Meta modelling		x	x			
		Simplified energy analysis procedure		x	x			
		Systems identification			x	x		

Table 4 Issues relating to calibration techniques. Adapted from (Coakley et al., 2014)

2.5 Cory Calibration Method

The manual method developed by Cory (2015) relies on the combination of two approaches. The first, by Raftery (2011) which proposes an evidence-based methodology i.e. a detailed and transparent audit process to create the initial model, and the second, by Wei, Liu, & Claridge (1998) which was further developed by Bensouda (2004), use 'energy signatures' to help identify incorrect input parameters.

2.5.1 Evidence based approach

Raftery (2011), approached the problem of calibration by advocating for an evidence based process using version control to track the entire process. This provides transparency and improves the reproducibility and credibility of the calibration.

As part of the evidence based methodology, Raftery used a detailed building energy audit (similar to a Type two audit) with additional monitoring equipment to gain a detailed understanding of the building being assessed. This information is processed in a hierarchy, with for instance empirical measurements considered more trustworthy than building documentation data. Changes to the model are only made if they come from evidence, meaning access to building information is of critical importance. Raftery also suggested a zone typing methodology, whereby the geometry of thermal zones in a building are modelled to be as similar as possible to the actual layout of the building.

The result of this is a detailed BEM, with occupancy patterns, operation schedules for HVAC lighting etc., building constructions, temperature and CO₂ sensor data, HVAC plant and control strategies all mapped in fine detail. The approach, while resulting in an excellent level of calibration has two main flaws; firstly, the level of data required is so extensive that the collection would be difficult for practitioners to achieve in a timely fashion.

And secondly, the approach does not provide a strategy for identifying incorrect inputs beyond the evidence based approach. So when information is found to be missing, incorrect or unavailable, there is no systematic way of updating the model. Generally, this is likely to be missing or incorrect HVAC

inputs, which can have a large effect on the electricity and gas prediction of the BEM (Berkeley, 2013). Thus the initial BEM is made to be as close to the actual building as possible, with as many energy end-uses being reconciled as closely as possible to the energy audit. This is where addition of the second process developed by Wei et al. (1998) and Bensouda (2004) can be used.

2.5.2 Energy signature method

Wei et al. (1998) first developed the energy signature method, a process whereby input parameter 'signatures' are used to help identify faulty, incorrect or unknown parameters in HVAC systems. Two types of signatures are used: calibration and characteristic signatures:

"Energy signatures are graphical representations of the differences in heating and cooling energy consumption expressed as a ratio to the maximum baseline heating and cooling energy consumption and plotted as a function of the ambient temperature"

- (Cory, 2015) from (Bensouda, 2004)

The energy signature technique has been used successfully in a number of calibration studies (Bensouda, 2004; Cory, 2015; Kandil & Love, 2014).

Lastly, it is important to firstly remove all non-HVAC energy uses from the analysis before beginning the signature process i.e. the energy use being assessed must belong only to the HVAC. This helps to isolate the HVAC parameters ensuring the process is not affected by non-HVAC related energy consumption.

2.5.2.1 Calibration signatures

Calibration signatures represent the normalised difference between the actual building and **initial model** energy consumption as a function of outdoor air temperature. The creation of these signatures can be expressed by the equation:

$$\text{Calibration Signature} = \left(\frac{-R^C}{M^C} \right) \times 100$$

Where:

R^C = heating/cooling residual (between actual and simulated)

$$M^C = \text{Maximum measured Cooling energy}$$

Equation 4 Calibration Signature

These signatures are then represented graphically (Figure 5), with the **difference** in energy between the **initial model** and the **actual** building plotted as a percentage against temperature. From this plot, the validity of the model's HVAC inputs can be assessed. If there was perfect alignment between the model and the actual building, then the plotted line would be 0% for all temperatures. The size of a variation, and the temperature at which it occurs, can provide guidance as to what changes could be made to improve the fit.

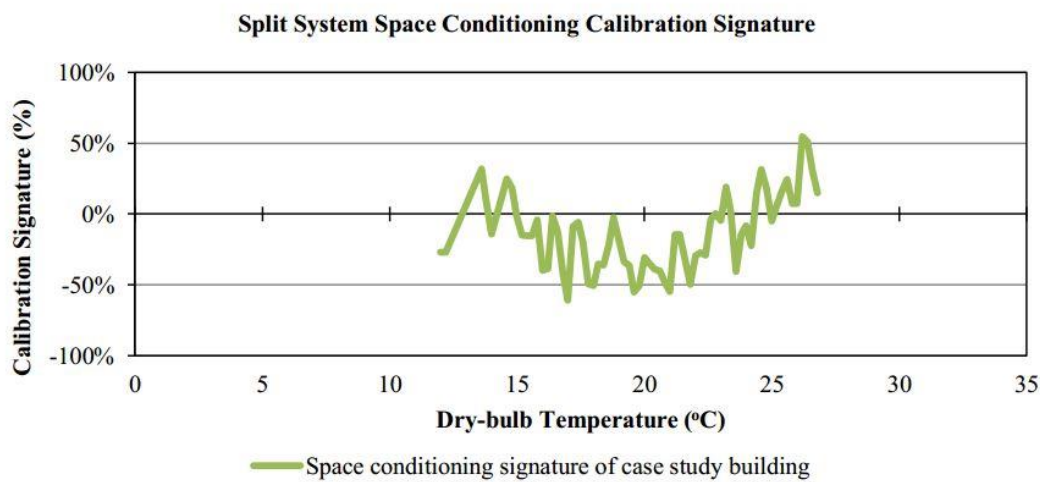


Figure 5 Example Calibration signature (Cory, 2015)

Figure 5 shows that energy use is generally over predicted in the model at temperatures below 15°C and above 25°C, and under-predicted between 15°C and 25°C.

2.5.2.2 Characteristic signatures

Once the calibration signature has been plotted, the characteristic signature can be plotted. Characteristic signatures are plotted for each individual input the modeller wishes to investigate, meaning there is often more than one characteristic signature. The characteristic signature is the normalised difference between the **initial model** and a new model which has one input varied, and is expressed by the following equation:

$$\text{Characteristic Signature} = \left(\frac{-C^c}{B^c} \right) \times 100$$

Where:

C^c = Change in heating/cooling energy consumption from the initial model and the new model

B^c = maximum baseline heating/cooling energy consumption

Equation 5 Characteristic Signature

The **change** in energy consumption is taken from the **new model** minus the **initial model** at the same temperature. This causes the characteristic signatures, each representing change to a single variable, to be of different shapes and sizes, which can then be compared visually to the calibration signature (Figure 6). Put simply, by comparing calibration and characteristic signatures for similarity and trend, incorrect inputs can be identified (Figure 7). This identification occurs by identifying a visual (or statistical) correlation between the two signatures.

Once an input has been identified as being possibly incorrect, the input value is changed depending on the energy use of the model when compared to the building.

For example, Figure 7 shows that the model is under-predicting energy use as the calibration signature is trending down as temperature increases. The cooling setpoint input (shown in the characteristic signature in Figure 6) can be seen to follow the same general trend as the calibration signature. Therefore, the cooling setpoint must be lowered in the model in order to increase the energy use so the model matches the building. Once the characteristic and calibration signatures correlate exactly (or as near as can be achieved), the input is considered corrected. The final goal is to get the calibration signature flat, or as close to zero as possible (Bensouda, 2004; Cory, 2015).

The energy signature process is reliant on the modeller choosing the right inputs by knowing which are assumed (and therefore uncertain). While this provides little insight into *how* to select inputs, it does serve to better inform the modeller by confirming or disconfirming their suspicions. The magnitude of change needed for each incorrect input is dependent on the modeller iteratively changing model and checking the correlation between calibration and characteristic signatures. It can also be used to confirm or disconfirm models used in automated approaches.

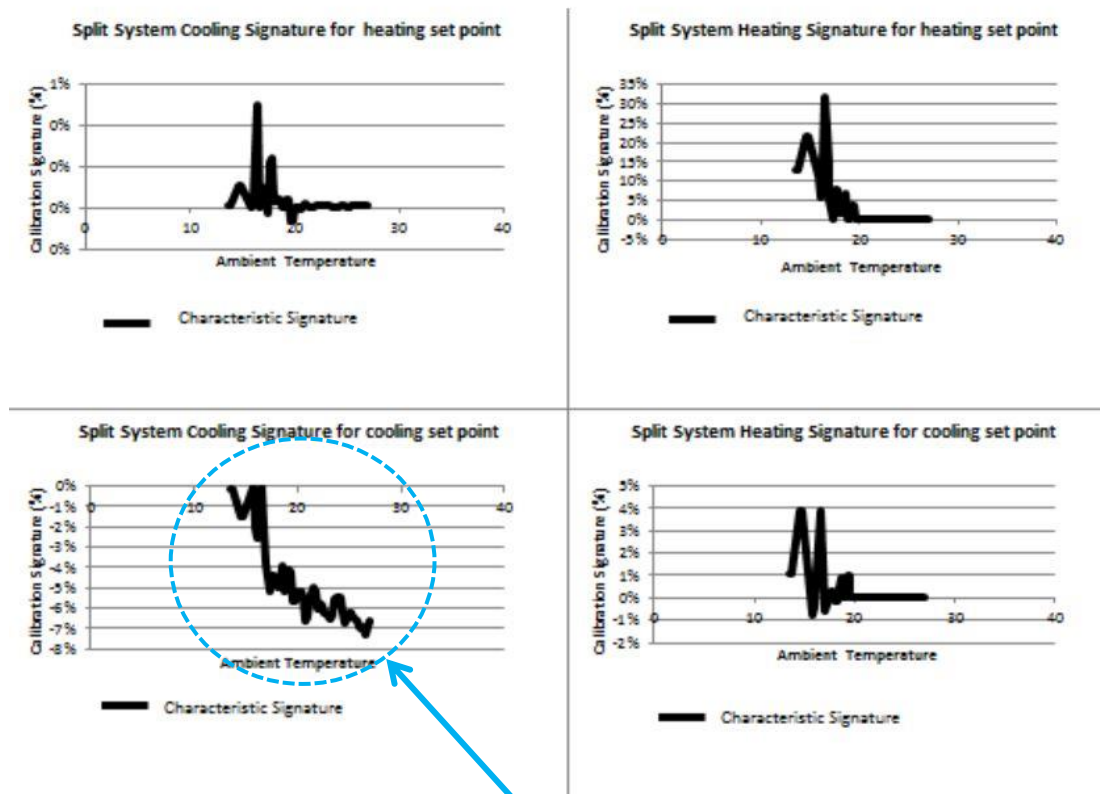


Figure 6 Example Characteristic signatures (Cory, 2015)

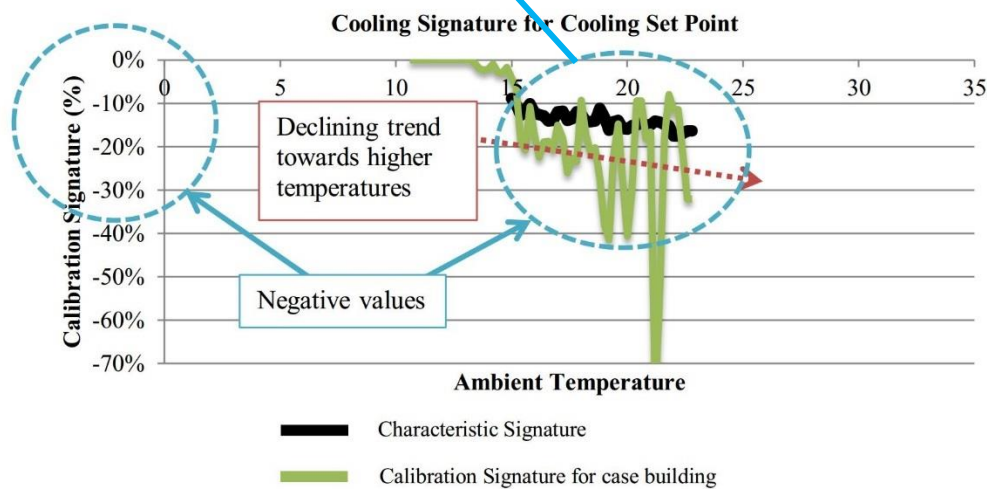


Figure 7 Example identification of an incorrect input parameter (Cory, 2015)

2.5.3 Calibration procedure

The model calibration procedure used by Cory is set out in 3 stages, as shown in Figure 8. Stage one is pre-simulation, stage two, calibration evaluations, and stage three is post simulation.

2.5.3.1 Stage one

Stage one involves gathering the required building information, such as would be found during a Type 2 building energy audit. This stage “involves a process of using genuine as-built information, surveys, and measured data to update the input parameters of the initial simulation model so that it closely represents the real operation of the building” (Raftery, Keane, & Costa, 2009). This information is then manually input into the model to as closely as possible match the building operation.

Once the model is updated, it can be simulated (Cory, 2015).

2.5.3.2 Stage Two

After the updated model is simulated the results are evaluated to see how closely it is calibrated with regard to the monthly and annual MBE and CV(RMSE) tolerances.

2.5.3.3 Stage three

Stage three involves taking the non-calibrated model and plotting its energy signatures. These are used to identify the incorrect input values from stage one. They also help to determine what input values might be correct for that particular input. The calibration signatures (**difference** between real data and initial model) are then tested against model characteristic signatures (**change** between initial and new models) to assess how closely they match. Stages 2 & 3 are repeated until the model is calibrated.

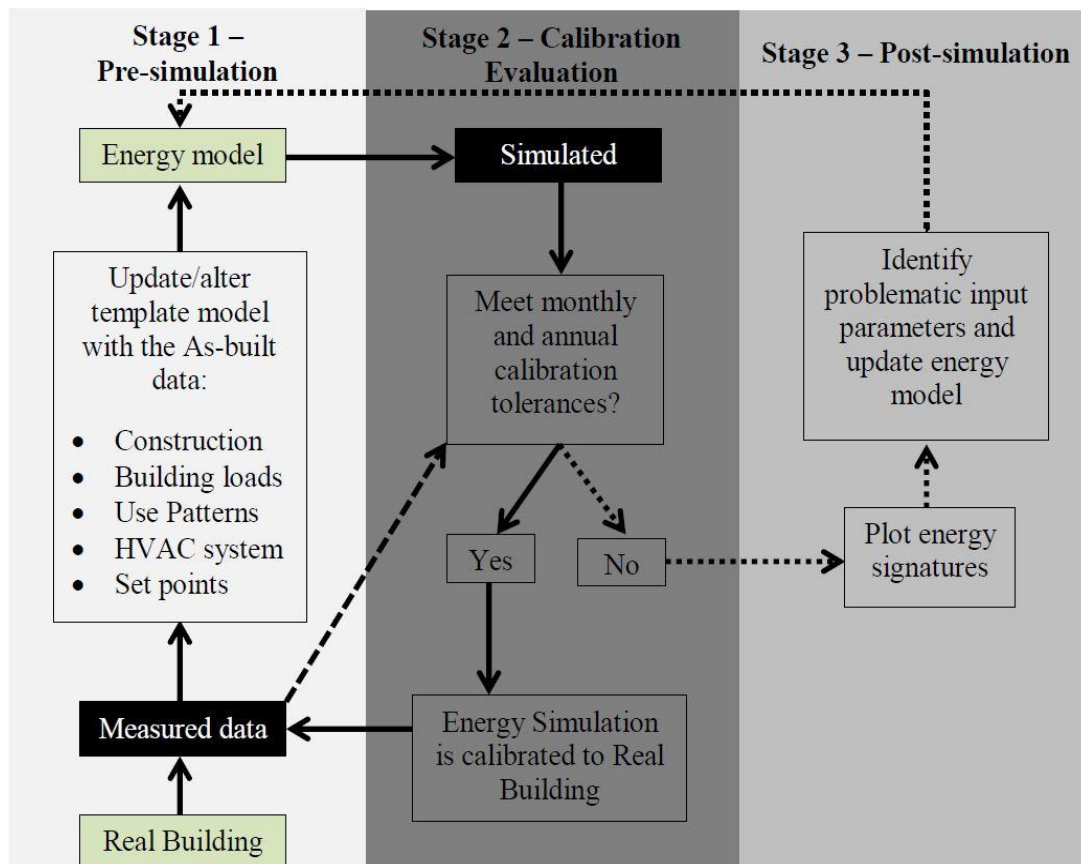


Figure 8 Calibration procedure from (Cory, 2015)

2.5.4 Advantages of the Cory model

The Cory model was developed primarily to satisfy the requirements of generating a representative model of the New Zealand Building stock based on 48 case study buildings (Cory, 2015). These models were calibrated using the monitored data collected from the BEES case study buildings.

The main benefit of this method is the ability to test the assumption inputs and have their results proved or disproved in a more systematic and scientific way than before. This helps to remove modeller bias and improve transparency and credibility while still keeping the modeller involved in the process. Provided the right signatures are selected, the level of modelling uncertainty is reduced.

2.5.5 Disadvantages of the Cory model

The validity of the calibration procedure relies on detailed records of energy use and local climate data. However, if the data collection has a low level of detail the calibration procedure is unlikely to be helpful e.g. if the time steps between data collection points (for both meter readings and climate data) are not close enough, they won't be able to provide the right amount of detail

necessary to differentiate the calibration and characteristic signatures. Additionally, they will not reflect the 'behaviour' of the building on a day-to-day level making it impossible to assess how a change in the model correlates with the actual building.

It is also possible that where the modeller does not choose enough parameters, or produces a model where more than HVAC related energy is being assessed in the energy signature, meaning the analysis may lead to possible false-positives if not followed correctly.

The Cory method also requires the user to have access to recent climate files from which to calibrate against as the usually used typical Mean Year (TMY) climate files are averaged over a period of 30 years of climate data and are therefore unrelated to the measured data used in the analysis. AMY (Actual Mean Year) files of weather data specific to the location of the building are a necessity.

2.5.6 Using the current Cory process

In its current form, the Cory method is a mixture of six different Microsoft Excel spreadsheets, each serving a different purpose. Beyond calculations and some limited rearranging of data, they do not provide users with a 'tool' as such. It is what might be classified as a 'pre-alpha' software. The process is entirely manual and linear, and does not provide much assistance for unfamiliar users (see section 3.3).

2.6 Reddy method

The Reddy method for calibration (T. A. Reddy et al., 2007a), while not the only method that uses automation, serves as a good grounding point to understand the differences between manual and automated calibration techniques. The method was developed as part of the ASHRAE funded Research Project RP10-51. It serves as one of the most cited and influential methods for calibration. The Reddy model uses an entirely different approach than the Cory model, though it still sets out to achieve the same basic goals. The premise for the Reddy method is that;

"A satisfactory overall calibration to the utility billing data will not guarantee accurate identification of the individual parameters in the simulation program. Our calibration methodology explicitly recognizes the fact that it is unlikely that any one calibrated solution can be deemed the "best" solution. Hence, our approach advocates that it is much more robust to identify a small set of most plausible solutions instead"

- (T. A. Reddy et al., 2007a).

The reason for this, as described by Coakley (2014), is that there is likely to be too much uncertainty as to which parameters are the most influential on the model regardless of the level of calibration achieved. As such, the Reddy model does not to create a single calibrated model, but instead attempts to understand what parameters are the most influential, by how much, and give a level of uncertainty in the results. It is heavily dependent on statistical analysis and probability functions.

The Reddy model is radically different approach to the Cory model as it is focussed on the importance of uncertainty in the model, and thus uses a large number of models to 'search' through a sensitivity analysis for the most plausible answers rather than iteratively improving a single model.

The analogy of fishing could be used to describe the differences in approach. If the Cory method is analogous to a single fisherman, then the Reddy method is more akin to a large fishing boat. The single fisherman searches in a small area with a tight focus gradually refining his cast until he finds a single prize fish, while the bigger boat trawls' the ocean and casts a huge net and pulls as many 'close to prize' fish as possible.

2.6.1 Reddy Calibration Procedure

The method uses an optimisation algorithm to 'search' for the most influential inputs in the model, and then refine them until the models are in agreement within the calibration tolerances. A simplified account of the process is presented below:

2.6.1.1 Stage one

The first stage is identical to Cory's as it requires the modeller to gather as much useful information about the subject building as possible, and to model the building as closely as possible to the real thing. Though this does not exclude the use of template models, it is generally intended that a detailed model is used.

2.6.1.2 Stage two

Stage two requires a reduction in the 'dimensionality' of the 'parameter space' (number of parameters that are uncertain) through the use of a walk through energy audit to identify the set of influential parameters such as building loads and schedules of use. Dimensionality here refers to the breadth of parameters that could be considered influential to the end results. Identifying these at an early stage can help narrow down this list. These are effectively the same requirements as suggested in the first stage of Cory's model.

2.6.1.3 Stage three

Next a broad parametric analysis is performed through the use of a Monte Carlo Sensitivity analysis based on a large batch of models. This then helps guide the process to find the best set of most plausible models. A set of 'strongly influential' inputs are found at the end of this stage.

2.6.1.4 Stage four

Next a more guided sensitivity analysis is performed by starting with the top "promising" solutions for the input parameters from stage three and systematically fine-tune the input values so as to improve the simulation prediction accuracy. What differentiates this step is that it starts with the estimates of the strong input parameters identified in the previous step and incrementally improves the calibration to reach a closer match with the actual data.

2.6.1.5 Stage Five

Identify the sets of most plausible solutions. These are the final models that are considered the 'most plausible calibrated models'.

2.6.1.6 Stage Six

Compute the level of uncertainty in these models using goodness-of-fit criteria

2.6.2 Advantages of the Reddy method

The major selling point of this process is that it gives does not purport to provide one definitive solution, but rather a range of most plausible solutions. Reddy describes a single calibrated model (where only one solution is proposed) as being erroneous as there is no way to really understanding that the inputs that went into it are indeed correct. Reddy's method therefore takes a 'safer' route of proposing a small set of plausible solutions, thereby giving a more robust answer.

2.6.3 Disadvantages of the Reddy method

There are two major drawbacks to the Reddy, and indeed any, automated process. The first factor is the level of resource use. The use of large batch simulations and computational algorithms which search for the best fit take a very long time. In the method presented by Reddy (T. A. Reddy, Maor, & Panjapornpon, 2007b), some case-study models took over 10,000 simulations to find the best solutions. Even with the use of cloud computing it can take days or even weeks to run several thousand iterations. This is often too long to wait, however as processing and bandwidth speeds increase, simulation through cloud computing and server farms may make it possible to get very quick results. The speed of simulation is ever improving, however the ability to manage and use a set of 'most plausible' models is less likely to be attractive to a practitioner.

The other major disadvantage of this approach is that there is little to no control over it. While it is possible to provide boundaries for the sensitivity analysis, it is possible that the actual input values fall outside of the ranges set, or that not enough parameters have been chosen. This is the infamous 'black box' problem (Coakley et al., 2014), where the user is dis-engaged from the process and does not necessarily understand how the optimisation is working or the algorithms that govern it. It may also provide little room for real-life scenarios such as the building being partially unoccupied for a year

2.6.4 Apidae Labs calibrator

The Apidae tool¹ works through a similar, though much faster, process as the Reddy method. It begins in the same way, with an initial model. The model is uploaded to the Apidae website along with utility data (the measured data to calibrate against). Uncertain parameters are selected by the user with upper and lower limits applied to the parameter value along with an intermediate step value e.g. fan efficiency values may be set between 50% and 90%, increasing in four 10% increments. Once all of the uncertain parameters are set into the tool, the process of simulation begins.

Where the Apidae tool departs from the Reddy method is in the way the sensitivity analysis is run. In the Reddy method, a Monte Carlo sensitivity test is run multiple times to identify 'strongly influential parameters'. The Apidae tool instead uses an optimisation algorithm to search for the most likely values. The Apidae algorithms are far more efficient than the Monte Carlo method resulting in a shorter run-time. For example, the Apidae Labs tool was able to run through a series of 17 different parameters, with 3 steps each, or 3^{17} (129,140,163) different possible models. The Reddy method would have to compute each solution iteratively, whereas the Apidae tools' algorithm discarded unlikely pathways along the way thus reducing the amount of time required to find the best solutions. This took the total simulation down to just 900 different models.

Apidae results are displayed using a parallel coordinate graph shown in Figure 9. A parallel co-ordinate chart allows the user to see the results of the optimisation by plotting the parameter values of each calibrated model. Each line through the graph represents a single calibrated model, thus it presents a set of most plausible solutions rather than a single calibrated model.

¹ <https://apidaelabs.com/>

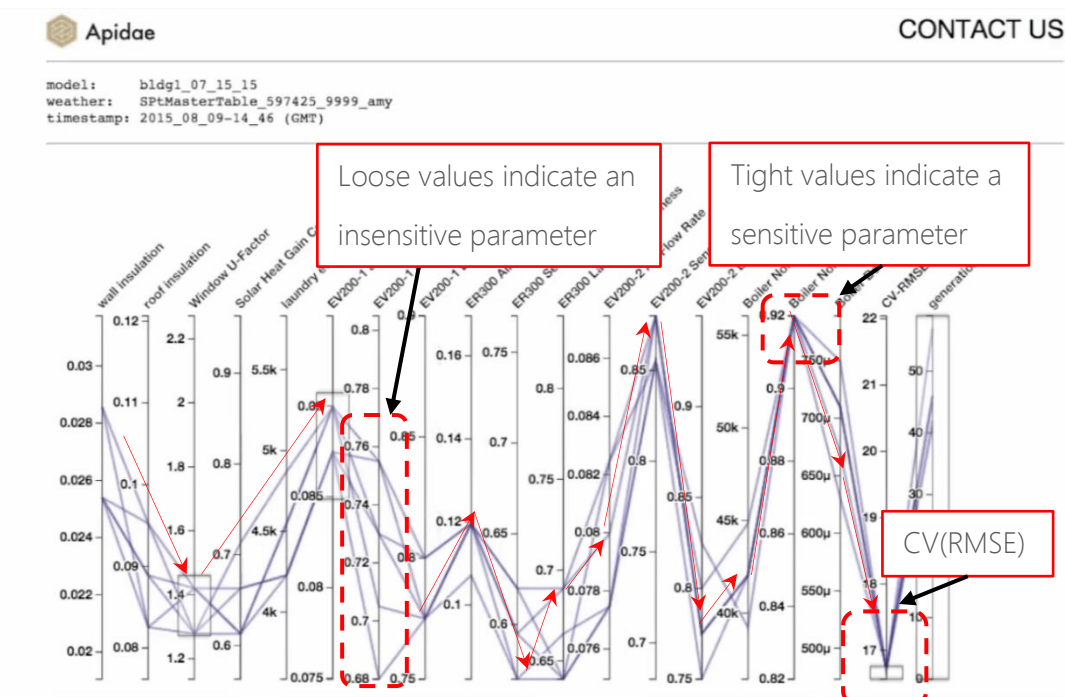


Figure 9 Apidae calibration tool results display and plot description

2.6.5 Differences in approach

The fundamental difference between the Cory method, and the Reddy and Apidae methods, is the amount of control the user has.

The Cory method requires the user to work through the models in a manual and linear process providing a single solution, with the user in control of each iterative step. The Reddy and Apidae methods use an automated process working in a non-linear way, providing a range of solutions with the user only able to select from the calibrated solutions. The sensitivity analysis or method that runs the calibration steps are automated to speed up the process.

2.7 Review of BEM Industry guidelines

Given the diversity of methods available in literature (see Table 3) it is interesting to find that few, if any, methods are presented to practitioners through BEM guidelines and protocols. A review of four simulation guidelines or protocols show that calibration is rarely described in any systematic way. These guidelines or protocols were selected because they present a general description of good practice for energy modelling and are publicly available.

2.7.1.1 NABERS Energy Guide to Building Energy Estimation 2011

Purpose: NABERS Commitment agreement guideline for energy simulation. The guide provides a basic checklist and protocol for NABERS assessors undertaking commitment agreements (a commitment to attaining four or more stars).

Calibration discussed: The document outlines a basic process for simulation of buildings looking to attain a 4-star or higher NABERS rating. It is largely checklist based, and provides little advice on validation or calibration. It instead offers a warning that simulators should not expect their models to reflect the performance of the actual building.

Reference: (NSW Office of Environment and Heritage, 2011)

2.7.1.2 The ABGR Validation Protocol for Computer Simulations 2005

Purpose: Predecessor to the NABERS Energy Guide to Building Energy Estimation guideline. Developed as a means for building energy assessors under the Australian Building Greenhouse Rating (ABGR) rating scheme (the predecessor to the NABERS rating scheme).

Calibration content: It describes the general issues with calibration and off-axis scenarios (operation and control failure scenarios) rather than discussing methods of calibration. It focusses heavily on common mistakes and areas to avoid rather than systematic methods to calibrate.

Reference: (Bannister, 2005)

2.7.1.3 ANSI/ASHRAE/IES Standard 90.1-2013 -- Energy Standard for Buildings Except Low-Rise Residential Buildings

Purpose: provides direction for practitioners on the modelling of new and existing buildings.

Calibration content: Like the NABERS guide, it provides a checklist of what to model, but has no advice on calibration techniques or the use of model validation beyond using software validated under the BESTEST procedure such as EnergyPlus (Henninger, Witte, & Crawley, 2003).

Reference: (ASHRAE, 2013).

2.7.1.4 ASHRAE Standard 209P Energy Simulation Aided Design for buildings except Low-rise Residential Buildings (2016, under review)

Purpose: *"To provide minimum requirements for providing energy design assistance using building energy simulation and analysis"* (ASHRAE, 2016)

Calibration content: This standard, though under review at the time of writing, is more focussed on enabling the modeller to analyse and examine the data outputs from the simulation software. It provides a very good description of calibration standards using MBE and CV(RMSE) tolerances as well as methods to calculate these figures. It also provides a number of basic techniques for the modeller to assess the closeness of match to real data such as linear regression analysis and its own definition of an 'energy signature scatter plot' (with temperature represented on the x axis and energy use represented on the y axis). It also suggests uncertainty analysis though does not provide a means or a reference to do so. These are not strictly calibration techniques as much as different ways of analysing the simulation output data. It does however, suggest that modellers provide a small narrative on the list of possible differences between real and simulated energy use, and a list of recommended steps to correct these faults.

Reference: (ASHRAE, 2016).

Another prominent guide, CIBSE AM11 Building Performance Modelling 2015 (CIBSE, 2015), was published in late 2015, however as it is a paid document for CIBSE members only, it was not able to be reviewed.

Most of these guidelines and protocols are focussed on new building simulation rather than existing. Often these guides and protocols are prescriptive, focussing on ensuring that the modeller follows a set methodology. They also tend to provide default inputs for unknown parameters such as lighting on/off schedules.

However, the rules around measurement and verification (M&V) have begun to require calibrated energy models as evidence of calculated savings (IPMVP Committee, 2001; Nexant Inc., 2008). How then can energy assessors make

significant savings as well as quality assure their models without a usable, affordable and achievable calibration process?

2.8 Summary

This chapter presents a review of the problems faced with ECM retrofit strategies in existing buildings in New Zealand. These are currently slow, spreadsheet based calculations which do not offer design flexibility. While the increased use of BEM could provide improvement, quality assurance remains a major issue. Uncertainty around the value of model inputs, access to building information and the impact of uncertain inputs all play a major role in the trustworthiness of BEMs. The literature suggests that calibration is a robust means of improving the quality of BEM in practice, however a range of different calibration techniques suggested in literature have not yet been translated into industry (Coakley, 2014; T. A. Reddy et al., 2007a). This is probably because of the time and expense need to complete the work which is simply not feasible for industry.

3 Research Method Part 1

Student workshops

The first research phase required the Cory process in its current form (a series of complex spreadsheets, energy models and calculations) to be assessed for usability, and then improved and streamlined into a more usable tool for the second research phase. The Cory method is complex and difficult to understand, therefore improvements are required if it is to be used and understood by industry experts for comment in research phase 2 (see chapter 4).

A Victoria University 400 level Building Science class, *BILD 423 Buildings and Energy*, uses the Cory signature process in its current form as part of its regular coursework. Students in this class were observed as they worked through the signature process in their assignment work to see what faults the process has and where improvements can be made. New and novice users were ideal candidates for this as they would highlight the major usability issues

3.1 Measuring Usability

Usability is a difficult term to interpret, even though it may appear to be simple in concept. How can it be measured? Before questions of measurement or assessment are asked, usability needs to be defined. Although there is still no complete definition, usability is generally described as “the capability to be used by humans easily and effectively” (Shackel & Richardson, 1991) and “the effectiveness, efficiency and satisfaction with which specified users can achieve particular goals in particular environments” (International Standards Authority,

1998). It is largely context dependent, with its requirements changing from task to task making measuring it extremely difficult.

In this research, usability is used as an umbrella term for understanding of the calibration process through an interface, rather than as a means of testing the interface itself. While many different means of testing usability are available (Hornbæk, 2006), few help to determine the utility of a piece of software (Johannessen & Hornbæk, 2014). Even when methods for testing utility or usefulness are available, such as the Utility Inspection Method (UIM) developed by Johannessen and Hornbæk (2014), they are not often particularly well described for novice researchers.

3.2 Class overview

The BILD423 course has two assignments during the 12-week trimester; the first requires students to conduct a type 2 energy audit on a large building with centralised HVAC as per NZS3598:2014. The second assignment is to then create both template and detailed models of the building using the collected audit data. Once the models were complete, the Cory energy signature calibration process was to be performed in order to calibrate the buildings to within the ASHRAE 14:2002 guideline. The class was observed using the process in order to document usability issues and any faults in the Cory signature calibration process.

The class, taught at distance, was provided with course documentation, lecture material, a reading list, tutorials, step-by-step guides to help complete the assignments as well as EnergyPlus template models on which to base their models. A tutor was also available to the students for one hour a week to assist them with their work and answer course related questions. Two students were enrolled in the class.

Ethics approval was obtained so that the two students could be shadowed as they completed their work during the course

The two students, who were known to the researcher, were asked to work as independently as possible during assignment two, so as not to influence each other. As the assignment took a total of eight weeks, it was thought that the

tool could be created in parallel with the course, making adjustments when and where the students gave feedback.

Unfortunately, the 2015 BILD423 class had a number of significant issues which prevented them from ever reaching the calibration process in the second assignment. Most significantly, the building they were provided with was not able to be calibrated using the Cory signature process method. There were two reasons for this. The first is that the building being used, a student hostel, had only a gas fuelled heating system. Gas metering for the building was monthly and so could not be used as it could not be synchronised with outdoor dry-bulb temperature on an hourly basis. In addition to this, temperature and fresh-air ventilation control was set by the occupants using thermostats and openable windows (or both in tandem), which meant a lack of the required central control. While it is possible to test non-central-HVAC buildings using the Cory calibration process, it still requires a means of accounting for energy use as a function of outdoor temperature, which this building was unable to provide.

An alternative building, provided by the researcher and having all the necessary documentation, was suggested so the students could complete the second assignment. This building also had both calibrated template and detailed models, allowing analysis to be performed on how the student's models matched the existing models. However, this was also rejected on the grounds that the students would not be able to understand the building in such a short period of time.

As such progress in the class was exceedingly slow, which was further compounded by a course official mistakenly not making the step-by-step guides available to the students. Despite these setbacks the students were shadowed each week, though it became increasingly obvious that they would not complete their work in time. The end result was that this research phase was postponed until the end of the academic year.

It must be noted that the class was ultimately successful. It was the first time being taught at distance and the students both received excellent grades. The description above serves to illustrate how the class did not meet the requirements of this research only.

3.2.1 An alternative proposal

In lieu of monitoring the class, an alternative was created by inviting the two students to participate in a 7-day workshop. The intent was to replicate the BILD423 class tasks over 60 working hours using a new building with students still completing the same assignment work. The workshop was run in the same manner as the class was intended to be run, albeit much shorter with the two students working full-time. The first three days were spent completing the Type 2 audit on a 14,000m² building in downtown Wellington, with the final four days spent creating and calibrating two energy models of the building, one template and one detailed. Figure 10 provides a Gantt chart for this work. The researcher supervised the work and provided additional help during the audit process to speed the work up.

Workshop Time Planner

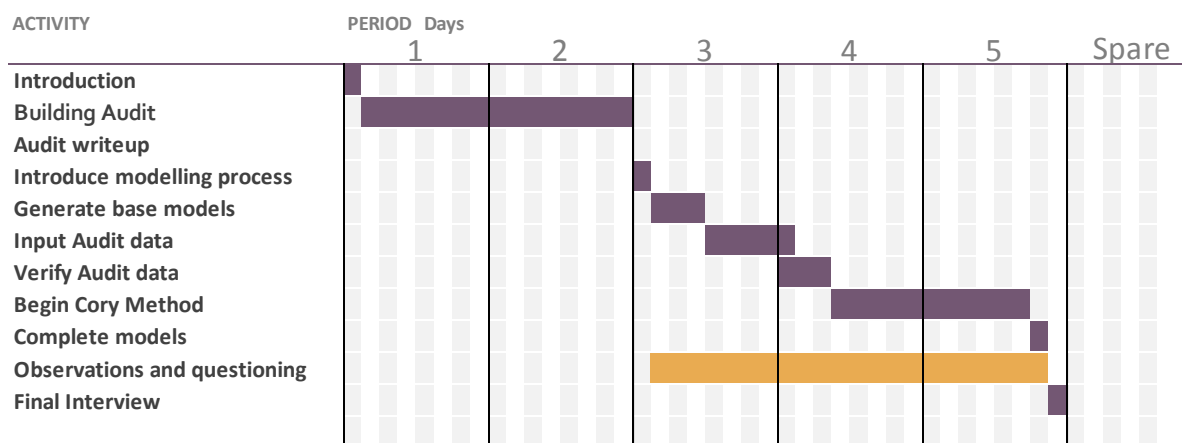


Figure 10 Student Workshop Timeline

This new workshop was successful with the students completing the audit within the first 3 ½ days, and the next stage completed to the point where the students were modelling and using the calibration process in full, though without producing valid models. This lack of a calibrated model was not an issue as the goal of the research phase was to investigate the usability of the calibration process, and to provide direction for its improvement.

3.2.2 Research method

The workshop used a number of approaches to assessing the usability of the process. In-workshop probing questions were asked of the students when they encountered difficulty, timers were kept of their actions and amount of time

required to perform an action, the number of mistakes were noted, and finally at the end the students completed a standardised usability questionnaire. At the end of the week long workshop an informal interview was used to get an overview of what went well, what went wrong, and what could be improved.

3.2.2.1 Probing questions

Two probing questions were asked of the students when using the Cory process.

1. *At your current stage, what aspect of the process do you find difficult to use or understand the Cory Process and why is this so?*
2. *In what way(s) do you think this process could be made easier or more user friendly?*

The purpose of the questions was to determine which parts of the process were difficult, and to provide a broad direction for improving and streamlining the process into a usable tool. The questions were asked in simple language so that the students would have a clear understanding of what was being asked.

A log for each student was kept taking note of the time it took to complete a step using the process. This was kept as an indicative measure of performance as the step-by-step method provided to the students was at times difficult or arduous to follow. An example of the log is shown in Figure 11. Times were kept to compare the time differences between template and detailed models.

Step	Observations Made	Start	Finish	Time taken
Tut 6 steps 9 & 10	<p>Making notes to make sure he gets the folder names correct for the calibration process. This is possibly something that could be automated – decreasing the amount of time the modeller has to go through the more tedious work.</p> <p>This process also includes changing the IDF outputs to hourly so that the calibration process can recognise the temperature differences between the weather file and the model.</p> <p>This process seems to take quite a while, as he is having to write down the new folder names.</p> <p>Following the process diligently with no major emphasis on the amount of time it is taking.</p>	10:40	10:50	10

Figure 11 Student workshop observation diary excerpt

3.2.2.2 Usability questionnaire

The System Usability Scale is a ten question survey using a five-point scale developed by John Brooke (1996). The SUS is a “quick and dirty” isolated single-use test to determine the single usability and satisfaction score of a product or service (Bangor, Kortum, & Miller, 2008). It does not measure effectiveness or efficiency as part of ISO9241 (International Standards Authority, 1998). The scale measures overall system usability with the following statements:

1. *I think that I would like to use this system frequently.*
2. *I found the system unnecessarily complex.*
3. *I thought the system was easy to use.*
4. *I think that I would need the support of a technical person to be able to use this system.*
5. *I found the various functions in this system were well integrated.*
6. *I thought there was too much inconsistency in this system.*
7. *I would imagine that most people would learn to use this system very quickly.*
8. *I found the system very cumbersome to use.*
9. *I felt very confident using the system.*
10. *I needed to learn a lot of things before I could get going with this system.*

Scores are reported from 0 to 100, though often reported as such, they are not percentage scores. A higher score indicates better usability. The SUS has reportedly been administered in articles and publications of varying degrees over 1300 times (Assistant Secretary for Public Affairs, 2013). It is noted as being a good choice for practitioners as it is independent of the interface being tested (e.g. Graphical User Interface or interactive voice response system used for telephone directories), for its quickness and ease of delivery, and for providing a single score metric which can be easily understood by a wide range of people. An empirical study of over 200 SUS surveys by Bangor et al. (Bangor et al., 2008) noted a number of interesting features of the SUS in practice.

Firstly, the scale tends to be scored positively, with a mean of 70.14 and a median of 75 (std.dev=21.71). Therefore, in order to score well on the scale a score of over 50 will be seen as acceptable (Figure 14).

Participants tend to score positively toward the higher end of the scale overall, despite statements 2, 4, 6, 8 and 10 being negatively focussed. Figure 12 and Figure 13 illustrate scoring patterns in over 200 SUS surveys. Individual scoring is generally above 50 (Figure 12) with overall survey score tending to cluster a standard deviation either side of the mean (Figure 13).

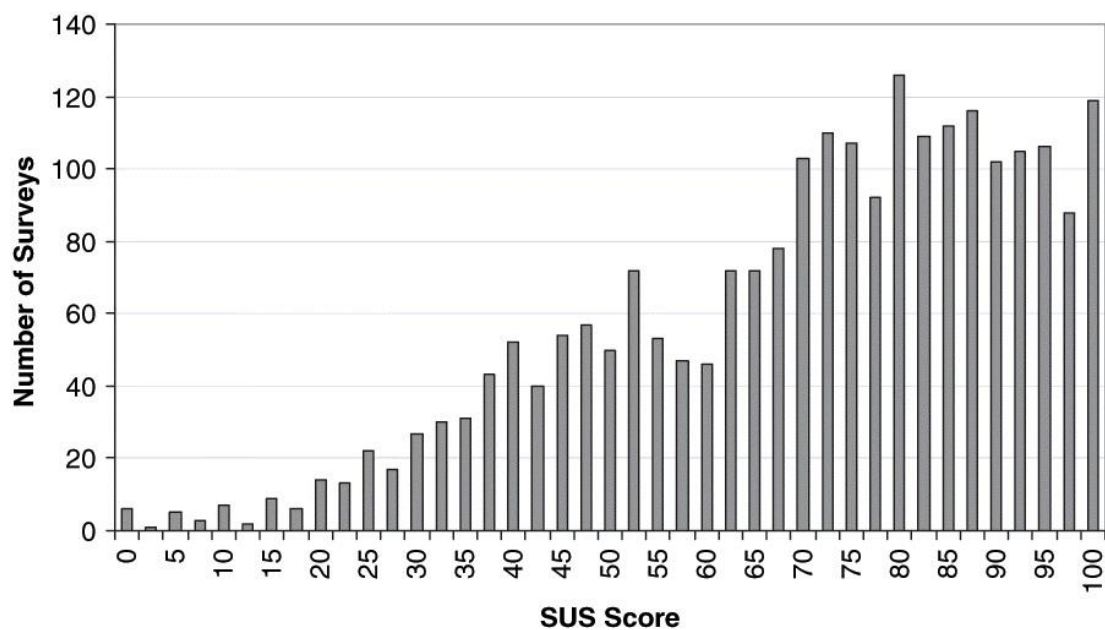


Figure 12 Histogram of individual SUS scores (n=2324) (Bangor et al., 2008)

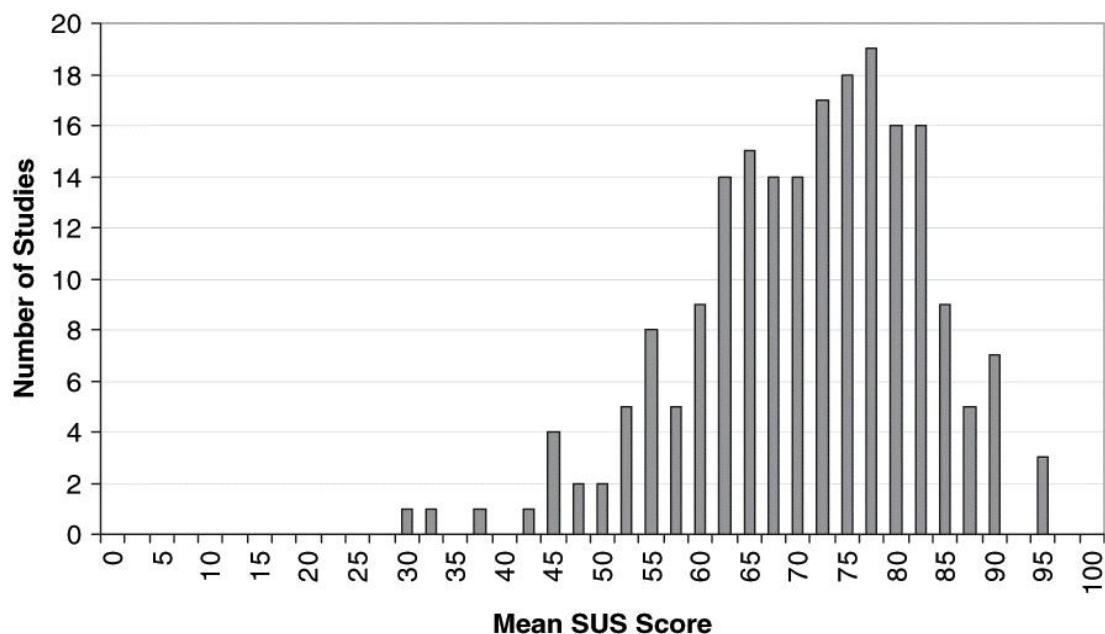


Figure 13 Histogram of study means of SUS scores (n=206) (Bangor et al., 2008)

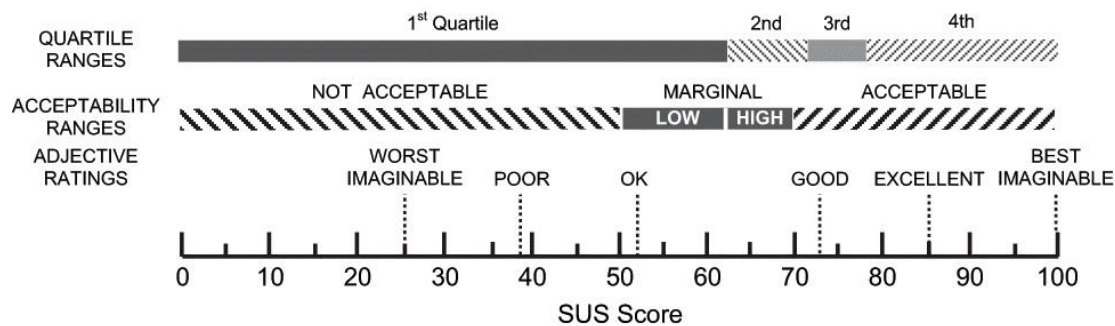


Figure 14 A comparison of mean SUS scores by quartile (Bangor et al., 2008)

3.2.2.3 Final interview

The final interview consisted of a 30-minute conversation about the workshop as a whole and how the process could be streamlined to be faster and more usable. This was kept informal and was used to provide extra context to the observations made of the students. This was an opportunity for the students to express their thoughts on the process more fully. Interaction with the students was kept informal, as the conversation was intended to provide additional insight into the observations made during the workshop.

3.3 Results

The workshop revealed a number of significant issues with the way the calibration process is presented. Neither student was particularly successful in using the process, despite the resources and step-by-step guides provided to them. Though it must be said this was not for a lack of trying. Both students were persistent and had an excellent work ethic. Largely it was found that the process of calibrating was unintuitive and was not particularly well refined.

Based on the observations, three significant things needed to change; the formatting of the spreadsheets, the number of time consuming and trivial steps needed to be reduced, as did the dependence on the user to understand Microsoft Excel to complete the work.

For instance, a large amount of copy/pasting date specific data between spreadsheets was required. While a trivial action, any miss-step would immediately throw the entire process out and cause the results to be nonsensical. This required the students to go back through the individual cells in the spreadsheets to track down what had gone wrong and search for a

possible solution. Consequently, the students focus was directed to the minor details within the spreadsheets rather than the calibration process itself.

What was most difficult about the process?

- Researcher

"While there were difficult aspects to the Excel interface, the majority came through the presentation. Being asked to use multiple sheets at any one time was very difficult to manage. You have to be fully engaged with the process the whole time. Additionally, you have to build slowly into the work; it's not something where you can pick it up quickly the next day"

- Student A

"I would agree with that, but also on some smaller points. For example, a lot of dates have to be managed, with distinctions made for business and non-business days. There is no 'sensible' distinction between weekends, weekdays and holidays which makes things confusing. When a number of these elements crop up, the whole process becomes mentally very draining. Automated processes that can identify these days would be much easier"

- Student B

These answers and following statements were centered on the way the calibration process was presented and formatted rather than which parts were difficult. Student A described the process as being a "very tight, tightrope" from which you could easily fall off. In this sense it became clear that the formatting and presentation of the process was having a major effect on the understanding of the process. The step-by-step instructions were followed exactly because they were unable to understand what the purpose of each step. Any deviation from this was met with concern. Instead it felt more like a paint-by-number exercise where the goal was to finish the tutorial, rather than what the tutorial was asking for. The end result of this is that the students were diligent, but not free-thinking or critical throughout the process.

"It seemed to be a very tunnel-visioned process, constantly I would be thinking 'this looks wrong, but I don't know why!'"

- Student B

3.3.1.1 Time and usability

A breakdown of the collective time spent by the students on various task is shown in Figure 15. The majority of time spent was in correcting mistakes handling the data i.e. moving/copying/shaping between spreadsheets. These two factors led are primarily responsible for the amount of frustration experienced by the two students and is reflected in their System Usability Scale scores. Both students scored the process as being very un-usable at 12.5 and 20 out of a possible 100 marking it in the 'not acceptable' category (Bangor et al., 2008).

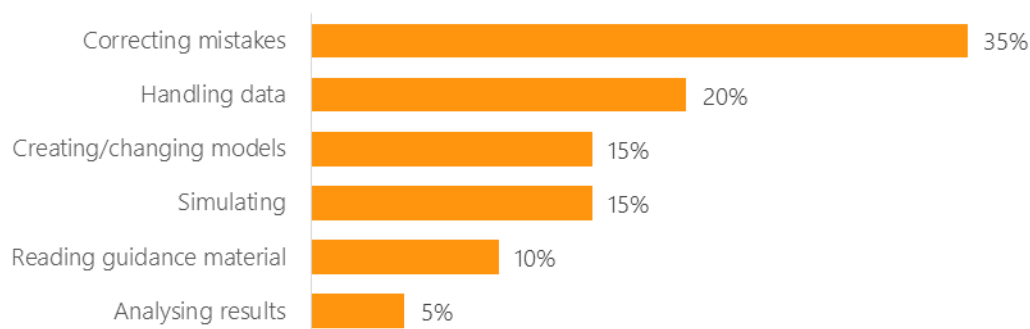


Figure 15 Student Workshops: Distribution of Time

3.4 Improvements

Following the workshop, the Cory method was revised and streamlined into a two-dashboard display (Figure 20 & Figure 22).

BEM specific improvements were made to the Cory method based on findings from Berkeley et al. (Berkeley, Haves, & Kolderup, 2015) and the results of the student workshops. General usability principles from Nielsen (1994) were used to guide more general usability issues.

3.4.1 Student workshops

Areas for improvement were noted during the student workshop with changes tested iteratively with the two student participants over the following weeks.

Three main areas of development were needed; automation, visualisation and general improvements.

3.4.1.1 Automation

Issue	High level of data management
Description	A significant portion of time (20%) was spent handling and managing data such as model inputs and model outputs between spreadsheets. This often led to mistakes being made which in-turn resulted in simulations acting in unexpected ways. The overall effect was work often being done twice.
Improvement	<ul style="list-style-type: none"> - As many processes as could be automated were. This was done by scripting macros and using pivot tables and other in-built Excel devices. Elements that were automated include: - Creation and altering of EnergyPlus input (IDF) files from excel rather than through the EnergyPlus software. - Automatic import, handling and visualisation of EnergyPlus result files through the use of scripting. - Introduction of error reporting on automated processes to ensure all automatic processes worked as expected. - Simulation time was reduced by increasing the number of concurrent simulations from one to eight through batch file creation. - 'Double backing' i.e. going back to correct mistakes, through the process reduced by limiting user opportunity for errors. - Folder and file searching within excel introduced to prevent manual searching and to ensure data from all batch files were introduced simultaneously and without error.

The main improvement increased automation of important yet mundane repeated tasks. Often it was found that students made errors when handling data and managing files. The intent was to decrease time spent as well as to ensure a more consistent result.

3.4.1.2 Visualisation

The visualisation of results and clarity of spreadsheets was the second largest issue with the students finding it difficult to manage the large number of similar spreadsheets required under the original process.

Issue	Visual clarities and inconsistencies
Description	Often inconsistencies were found in colouring and graphs across different spreadsheets. Additionally, a large number of spreadsheets were required making the process confusing.
Improvement	<ul style="list-style-type: none"> - Consolidation of spreadsheets to improve visual clarity. A single spreadsheet was used with all elements relating to the energy signatures and calibration metrics spread between two sheets. This was the main dashboard display. - These two sheets filled a single screen so no scrolling or other movement was required. This helped 'anchor' the user to a single dashboard display with all data, models and graphs being edited and viewed within. The book 'Information Dashboard Design' (Few, 2006) by noted information visualisation specialist Stephen Few was used to guide this process. - Visual clarity and consistency were increased by standardising the two dashboards. A particular style using consistent fonts, colours, scales and nomenclature were used to aid this process.

Other general issues relating to the small annoyances and faults were fixed through iterative testing.

3.4.1.3 Demonstration tool for use in industry user testing

The tool was then developed as a demonstration tool, which could only run a template model (see Figure 16) developed as part of the BEES research project (Cory, Gates, & Donn, 2011). The tool is to function as proof of concept and to help aid the understanding of the Cory calibration process in the industry workshops. A range of eight template model input parameters were selected for tests; three measuring outdoor air rate, three relating to the fans, one to cooling setpoint temperature, and one to chillier Coefficient of Performance (COP) efficiency.

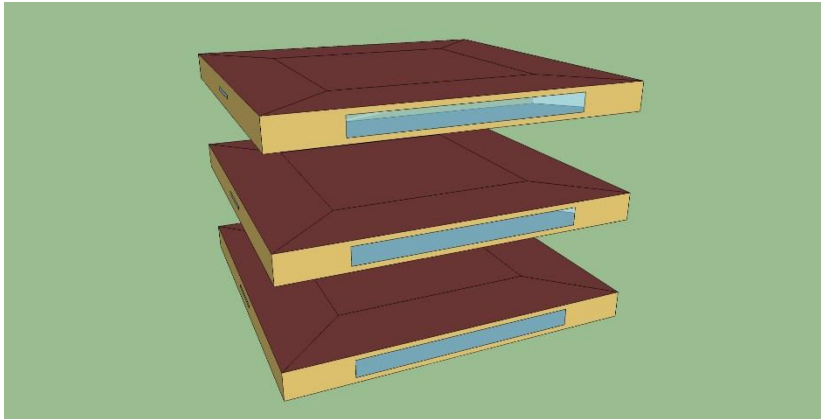


Figure 16 BEES open plan template model

3.4.2 Principles of interface design

A number of design considerations for graphical user interfaces for use in BEM are suggested by (Berkeley et al., 2015). Berkeley et al. describe the requirements for user interface design specific to BEM as part of their research on modeller assumptions on BEM accuracy (see section 2.3.3.1). They provide five requirements for successful user interface design to limit modeller uncertainty.

Firstly, inputs are to be simplified wherever possible. The original process did not provide a means of inputting information. This has been remedied with the introduction of simple type and button-click operations for all inputs. E.g. input information can be input into the dashboard display and updated with a click of an on-screen button.



Figure 17 Dashboard inputs and button

Secondly, the range of possible inputs is to be made clear to the user. I.e. provide a clear range of appropriate inputs for the parameter being assessed. For instance, providing a likely input efficiency rating of 60% for fans and a cooling setpoint range of 21-25°C providing warnings for inputs beyond typical limits would also help guide the user to what might be a likely input. This was intended to be implemented but was mistakenly left out of the finalised tool.

Thirdly, communicate functions clearly to the user. Notification windows and other signposting help to achieve this. This principle was not as important as

the Cory process only had a small number of functions, inputting data, saving and creating new models and importing results. These were all made possible through clear sign-posting and notifications.



Figure 18 Dashboard signposting of process steps and actionable buttons

Fourthly, make the entry of geometry more flexible – not an issue in this instance as no geometry is being modelled.

Lastly, clearly demonstrate what the users have entered by providing feedback. Feedback was provided at the end of tasks with message boxes popping up indicating success or failure.

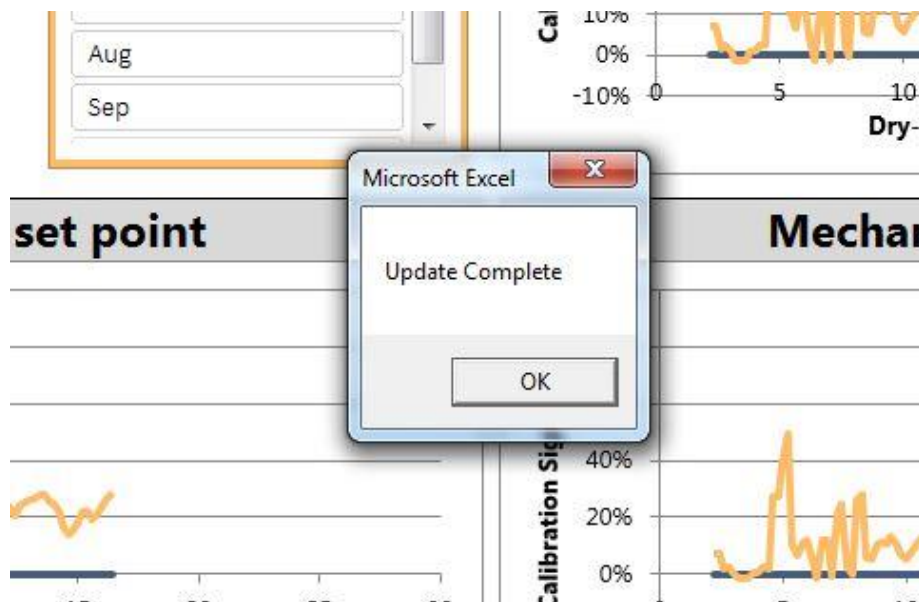


Figure 19 Dashboard notification display on task completion

3.4.2.1 Completed tool

The completed tool interface (see Figure 20 and Figure 22) used just two dashboards for the user on one spreadsheet as opposed to the previously used

six. The first dashboard provides all of the functionality of the tool, with import/export and updating functions all scripted into the buttons in the upper-left corner of the display. The user is able to control, manipulate and simulate up to 10 models at a time through this dashboard, as well as looking at the signature graphs used in the assessment.

The second dashboard shows the hybrid models calibration levels (the hybrid model is the working model where all updated inputs are applied).

3.5 Completed tool

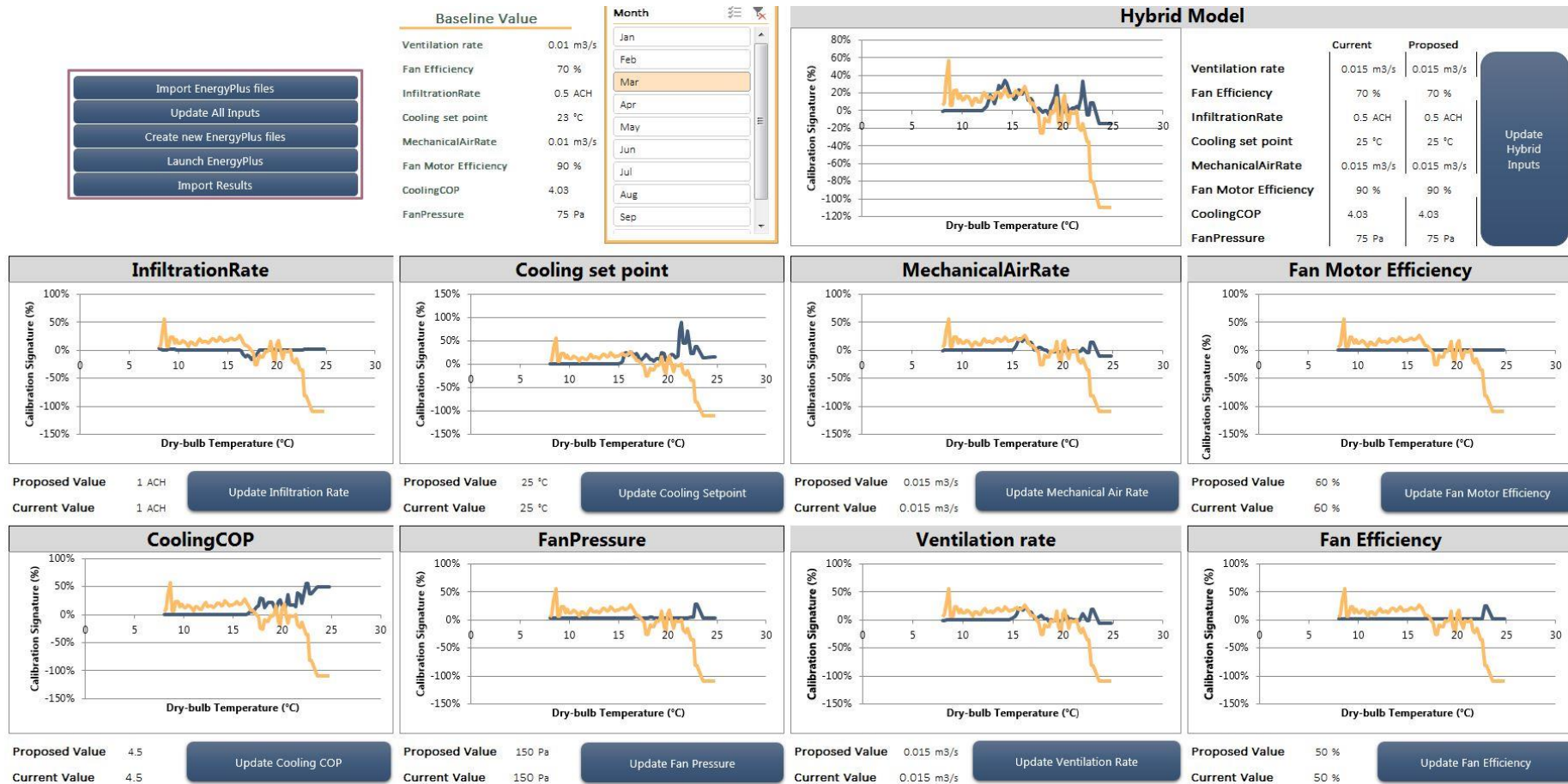


Figure 20 Streamlined Cory process tool – Dashboard one

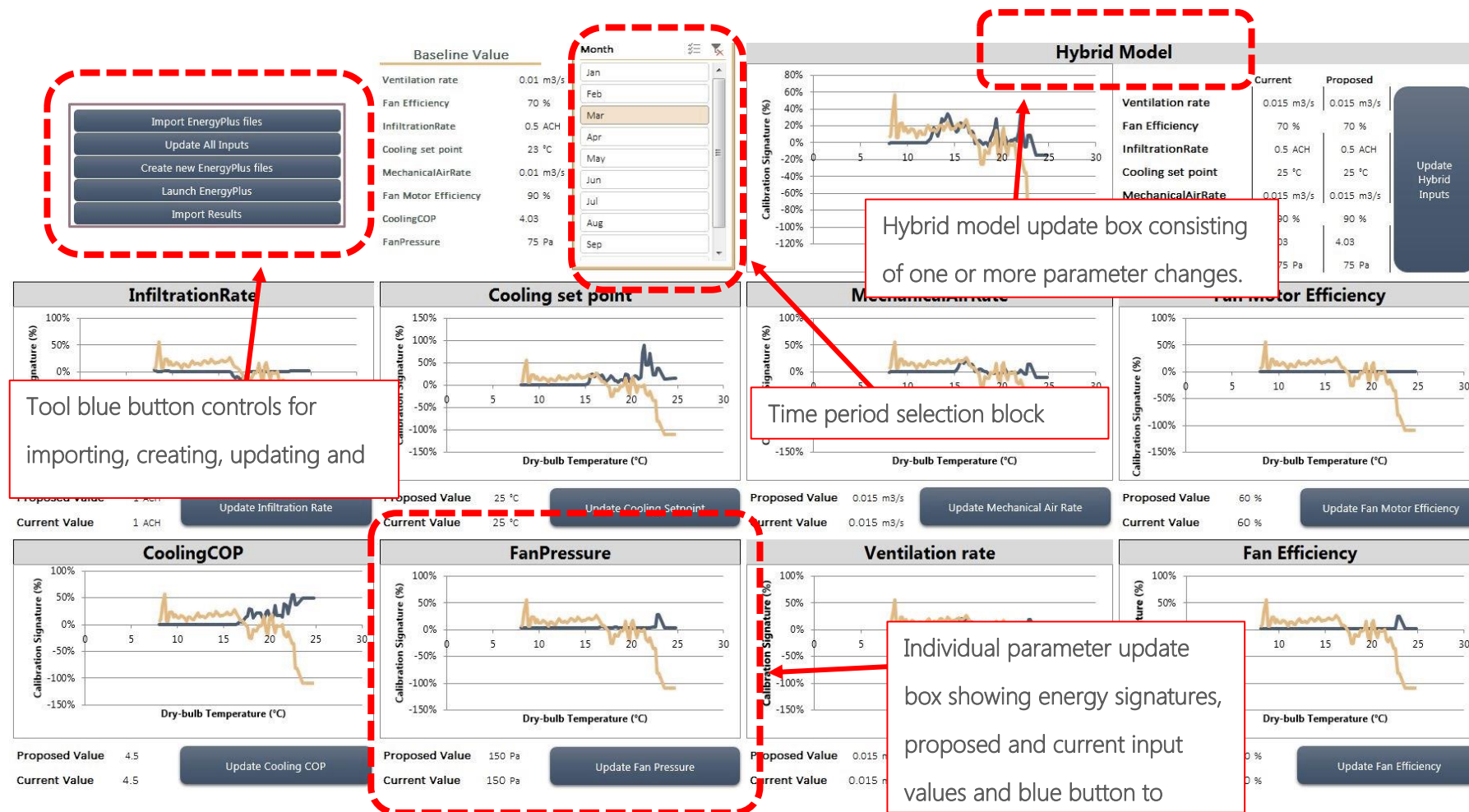


Figure 21 Dashboard one – Interface features

Monthly Totals	Actual	Baseline	Hybrid Model	Calibrated month?
Jan	148,288	136,420	150,301	✓ 1.3%
Feb	143,185	123,904	134,284	✗ -6.6%
Mar	152,332	123,310	137,060	✗ -11.1%
Apr	135,487	123,411	140,389	✓ 3.5%
May	141,624	123,120	135,179	✓ -4.8%
Jun	136,469	113,226	123,240	✗ -10.7%
Jul	148,045	123,757	127,263	✗ -16.3%
Aug	143,496	119,430	120,303	✗ -19.3%
Sep	140,435	119,419	128,589	✗ -9.2%
Oct	129,844	120,953	135,958	✓ 4.5%
Nov	123,951	120,045	132,795	✗ 6.7%
Dec	113,728	127,650	142,564	✗ 20.2%
TOTAL MONTHS CALIBRATED MBE				4
TOTAL CV(RMSE)				11%
TOTAL MBE				-3%

Item	Actual (kWh)	Hybrid Model (kWh)	Difference (%)	Difference (kWh)
Lighting	413,093	428,165	4%	15,071.52
Equipment	724,736	736,745	2%	12,008.63
Servers	163,724	164,839	1%	1,114.62
Lifts	35,478	35,125	-1%	(352.98)
Kitchenette	24,496	24,258	-1%	(238.41)
HVAC	235,209	216,964	-8%	(18,244.60)
Total	1,656,885	1,607,925	-3%	(48,959.51)

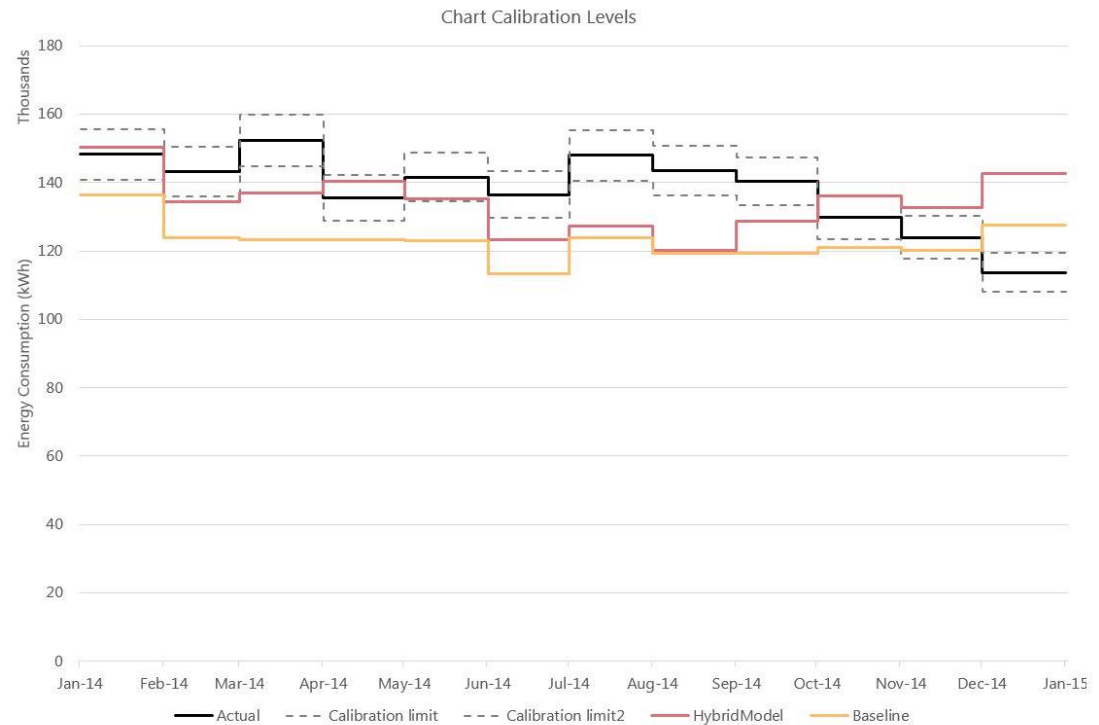


Figure 22 Streamlined Cory process tool - Dashboard two

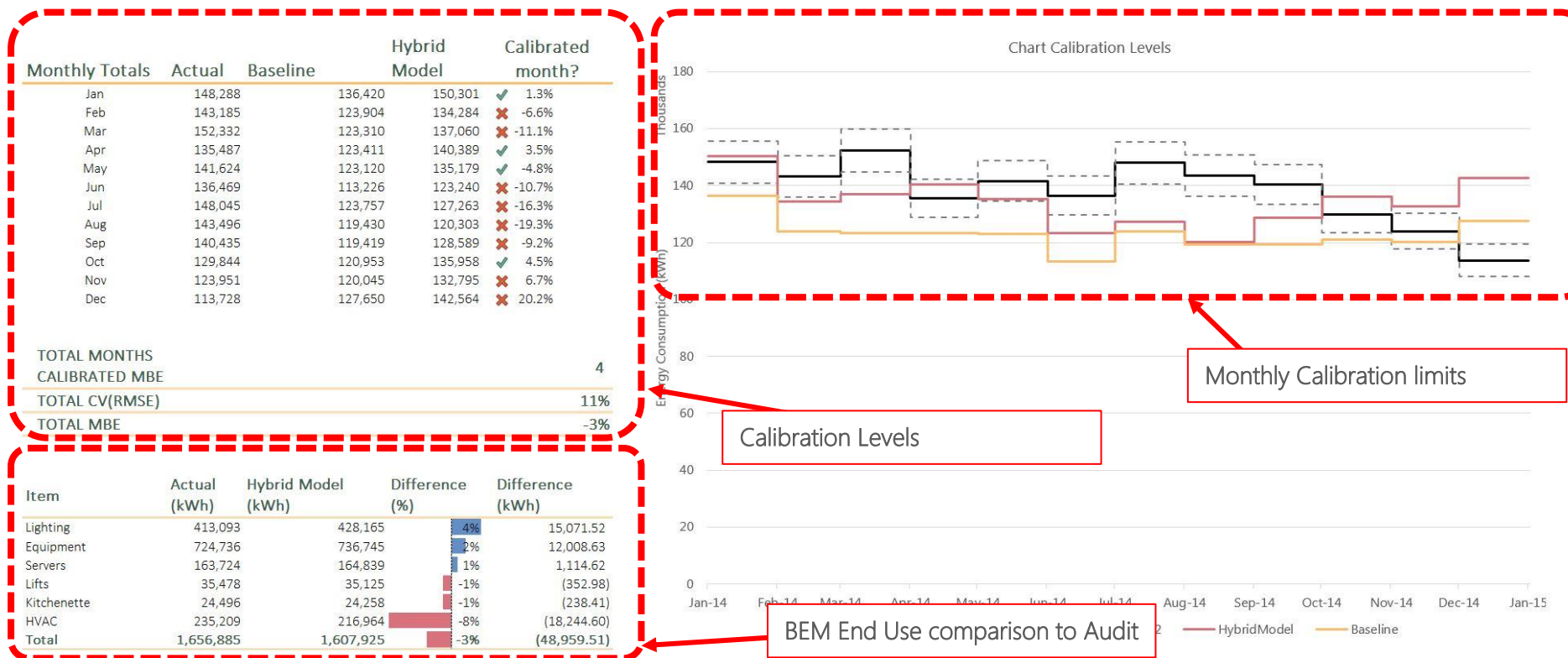


Figure 23 Dashboard two – Interface features

4 Research Method Part 2

Industry Workshops

This chapter examines the research questions, outlines the use of semi-structured expert interviews and provides the rationale for using this method over other alternatives. It also addresses issue of sample size and reaching data saturation. It provides transparency by describing the interview protocol and process, and how these were analysed and synthesised using thematic analysis.

4.1.1 The use of qualitative research in building science

The use of qualitative research is not common in the so called 'hard' sciences as there is a "widely held conviction that only quantitative data are ultimately valid or of high quality" (Guba & Lincoln, 1994). This is because most research undertaken in this field is quantitative.

Despite this, there are still a number of valid critiques to the use of qualitative research. Firstly, while literature provides many sophisticated means which can be used to calibrate BEMs, the uptake appears slow. This research is examining the use and relevance of BEM calibration tools in industry as a whole, not any one individual tool. It is not looking to say which tools are best or by how much, but rather why is industry in the position it currently finds itself in, and how can calibration tools be improved to change this? While a quantitative study comparing tools could reveal which calibration technique is the best, it would not shed any light on *why* they are not being used in industry.

Secondly, critics of qualitative studies tend to be sceptical of the reliability, validity and representativeness of results. The accusation that in order to be rigorous a qualitative study must use a statistically representative sample of random cases/participants is common (Bernard & Ryan, 2009). But this misses

the point of exploratory research. Exploratory research looks to get a range of cases to help uncover themes and patterns. It does not look to find what is typical or suggest that the findings are true in all cases – that comes later. Furthermore, it is not always appropriate or even possible to get a statistically representative selection of cases, but that does not mean that the research is not valid (Bernard & Ryan, 2009).

4.2 Understanding the research questions

The first research question, *'is a lack of a fast and usable calibration process inhibiting the use of BEM in industry'*, looks to both gain an understanding of current practice and its implications. Specifically, it looks at how the use of simulation, with emphasis on quality assurance through model calibration, is currently conducted. This is a primarily deductive research question that is setting context to test the problem statement, and then confirming or disconfirming it (Bernard & Ryan, 2009). The aim of this research question is to gain some context for why the hypothesis is or is not true, and what causes might lead it to be so.

The first research question also gives context for the second research question, *would the introduction of the Cory method improve the outcomes of energy modelling in industry?*, which is primarily inductive in nature. This is to say that observations and patterns found in the collected data (about the Cory method) are used to formulate hypotheses and conclusions about the real world (Bernard & Ryan, 2009; Goddard & Melville, 2004).

The aim of the industry workshops is to:

- a) assess the suitability of the streamlined Cory method for use in industry, and
- b) formulate hypotheses about the design of calibration tools for future research.

4.3 Industry Workshops

4.3.1 Data collection method

In-depth semi-structured interviews with expert industry professionals are the main method of data collection used for this part of the research. Interviews have been used for three reasons.

Giving context: Firstly, there is no information on the degree to which industry professionals use BEM for existing building energy efficiency projects, so this must be established. The use of BEM for retrofits is quite limited in industry at the global scale (Tupper, Franconi, Chan, Hodgins, et al., 2011), which, as a corollary, suggests that its use in New Zealand with its relatively small industry is likely to also be limited. Personal correspondence estimated this population to be around 8 to 10 firms and professional individuals (personal correspondence with NZGBC October 6th 2015, and EMANZ October 7th 2015). There are only a few firms and individual professionals using BEM for existing building modelling, as discussed in section 4.4.3, so the sample population will be limited. This limits the possible data collection techniques due to the small population and need for expert participants.

Tool understanding: Secondly, the understanding of the tool itself needs to be clear and unambiguous. The student workshops found understanding the process is not simple and is not immediately intuitive. A survey is not equipped to provide the detailed description of the Cory calibration process even when accompanied by supporting documentation (Bernard, 2011). It may not be understood equally by all respondents meaning the results may be inconsistent. Being present alongside the participants as they use the tool allows any misconceptions to be quickly clarified, and also ensures that all of the participants have an equal understanding of the process thereby limiting any possible imbalance or disadvantage between participants.

Broad responses: Thirdly, as the research is exploratory and not confirmatory, the respondents need the ability to express their views in more ways than a survey or questionnaire can provide i.e. open answers, including through verbal exchanges with the interviewer. While written responses may help to mitigate this, there is no guarantee that the results can be interpreted correctly by the researcher.

Focus groups as a primary research tool were the most obvious alternative method for this research. While focus groups have some advantages over in-depth interviews, they were not used as the main data collection method following on three considerations from Wilkinson (1998);

The first consideration is the purpose of the research. Establishing how industry currently operates and what techniques they use to calibrate BEMs needs to be on an individual basis. A lot of this information is commercially sensitive and not suitable for open discussion, particularly amongst direct commercial competitors. While focus groups would provide the same data interviews would (Wilkinson's second consideration), the practicality of recruiting, and organising expert participants is very difficult (Wilkinson's third consideration). Furthermore, if one participant cannot attend the group discussion, the reliability of the collected data will be lost reduced owing to the small population.

One other qualitative research methodology that might have lent itself to this research is Grounded Theory (B. Glaser & Strauss, 1967). Grounded Theory is an inductive method where the researcher investigates the subject area without a set sense of what the findings may be. It uses constant comparisons of the collected data and 'theoretical sampling' (where new cases are added until no new concepts arise) (Bernard & Ryan, 2009). This process is systematic and has very set rules about gathering and analysing the data, but it also allows the researcher to 'follow their nose'

While the use of Grounded Theory for this research was initially attractive, it could not be used as one of the central tenets is the requirement to analyse the data to produce theory free from bias and preconception (B. Glaser & Strauss, 1967). The research questions in this research were born out of understanding of the literature and so could not satisfy this requirement.

An alternative approach would have been to establish current industry practice through a survey, and have expert professionals partake in a single focus group workshop. This would have had the advantage of canvassing a wider population to understand current practice, possibly including those who do not undertake BEM as part of their energy efficiency projects. It would also allow for a group discussion *between* experts in industry with less chance of

commercially sensitive information being brought up. However, this option was not pursued as the first two research questions are considered to be linked. Understanding the impact of the Cory method in industry and the future of calibration tool development in isolation from current practice would not provide the necessary context.

Assuming enough respondents could be surveyed to, gain a representative answer on the use of BEM in current practice, this would still not provide in-depth answers as to why certain factors are important, which is critical to this study.

4.4 Sample size requirements

‘How many interviews are enough?’ is a difficult question to answer, particularly when, as with this research, the main research questions are exploratory. In deductive methodologies sample size is dictated by the ability to statistically analyse the results so that the conclusions can be confirmed and generalised (Bernard, 2011). However, in this case the research is about a specific domain, that of industry BEM use and calibration tools, therefore it requires a closer look at the group being assessed with careful selection of its participants.

Given this, it is important to ask two questions when considering the total sample size required; firstly, are the findings based on the data collected going to be measurable and reliable? And secondly, are the results going to be generalised up to represent the wider population that is being studied? These are known as the internal and external validity problems. In this research, the question of measurement is dependent on the sample population, which in this case is limited. However, the answer to the second question – the external validity problem – is dependent on whether or not the things being studied are part of a probability sample (Bernard & Ryan, 2009).

4.4.1 Probability and non-probability sampling.

Two types of sampling exist; probability and non-probability sampling (Bernard & Ryan, 2009). In order to determine which type of sampling technique is required of the research project, Bernard and Ryan (2009) propose a three rule test. The first rule asks if the research objective requires the researcher to

estimate something about the wider population, if so, probability sampling is to be used. In this case the research objective is to explore the ways Energy Efficiency assessment and BEM calibration is conducted in industry currently, and whether the adoption of the Cory calibration process would improve outcomes. This research is not carrying out a usefulness test (though it is using elements of usability methodologies to answer the research questions). So in this instance, the objective is not to estimate a parameter size.

The second rule is to ask if there is a need to provide a statistically representative sample. If so, probability sampling is required, if not, non-probability sampling should be used. In this case the sample population from which to draw from is low (twelve) as explained in section 4.4.3. If it was possible to recruit all twelve participants, this would be sufficient for statistical analysis. The third rule simply asks the researcher to systematically apply an appropriate sampling method, such as purposive sampling, when using non-probability sampling (Bernard & Ryan, 2009).

Purposive sampling is often used when the research is field oriented and unconcerned with generalisability (Bernard & Ryan, 2009). Purposive sampling can be used for cases with small but intensive case studies with limited population samples such as is the case with this research (Bernard, 2011). In purposive sampling selection of participants is at the researcher's discretion rather than randomly, however this can bias the results. In other words – you take what you can get but are uncertain of the bias.

4.4.1.1 Sample size and data saturation

Determining bias and ensuring that the research is exhaustive is usually determined by the sample number. The sample number is highly dependent on the population available. The aim of a sample in non-probability research is to reach the point of data saturation, the point at which no new information is discovered (Guest, Bunce, & Johnson, 2006). The term 'saturation' is not a particularly well defined term for the applied science researcher. As Guest et al. point out, there are very few published guides to say when, numerically, data saturation can be considered to have been reached thereby leaving it to the judgement of the researcher.

Marshall, Cardon, Poddar, & Fontenot (2013) suggest three methods of defining a level of saturation. Firstly, by citing recommendations set by prominent methodologists in the field, secondly by defending sample size using precedent, and thirdly by justifying using statistical means.

Seidler (1974) asserted that in purposive sampling, where the participants or 'informants' are selected individually by the researcher, a minimum of five participants are needed to help limit bias. Additionally, in their analysis of 60 interviews, Guest et al. (2006) found that their first six interviews revealed 80% of the results, with the data being considered saturated (92%) after twelve interviews.

Nielsen (1994) suggests that over 80% of all usability issues can be identified with the first five participants. Though this result is usually associated with usability evaluation experts, the finding is consistent with Guest et al (2006) in that the first few participants will elicit the majority of possible responses.

The concept of data saturation is not without its critics. In their critique of the over-emphasis use of 'saturation' in qualitative research O'Reilly and Parker (2013) challenge the need for saturation and question its applicability to all situations. O'Reilly and Parker (2013) argue that the adequacy of a sample is not determined by the number of participants in it, but rather by the quality of information found and the transparency of the method. They go on to discuss the need for transparency, breadth, depth and saturation as appropriate for the research being conducted. Their publication calls for researchers to ensure that their work is transparent and unambiguous in its methodology, has depth to its interviews, breadth in its participants (where possible) and is clear about any limitations and difficulties encountered in order to satisfy the need for scientific rigour (O'Reilly & Parker, 2013). Overall, validity in probability sampling is about selecting a representative sample and exhausting research area through data saturation. This area looks to explore this area through a small number of in-depth cases. Validity in non-probability sampling is not governed by the same rules necessarily and can be achieved through a small number of in-depth cases by providing transparency to the process.

4.4.2 Sample design – defining the expert interview

In their book *Interviewing Experts* (Bogner, Littig, & Menz, 2009) provide a very good analysis of the problems faced when interviewing experts, particularly with respect to defining an expert and the role of the interviewer in the interview. At the beginning of chapter two they note that not only are expert interviews difficult to conduct, but that they are also suspected of tending toward impressionistic results, rather than standardised quantifiable results, because they tend to be led conversationally by the interviewee (the expert) with occasional interjections by the interviewer (researcher). The question of quantifying the results is clearly a difficult one if the sample population is low (as is generally the case when considering expert participation). However the question of the experts giving impressionistic responses (which may only be relevant to them and their understanding) is of more immediate importance for this research. This discussion tends toward questioning the rigour of the methodologies employed when conducting interviews with experts so that the research is not only rigorous in method, but also provides a clear and unambiguous set of results.

The first question, that of the quantification of results, is dealt with by the use of expert interviews as an *exploratory* tool seeking to orient the research and to generate hypotheses, providing a clearer understanding of the problems faced (Bogner et al., 2009). In other words, it is used to explore the subject area with a reasonable degree of openness as well as to lay foundations for future research through the generation of hypotheses and research questions. The goal is to 'sound out the subject area' thus allowing for conversational and free-flowing interviews where unexpected changes in interview direction are allowed. As identified in the literature review (section 2.8), there is a disconnect between building science academia and industry practice. Though the importance of sample size remains, it does not necessarily require statistical analysis of the data to remain rigorous, however it should be explicit in its lack of quantitative analysis. This approach is used in this research as opposed to the *systematising interview* which uses expert interviews to extract information for the purposes of gaining complete information which would be more conducive to statistical analysis.

Under the tenets of exploratory research, it is important to gain as wide a range of views as possible on the topic area. This can be most easily done by having a wide range of research participants. However, this research explores an area which tends to be particularly niche, thus limiting the sample population.

4.4.3 Sample design – identifying the population

The population used for this research is the New Zealand engineering and building science industry as they are the most likely to use BEM in their current and future work. This includes individuals who would undertake energy efficiency assessments, provide EE retrofit consulting and or NABERSNZ accredited practitioners. While this population might seem non-specific and perhaps quite broad, the reality is that there are only a limited number of practitioners with a good level of experience with BEM and simulation in practice. This assumption is based on industry networking with professionals in New Zealand, Australia and the United States whom have all stated that the use of BEM, and especially calibration of BEMs, as a means of assessing and designing ECMs to a building is not particularly common in industry. This has resulted in only a small percentage of firms or professionals that have the ability to undertake the work. By way of example, an architect would normally not engage in this line of work.

The shortlisted candidates were assessed for their appropriateness for interview based on a number of inclusion criteria. The candidate:

1. Is considered by industry bodies and other professionals to be highly proficient in the design or assessment of energy efficiency projects and building energy modelling and;
2. Could be considered by industry bodies and other professionals as being able to reflect on the wider industry and the future direction of BEM in industry in New Zealand.

The requirement to not have industry participants who were novice or new to BEM and its links to EE assessment and ECM suggestion was based on mock interviews with the student participants in phase one of the research. When the students at the end of their workshop were asked about the calibration tool, using broadly similar questions to what would eventually be used with the

industry participants, they were unable to extricate themselves from their recent experiences i.e. their responses to general questions were often simple single word answers that did not explain why they felt this way or what implications this could have. This makes this point quite contentious for two reasons. One, it is inherently exclusionary to current building scientists, engineers and building designers who, for the purposes of this research, are not considered experts in the wider field, and two; because of the difficulties in defining an expert and taking their word 'as gospel' over the word of other excluded participants.

In order to develop a list of this small population, two industry associations and a number of industry experts were contacted to discuss the research and to identify individuals who would be able to provide knowledgeable insights into the subject area. The industry associations were the New Zealand Green Building Council (NZGBC) and the Energy Management Association of New Zealand (EMANZ). The NZGBC has a large network of building modellers as part of their internal auditing requirements for Green Star projects, and EMANZ has a large member based organisation which specialises in energy management and conservation. Networking through industry and academia were also used to find potential participants. The final short list had twelve names, ten from New Zealand and two from Australia. Removing duplicates, the NZGBC listed five names, EMANZ two and networking found four additional names.

4.4.4 Sample selection and limitations

Of the twelve potential participants shortlisted, three did not respond to the invitation, four declined for logistical reasons, one declined but recommended a replacement (who was already on the shortlist), and four accepted the invitation and participated in the interviews. While this is a good participation rate (33%), it is unlikely to meet the minimum requirements needed for saturation and sampling bias as set out by precedent work such as Guest et al. (2006) or Seidler (1974).

4.5 The Workshop and Interview process

The industry workshops were used as the primary means of data collection to answer the main research questions. Figure 25 gives the workshop timeline,

showing the use of a pre-interview and a post-workshop interview. The industry participants were asked in a pre-interview to discuss their current practices as they related to EE assessments in buildings, and the use of BEM in those assessments. This part of the workshop was designed to assess similarities and differences between professional practice in New Zealand, and what information was available through literature and personal correspondence with other professionals in Australia and the United States.

Upon the completion of the pre-interview, two BEM calibration tools, a commercially available optimisation tool by Apidae Labs (section 2.6.4), and the streamlined Cory tool were presented and trialled in a short hands-on workshop. In order to assess the relative merits and limitations of the two tools, the participant needed to be able to interact with the respective tools in order to understand the differences between them, and how these differences might affect their adoption in industry.

The participant was taken through the following steps;

1. They were first introduced to the definitions of calibration used in this thesis, and was made aware of calibration issues as discovered within the literature.
2. A quick walkthrough of an exemplar BEM was shown to them. This came with accompanying audit data (the same building and information as was gathered in the student workshops).
3. And finally, the two calibration tools being compared and contrasted.

The participant was first introduced to the definitions of calibration used in this thesis, and brought up to speed on calibration issues as seen in the literature. Then a quick walkthrough of an exemplar BEM was shown to them. This came with accompanying audit data (the same building and information as was gathered in the student workshops). Finally, the two calibration tools were compared and contrasted.

The workshop lasted for approximately and 3 hours, with time spent evenly between the two tools. The goal was for the participant to use and 'play' with the tools by attempting to use, and thus calibrate, the exemplar BEM. They were asked to consider the relative merits, features and limitations of each tool.

Successful calibration of the model was not assessed as understanding the difference in 'approach' between the two tools was more important. In the case of the commercially available tool, a 5-minute introduction video made by the developers was shown to introduce the core concepts and approach to the tool. In lieu of a video, a PowerPoint presentation was used explaining the Cory method to a similar level of detail.

After a short break, a post-workshop interview was conducted exploring how the implementation of the tools, or features of the tools, might impact their current practice. This interview provided insight into how the introduction of the tools might improve current practice, and into the direction future tool design.

4.5.1.1 Lessons learnt from the student workshops

When designing the industry workshops, lessons learnt from the student workshops were used to establish how much could feasibly be put into a three-hour industry workshop.

The Cory method is split into two parts; building the initial model based on the energy audit and then calibrating it using the signature process (section 2.5.3). The energy signature analysis steps were made the main focus of the new tool and industry workshops due to being the unique aspect of the calibration tool and the hardest part in the process to understand and to iterate through. The first and second steps (creating the initial BEM and matching it against the audit data) is considered to be a necessary step in any modelling procedure and so was not included in the industry workshop. This resulted in the exemplar BEM requiring only the final step of the Cory method to be calibrated (see Figure 24).

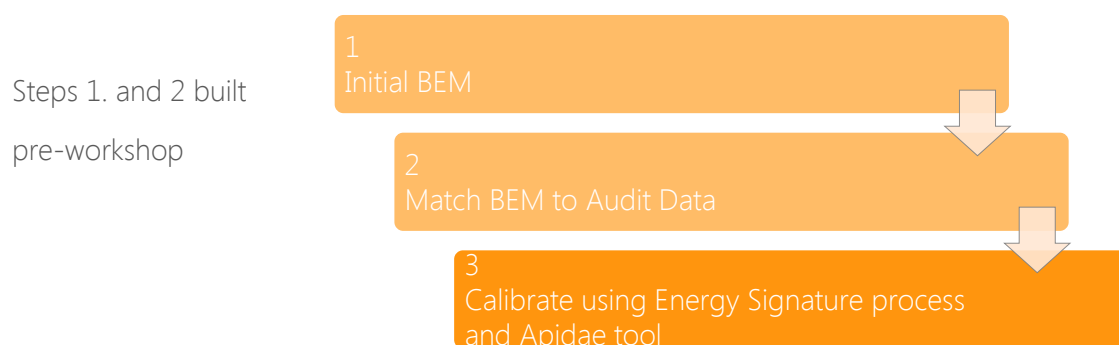


Figure 24 Exemplar BEM development

Industry Workshop Time Planner

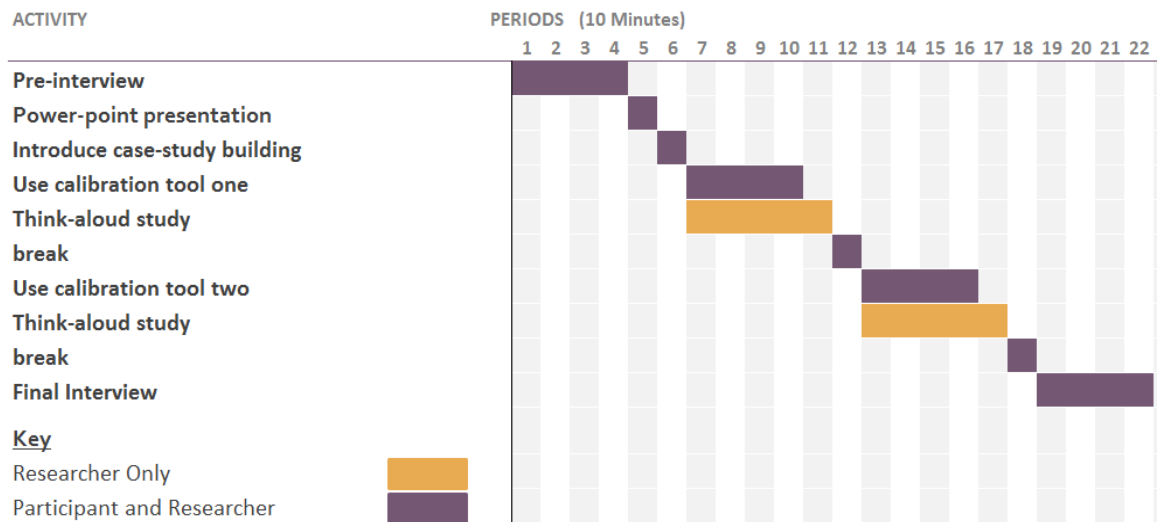


Figure 25 Industry Workshop Timeline

4.5.2 Interview Design

The interview questions were focussed on three main areas:

1. The state of current practice and the role BEM and calibration within it.
2. Impressions of the streamlined Cory tool
3. Requirements for future uptake of BEM and calibration tools

As a result of the small number of Industry professionals able to take part in the workshops, a simple semi-structured interview guide was developed. This was designed to allow flexibility within the answers, while also providing structure to the interview so that the responses across all four workshops could be compared and contrasted. Question examples from the interview guide, for both pre and post workshop interviews can be seen in Table 5. The participants did not receive a copy of the guide.

Pre-workshop question areas		Example Question
Background and practice		Please tell me a little about your background and role as it relates to commercial buildings and energy.
Current practice	Energy Efficiency Assessments	Please describe for me your standard process for calculating and recommending Energy Conservation Measures for a large commercial building
	BEM and Calibration in current practice	How do you determine whether or not a model is calibrated? How would you usually go about correcting a model that fell outside of this range?
Post-workshop question areas		Example Question
Impressions of Cory tool	Tool Impressions	What are your impressions of the tool itself?
	The role of the user	Do you feel that the modeller's bias (uncertainty) and ability to unduly tinker with the model are sufficiently limited with this process?
	The role or uses of the tool in current practice	Do you feel that uptake of the calibration process would help you to make more informed decisions when suggesting energy saving opportunities?
Requirements for uptake	Other tools for calibration	Are you aware of any other tools that look to aid modellers in creating calibrated models?
	Requirements for uptake	What are the current barriers of use against Building Energy Model calibration in Industry?

Table 5 Interview guide categories

The interview was allowed to flow naturally in a conversational manner. This allowed the experts to put their points forward and meant the discussion could focus on elements they thought were important. If questions in the interview guide were not answered in the conversation, they could be brought in when the discussion had stalled. The questions asked in the guide were quite direct, but were used to ensure all discussions were 'anchored' in the same place while still allowing them to flow freely, as discussed in Bogner et al., (2009).

The use of the interview guide was to direct conversation, but also to help link the two research questions together. This was important as it provided the participants a basis to consider their opinions.

As the first research question was largely deductive, the interview guide was closely adhered to in each interview with only the occasional deviation. Usually deviations occurred because the participant had suggested a new or otherwise

relevant piece of information that was not present on the guide. This was also the case for the second research question which is primarily inductive. While this may seem at general odds with some modes of inductive inquiry, the topic area could easily expand far beyond the scope of this research and the desire was to keep the answers to the second question relevant to the first question responses (though not so far as to limit their answers). This, coupled with the limited number of participants, meant that the interview process was kept as consistent as possible. Terminology was kept consistent with industry and academic terms. This helped set the scene for each question allowing additional probing questions when responses were unclear, limited or warranted further investigation. See section 9.1 – Appendix A for the complete interview guide.

4.5.3 Interview Interaction

The relationship between the interviewer and interviewee in expert interviews is an important one as it can dictate the course of the interview and the responses given. Bogner and Menz (2009) suggest that if the interviewer is seen as a co-expert, or an expert from another domain (e.g. academia), he or she will be treated as an equal which can lead to useful 'trading of information' that may not have been revealed otherwise. While this does not rely on complete congruence or equal knowledge between interviewer and interviewee, the general premise is that the interview can become less formal, more conversational, and thus moves away from a 'consultation' from the expert interviewee.

While this 'trading of information' and conversational tone between interviewer and interviewee can be seen as problematic (for example when the interviewer fills the space of the interview), it does allow the interviewer to present preliminary findings, thoughts and positions to which the interviewee can suitably respond. This allows the interview to validate or explore their positions more effectively (Bogner et al., 2009). This approach can be taken as the research is exploratory and not confirmatory (or as previously mentioned *systematising*). In this research, if a point was raised that was inconsistent with literature, contentious, or not well understood by the researcher, this trading of

information was used to explore with the interviewee the reasons why their position should be considered the best.

4.5.4 Workshop Interaction

As discussed in section 3.1, usability is a very difficult thing to measure reliably (Hornbæk, 2006). This research is not looking to evaluate the usability of the streamlined tool, but rather whether the calibration process has been communicated to the participants effectively. In this sense it was important that the participants were guided through the process as they 'played' and learnt by doing.

With respect to a complex and time intensive task such as energy model calibration, measures of effectiveness highlight how well the participants have understood what is required of them, and how they have prioritised their time toward the tasks set. Most usability studies require users to undertake simple tasks such as ordering tickets on a concert hosting website. In these scenarios no instructions are provided to help the user. It is assumed that the interface will provide a clear and intuitive method. Due to the complexity of the tasks required to be undertaken in this study, an intuitive interface could not be relied on, often the process had to be explained even after the original presentation. That is not the case in this research. Guidance was necessary even after the process was explained to the participants due to the complexity of the tasks. This was provided by the inclusion of a small set of instructions which broke a full 'cycle' of the Cory method down into smaller, simpler tasks. Additional guidance was provided by the researcher only as required.

4.5.4.1 First tool - Streamlined Cory workshop

The Cory method workshops were run similarly to 'think aloud' studies (Nielsen, 1994) which are often used in laboratory usability studies of various software. In a typical study the participant is usually a specialist evaluator experienced in usability research whereas in this research, the participants were experts in energy efficiency and modelling rather than usability assessments. While an important point for most usability research, it is not of major importance for this research as the goal was for the participant to raise

questions or points about process and understanding of how the tool worked, rather than to point out explicit usability issues.

The think aloud study, or cognitive walkthrough as it is sometimes known, is a process where the user undertakes a series of pre-set tasks through a piece of software and describes their experiences as they work through it making positive and negative comments throughout (Lewis, Polson, Wharton, & Rieman, 1990). This research slightly modified the 'think aloud' study slightly so that the participants/users could undertake the 'cognitive walkthrough' at the same time as using the tools rather than before or after. I.e. describing their experiences as they work through it making positive and negative comments throughout. It has also been suggested by (Berkeley et al., 2015) that think aloud studies could be of use when investigating interface and operability of BEM software.

Usually the think aloud study is used on simple interfaces. In this case the two tools being compared are not simple. They have complex graphical user interfaces, complex and very different approaches to the task of calibration and each requires a breadth of background knowledge on subject to be able to use them effectively. However, as stated, the goal was to ensure *understanding* not to assess usability. Therefore, the researcher was available to ensure each interview participant had the same understanding of process and difference in approach. Interactions were kept similar to the conversational tone of the interviews, however the expert/co-expert relationship switched as the researcher was now in the more familiar position.

At the end of the first workshop the participants filled out a System Usability Scale to describe their overall impression of the tool. One participant was unable to complete the SUS due a computer issue on the day of the assessment, although the participant was unable to complete the SUS the workshop continued as planned and every other aspect could be used included any comments made during the process

4.5.4.2 Second workshop – Apidae Online Calibrator

A second calibration tool, available online, was used as a counterpoint to the streamlined Cory method. An automated optimisation tool using a process

similar to the Reddy calibration procedure was desired (see section 2.6). A small set of criteria were made to find an optimisation tool able to be used for calibration of EnergyPlus models. These criteria are designed to help find the best comparative technique that could be used as a counterpoint to the Cory method of calibration.

1. Able to be performed with relative quickness and ease
2. Is readily accessible (not requiring re-working or formatting for use)
3. Readily learnable (within a period of an hour)
4. Provides a contrasting approach to the Cory method

The only tool that was suitable for this task was the Apidae calibrator (see section 2.6.4). The tool is fast, more easily understood than other tools (like GenOpt or JePlus, which are primarily used for research purposes), and has a short but comprehensive tutorial video available.

The Apidae tool workshop was less restrictive than the Cory method workshop. The focus was to provide a different perspective to calibration than the Cory method. The participants were asked to take note of the features of the Apidae tool (automation, optimisation, visualisation of results etc.) rather than the optimisation (algorithm) mechanics of Apidae calibrator work. Again, the tools were not being tested against each other, rather contrasted against each other from the point of view of the practitioner. This meant that the time spent assessing the Apidae calibrator was focused on its features rather than process.

In both the interviews and workshops breaks, in combination with the conversational tone, meant that the interviewees were not overly stressed or fatigued.

4.5.5 Data recording and transcription

Each interview was audio recorded and transcribed by the researcher. Each interview was transcribed verbatim so that content was not lost. Grammatical mistakes, fillers like 'err' and 'um' were removed along with other messy elements. The transcripts were tidied so as could be easily read and understood as they were being assessed for content only and did not need the detail required for research such as conversation analysis (Bernard & Ryan, 2009).

In total, nearly ten hours of recording was transcribed across the four interviews, with notes and memos made throughout making for a rich and detailed dataset.

4.5.6 Ethical considerations

Ethical considerations were highly important in this research. As participants were asked to discuss their professional work, they were assured that none of what they said would be met with favour or judgment and would be held in confidence. Participants signed a confidentiality form respecting their right to not be named, and that any naming of other professionals or other firms would also be held in confidence. The names of people, projects, buildings or other identifiable information was changed to pseudonyms.

Networking within such a small sample population was considered as a possible confidentiality risk, so participants were asked not to discuss the contents of the research or workshops with anyone else.

4.5.7 Research limitations

Such a small sample (four) is a major limitation of this research. The sample does not satisfy the generally accepted requirements for data saturation (Golafshani, 2003). Though it is exploratory and in-depth, it is unlikely to be exhaustive of the entire research area. However, that does not mean that the results are any less valid or important. It means that they cannot be considered generalizable to the wider population. The results are also tentative. However, they are guided by professional opinion and experience meaning that they *can* be used as a basis for further inquiry.

4.6 Interview Data analysis

Qualitative data analysis is a broad and sometimes vague area of research despite having a wide array of techniques and guides (Bernard & Ryan, 2009). This research is investigating qualitative data (interviews) in a qualitative way (by searching for meaning in the data without accompanying numerical evidence for those meanings). I.e. it looks to identify themes in the interviews and present ideas and arguments as to why they occur (Bernard & Ryan, 2009).

As stated earlier in section 4.4, this research uses a small number of in-depth cases making quantitative analysis of the results impractical. Interviews have long been established as a technique for gathering and collecting data about individual cases, cultures and domains, thus there is an equally long history of techniques used to analyse them. As such, refining this list down to find one analysis technique that best answers the research goal is important.

This research looks to understand *how* industry works, *why* it operates in the way it does, and *would* the adoption of the Cory method help increase the use of BEM to improve energy efficiency outcomes. It also asks what direction calibration tools need to take in order to be useful and usable. Choosing a research approach that answers these questions most effectively is important as it defines how the dataset will be viewed, coded and reported. A wide array of methods exist (Bernard & Ryan, 2009), however they tend to be very rigid in approach, and in many cases have requirements that cannot be met in this project in order to work e.g. more samples are preferred in content analysis (Bernard & Ryan, 2009).

When assessing the suitability of qualitative analysis methods, it is often not as simple as choosing the 'best fit'. Saldaña (2015) asserts that the best analysis techniques of qualitative data varies greatly between research projects. While many researchers may be inexperienced with qualitative research, they tend to have a good idea of what direction their research should take, even in the case where no single technique is totally appropriate. As each study is unique, a rigid approach may not always be suitable.

4.6.1 Thematic analysis

Thematic analysis (Boyatzis, 1998; Attride-Stirling, 2001; Braun & Clarke, 2006) allows the research questions to be answered with enough flexibility to explore the research area. The origins of thematic analysis are something of an oddity in qualitative research as they cannot be clearly pinned down to any other specific approach, however, thematic analysis has many parallels (Attride-Stirling, 2001), particularly to grounded theory (B. G. Glaser & Strauss, 1967). As stated by Attride-Stirling, "[Thematic analysis] aims to explore the understanding of an issue or the signification of an idea, rather than to reconcile conflicting definitions of a problem" which seek to "unearth the

themes salient in a text" (Attride-Stirling, 2001). For example, the network shown in Figure 26 show a hierarchical network of themes used to find the central ideas hidden in the data. The network works from the bottom up, gradually joining and refining themes until the central idea is found.

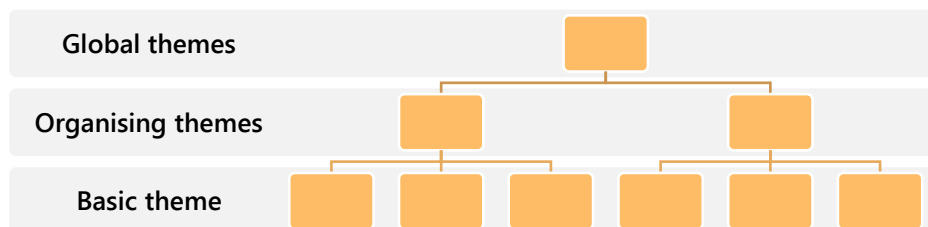


Figure 26 Thematic analysis approach

As such, thematic analysis is a method for identifying, analysing and reporting patterns or themes within the data in 'rich' detail (Braun & Clarke, 2006). This makes it an appropriate approach for cases with small but rich samples such as this. Thus themes are found across the dataset and are not isolated to a single participant. Themes may be big or small, focused or broad and may highlight areas of disagreement or dissonance across the participants (Braun & Clarke, 2006).

However, as described by Braun & Clarke (2006) it is an approach which is often left unclear by researchers making evaluation, replication and comparison difficult. To counter this, the methods described by Braun and Clarke (see Table 6) will be followed closely.

Phase		Description
1	Familiarise the data	Transcribing data, reading and re-reading the data, noting down initial ideas (memoing)
2	Generating initial codes	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code
3	Searching for themes	Collating codes into potential themes, gathering all data relevant to each potential theme
4	Reviewing themes	Checking if the themes work in relation to the coded extracts (Basic theme) and the entire data set (Organising themes), generating a thematic 'map' of the analysis
5	Defining and naming themes	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme
6	Producing the report	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis
<i>Table 6 Thematic Analysis procedure (Braun & Clarke, 2006)</i>		

Basic themes were found by reading through the basic descriptions of the codes used, and linking similar ones together using hand-drawn network maps. An example of this can be seen in Table 9. These helped to categorise the list down into a few basic themes. The hierarchy consists of basic themes, organising themes and global themes which are also shown in Table 9.

4.7 Summary

This chapter has presented the use of semi-structured in-depth interviews with industry participants. Additionally, workshops are used to help investigate the 'usefulness' of the streamlined Cory method. The manual approach of this tool is also contrasted with the automated Apidae calibrator tool to provide additional perspective on the approaches of other calibration methodologies present in literature, most significantly the Reddy method.

5 Findings

This chapter presents the key findings from the industry interviews in research phase two. These are presented in four parts; the first describes the thematic analysis process used in further detail, the second explores the pre-interview and how the participants currently undertake energy efficiency assessments, BEM and quality assurance procedures they currently undertake when modelling. The third explores their thoughts on the modified Cory calibration tool, its usability and potential 'usefulness' in industry. The last section links the two interview sets together and looks at the wider use of modelling and calibration with respect to current practice.

5.1 The analysis process

5.1.1.1 Step one: Familiarising the data

Familiarisation and immersion with the data was completed through listening to the raw audio recordings multiple times and reading and re-reading the transcriptions.

After the recordings were transcribed, they were listened to in full a number of times before and throughout the analysis phase. This was often accompanied by reading the transcriptions and as advised by Braun and Clarke, done in an active way by searching for meanings and patterns in the text with the creation of small memos and notes about possible themes. These memos were collected no matter how unimportant the theme felt at the time.

5.1.1.2 Step two: Coding the data

The data was coded soon after the data was considered familiar. An initial list of codes was developed using descriptive coding, a simple system to summarise a short passage of data (Saldaña, 2015). Saldaña recommends it as an excellent start for initial coding. Codes were applied to whole sentences and phrases where possible to keep the analysis tight and simple.

Transcription Example	Code
<i>Typically, would model the system when we've got 7/8 sub-systems within the building itself¹.</i>	¹ Modelling detail
<i>Typically, chillers, pumps, DHW, fans, common light power, lifts etc. for base building. We try to make sure the building is metered in a way that can be replicated in the simulation which is a critical point at the start of the project².</i>	² Metering
<i>Table 7 Example of initial coding</i>	

The entire data set was coded through in full resulting in 38 codes for the pre-workshop interviews and 35 for the post-workshop interviews. The data was coded for as many themes as possible at the beginning, with codes often being nuanced and sometimes overlapping. Coding continued through all of the dataset with new codes constantly being added. After this first cycle of coding, the codebook was reviewed and tabulated to see which participants said which codes and how often the codes were used (see Table 8).

Themes and coding analysis	Occurrences				
	Interview 1	Interview 2	Interview 3	Interview 4	Total
Modelling detail	3	3	6	5	17
BEM input assumption	3	2	2	3	10
BEM limitations	2	3	3	2	10
Experience	4	2	1	1	8
Building Knowledge base	1	1	5	1	8
Building control	4	1	0	2	7
<i>Table 8 Example code table from the pre-workshop interview</i>					

The codebook was then re-assessed for redundant codes and codes that were too similar. The dataset was then re-coded. At this point the codebook was ready to be searched for basic themes. Multiple codes could be used across themes as there are a lot of inter-relationships and cause and effect scenarios.

5.1.1.3 Step three: Searching for themes

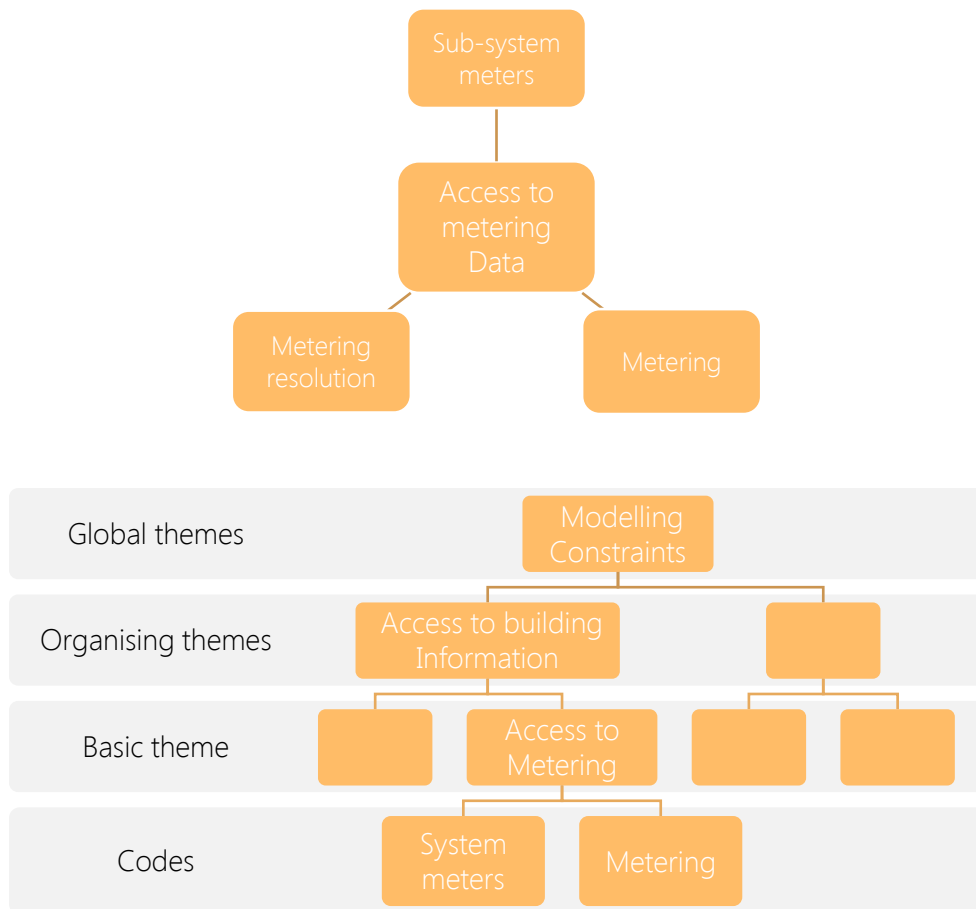


Table 9 Example thematic analysis network map

This process is improved by the use of a secondary cycle of coding. Axial coding is used to determine which are codes are the most important ones by using a central category (or in this case theme) which best describes the linked first cycle codes (Saldaña, 2015). i.e. the basic theme is literally considered the 'axis' of axial coding, with the first round of coding linking into it like spokes on a wheel. Thus, first cycle codes are components of this axis or basic theme. At the end of step three, 16 basic themes were found.

5.1.1.4 Step Four: Reviewing themes

Step four refined the 16 basic themes found in the previous step. This was done by collating the coded extracts together and reading the themes horizontally, that is only looking at coded extracts relating to a single theme from across all four interviews and checking them for thematic consistency. The aim of this step is to refine, rename or omit themes that did were incoherent or

could not be distinguished from another theme. This process took the number of basic themes down 16 to just 10, as shown in Table 10. The final thematic network maps are given in Figure 27

Limited need and Limited budget	Non-standard auditing and engagement	Limited time available for design	Lack of a 'usable' simulation software front-end	Ability to understand the buildings' control structure
Available tool chest is good	Ease of use and time constraints	Lack of guidance and protocols	Gap between literature and industry	Control representation in software

Table 10 Basic Themes

5.1.1.5 Steps Five and six: Defining and naming themes, and reporting

Step five requires the basic themes to be assessed against each other and against the dataset itself. Basic themes are made distinct from one another so that there is no overlap. For example, the themes of 'metering' and 'sub-metering' can be combined into a single basic theme. Step six is reporting the findings and themes by presenting examples and providing analysis on why they are significant.

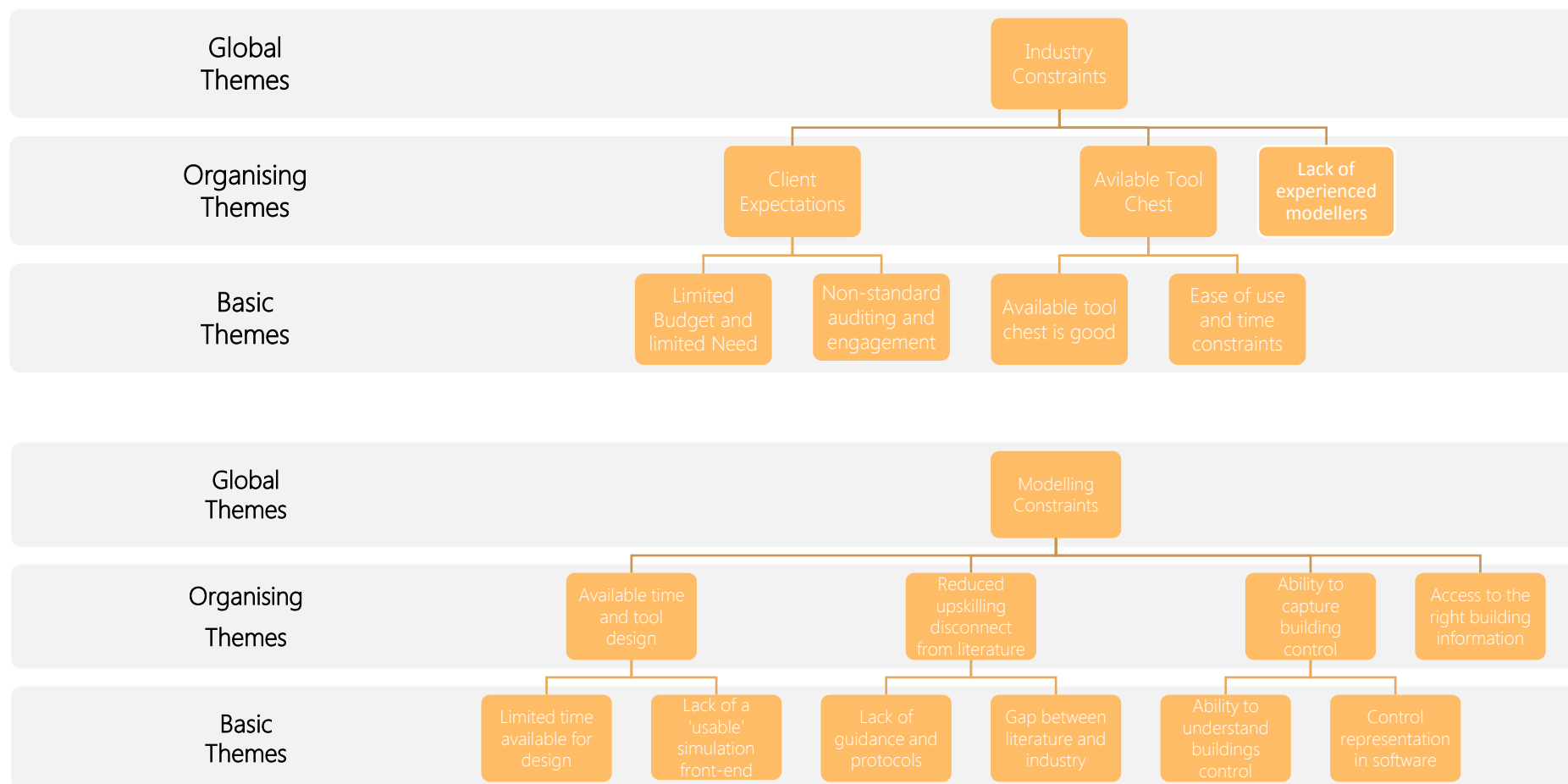


Figure 27 Thematic Networks - Industry and Modelling constraints

5.2 Pre-workshop Interviews

5.2.1 Findings using thematic analysis

The pre-interviews were used to establish how energy efficiency work is carried out in industry, and how the use of BEM relates to that. Two global themes (see Figure 27) emerged; Industry constraints and Modelling constraints. The first theme has been entitled *Industry constraints* relating to constraints placed on the adoption of BEM in industry which are directly related to the market or current practice. Investment in EE, processes in energy auditing and project objectives and constraints dictating the way an assessment is carried out. It was found that building tuning and systems optimisation that is simple to change, readily accessible and/or inexpensive (what might be deemed as 'low hanging fruit'), are more common than full-scale retrofits. This usually means that the minimum level of building data required to make a BEM is not available to the engineer. Thus static spreadsheet-based empirical models, which have been well developed and tested for simple one at a time changes are used. Due to time and budget constraints, BEMs become a second choice option for the majority of assessments. In addition, industry experience is that BEM appears to be undervalued in New Zealand compared to other such as Australia and the United States (personal correspondence with NZGBC October 6th 2015, and EMANZ October 7th 2015).

The second global theme was entitled *Modelling constraints* and is based on the ability to actually model and produce a BEM. When simulation is used, the opportunity to engage in any form of systematic calibration process is rare. Generally, one in-house review examining the plausibility of the model and its inputs is as far as any quality assurance measure goes. These assessments are similar to those performed by researchers in the 1980s (see section 2.4) and do not use statistical measures to assess accuracy. When modelling is used, it is generally for small and simple projects with many unknowns, or in large and complex projects such as whole-building retrofits. In the former type, BEM is generally used as either a diagnostic measure (tuning the building based on the simulation results), or for M&V (where the model is used to verify energy

savings). In the large and complex projects, BEM is used as a design tool to interrogate possible ECM alternatives, or more commonly as an M&V tool to forward predict energy savings across major retrofits.

Lastly within the global theme of modelling constraints is the importance and difficulty in capturing and replicating building control in BEMs. This was presented by all four participants as a major issue and would appear to be the most misunderstood, under-appreciated aspect of Building Energy Modelling in the literature. This limitation is due to two things; the ability of the modeller to gather, understand and replicate the interactions between building systems as controlled by the Building Management System (BMS), and the ability of the simulation software to replicate real world control logic and algorithms.

The analysis suggests that both of these global themes contribute to the limited use of BEM in industry thus creating a credibility gap for practitioners. BEM is seen as a higher risk option that has more potential to go wrong without necessarily producing a better result than empirical models, especially on smaller projects which make up the majority of the EE projects. The increase in time and cost to use BEM as well as the potential for embarrassment is a major factor in this decision (as stated by participant 4). However, all four practitioners indicated that while the inclusion of BEM was necessary on some projects, it would benefit virtually all projects.

5.3 Industry Constraints

Industry Constraint of the use of BEM for energy efficiency assessment and retrofit in existing buildings is the first global theme relating to question one. All four interviewees held the position that most of the work requested by building owners is too limited to warrant the use of BEMs. The analysis identified two organising themes under this global theme, each with two basic themes.

5.3.1 Client expectations

5.3.1.1 Limited need and Limited budget

Understanding and the client's needs was the key element to question one. The market for energy efficiency assessment plays a central role in the use of BEM.

Among all four participants it was found that often the key drivers in Energy Efficiency assessment were time, hindrance and cost.

The work requested by BEM clients varies greatly. While standard auditing practices were conducted by the two most experienced participants, the others are more often engaged in system optimisation work. These are optimisations of the existing systems, most often a single system in isolation, rather than an explicit search for ECMs that might be found in a type two energy audit. This represents a specific approach rather than a general one. In other words, inexpensive tweaks to the existing building are preferred by clients to larger and potentially more costly changes. These tweaks are often in the form of system optimisations to systems such as the HVAC e.g. returning old or faulty HVAC to its original state or improving the system efficiency. These can be for diagnostic, repair, or replacement reasons but tend to be limited in scope. While this may not be a new phenomenon, it is clear these contracts often do not require, or acknowledge the possible, use of a BEM.

However, it is important to note that this was suggested by all participants to be the opinion of the client and not the consultant. In fact, all participants advocated that the use of BEM on existing buildings could have a positive effect. When asked what proportion of projects required or could benefit from the use of BEM, all responded that few projects ever *required* the work, but that all could benefit from it. Generally, this benefit would be seen as a means of investigating different design opportunities or as an M&V procedure. The decision to include a BEM is, however, down to the client's expectations, requirements and budget. This difference of opinion between the two parties was summarised quite succinctly by one participant;

"The market really doesn't pay for the work"

- Participant 1

This participant further presented that in many cases they would create a BEM for their own purposes – usually research related. They went on to say that often a budget will not recognise BEM or calibration (regardless of whether it is used or not) making it an oft unrewarded venture.

While this response was given by three of the four participants, the fourth suggested that in their work BEM was used in the majority of their projects big or small, though the cost and scale of these models was significantly less than a typical client model. This participant identified as having more simulation experience than typical engineering experience and it was found that they kept more current on simulation practice than the others. They went on to suggest that modelling from the position of a building scientist was far more likely than from the position of an engineer, and that education and training are central to the use of BEM as a normalised process in industry.

"[On] the projects we work, less than 5% have a simulation energy model in the upgrade and generally the simulation [is used] where an empirical model doesn't suffice."

- Participant 2

What requires further investigation is the attitudes towards energy efficiency from building owners. The reasons for the perceived limited need of the use of BEM on existing buildings appear due largely to unclear expectations and benefits for building owners. It would appear the goals of building owners tend to be focused on returning the building to a more efficient status at minimal cost. Projects that use BEM to improve EE to more than this level are less common because of the added difficulty, cost, uncertainty and perceived opportunity cost. This route is seen as less uncertain and less arduous than a full scale retrofit. Therefore, investigating design options for large scale change through the use of BEM is not considered to be worth the effort

5.3.1.2 Non-standardised auditing and energy efficiency engagement

As building owners take a more conservative approach to Energy Efficiency assessments, so too does the use of auditing in the assessment. All four participants conduct EE assessments in slightly different ways, often as a variant on Energy Auditing standard NZS 3598:2014 (Standards New Zealand, 2014). Four types of audit were identified between the practitioners; Types one, two and three audits and a 'systems optimisation audit' that sits somewhere between a type one and two, which was explained by one of the participants:

"[A] Systems optimisation audit [is] neither a type one nor two, but a hybrid between the two. You don't have to follow the New Zealand standard for auditing but you employ some of the techniques [to get similar outcomes]. You're looking for energy saving measures and you have to have a good idea of where the energy is going in the building but you don't need to necessarily calculate the exact breakdown of where the energy is going"

- Participant 4

This might be classified as type three-lite, as it sits in direct competition with the traditional type three audit. It borrows elements of the type two audit without needing to adhere to the full requirements of the type three standard.

The auditing procedure used directly impacts the amount of building information the practitioner can collect. The narrower the scope of the project, the more limited the amount of time to understand the building and how the systems interrelate. Not only does this approach reduce the number of ECMs that can be identified, it also limits the possibility of creating a BEM for the project. Instead more traditional empirical models are used to quantify savings.

5.3.2 Available tool chest

5.3.2.1 Available tool chest is good

In one interview (Participant 3), the term 'tool-chest' was used to describe a body of well established, tested and reasonably accurate empirical models used to assess energy performance. This 'tool-chest' is based on the static steady-state modelling described earlier in section 2.2.1. These models, though simple and inadequate for complex ECM estimation, serve as good starting points and rule of thumb estimates. These models range from simple R-value calculations, to more complex equations mapping out efficiency curves for HVAC plant. The idea of the 'tool chest' is the range of problems that can have a commonly applied solution, such as widening the setpoint deadband to reduce heating and cooling consumption.

"All could benefit from the [BEM]. But the simple matter is time and money prevent this from happening. And to be honest rules of thumb will get you 90% of the way there in most buildings"

- Participant 3

Three of the participants use empirical models for 90% to 95% of the work they do, with the fourth around 60% to 70%. These models work very well for simple, low risk projects and can provide reasonably reliable answers. As such there is little incentive to move away from them, particularly when clients are requesting simple changes such as system optimisation, or the replacement of a single bit of HVAC plant such as a chiller. These calculations also provide for a very well evidenced calculation, where the total number of calculations is low, making error checking a quick and straightforward process. Using empirical models is less risky and uncertain than using a BEM which requires a larger amount of accurate building data before the results can be considered reliable.

"The times where a BEM is useful is firstly in new build or when you're doing a refurb[ishment] that is so deep that you're throwing away so much of the system that you cannot get useful information from it"

"We can empirically determine how the existing systems are working, but it doesn't help us understand when we put a new HVAC system into the building."

- Participant 1

While BEM is considered one tool in the tool chest, it is not seen as the 'go-to' option. The preferred option from three of the four participants was to use BEM as a last resort option. The reason put forward for this is that empirical models are able to get a good enough result for the work that clients are asking for with fewer risks. The exception comes when the project has too many unknowns to make empirical models unreliable or when the project scope calls for a retrofit too deep for empirical models to be of any value – in either case it is a question of the amount of building information.

5.3.3 Lack of building experience amongst modellers

All four participants noted the need for experienced modellers to use the Cory method rather than inexperienced modellers. One participant, when asked if the process would help improve the validity of their models replied:

"[if you're an experienced engineer] sure, 100%. But if you're looking at a graduate student and they're got very limited experience in modelling, and they're given a job, and that job wasn't verified, they'd get it wrong 99 times out of 100"

- Participant 4

This was counterpointed by another participant who mentioned a 'glass ceiling' for energy modellers in practice, suggesting that it is difficult for modellers to progress beyond a certain level.

"It's a terrible truth of simulation that people [who] do simulation work, and the people that manage simulation work, are very different people. So that the people that manage may have a lot of experience in buildings. But the people that do it are often not much more than code monkeys"

- Participant 1

These points highlighted an important aspect of modelling and simulation in practice. Experienced engineers and building scientists are less likely to be 'hands-on' with BEM for existing buildings because of the low value of it BEM models compared with their relatively high charge-out rates. This leaves them managing and checking models created by people experienced in modelling, but not in buildings. I.e. they understand how to operate the simulation software, but are less experienced with how buildings work and how different building systems interrelate.

"A lot of the people who do simulation models are graduates who don't have experience. It's difficult to find someone in the commercial sector who has a lot of experience. The career ceiling is too low and then where do you go from there?"

- Participant 2

5.4 Modelling constraints

Modelling and technical constraints were also a major contributor to question one. These constraints are from the perspective of the modeller and the challenges they face in process of create BEMs. Available time and tool design, an apparent disconnect with literature and BEM and research, access to building information, and the ability to replicate building control systems in simulation tools were the key contributors to this.

5.4.1 Available time and tool design

A lack of time spent simulating and testing design options is one of the key constraints affecting the adoption of BEMs. Much of the time set by the project constraints will be spent on gathering building documentation and other small yet demanding tasks. The time required to create a BEM of a large 10,000+m² building, to produce 'meaningful' results, was estimated by two of the participants as anywhere from 40 to 60 hours. While the time required will differ depending on modelling experience, software, building familiarity and how much of the building is being modelled, it does serve to illustrate the amount of time required to get to an 'initial' level of model accuracy, let alone a calibrated model. However, time is not just limited to the actual modelling of the building.

Simulation software front-end design is a major factor in this. There are trade-offs between flexibility and ease of use - flexibility to change the model at the users' discretion versus having a simple, uncluttered, usable interface. For example, EnergyPlus has flexibility, but is not as easy to use as Sefaira (a front-end to EnergyPlus), which in-turn does not provide the same functionality or flexibility as the base EnergyPlus programme.

The easier route is almost always taken; however, a large proportion of time will still be spent in other mundane tasks such as creating spreadsheets for data analysis. This has a direct effect on cost as the modeller has to spend more time on simple yet important work limiting the amount of time available to them for design testing and analysis.

5.4.2 Reduced upskilling and disconnect from literature

Little work is being done to further communicate practical simulation guidelines and protocols for industry. While it is primarily industry that constrains this, the effect on simulation skills and knowledge in practice is profound. In New Zealand, there remains a lack of awareness or ability for practitioners to move from empirical modelling to simulation to achieve energy efficiency retrofits. While, as stated before in section 5.3.2, there are times when empirical models are better suited to a task than BEM, the lack of upskilling in BEM may be affecting the use of BEM as a main tool for complex design strategies. Why bother learning to use simulation for the purposes of modelling existing buildings if clients will not pay for it? This is further compounded by a lack of available guidelines for good calibration and tools to aid in model calibration.

The gap between the user and the literature is vast. The array of approaches given in the literature from energy signatures, optimisation, Bayesian calibration, and statistical techniques (see section 2.4.2) is bewildering even for academics, let alone practitioners. The range of approaches in literature helps to show the importance of calibration to BEM, though does not advance its use in practice. Though it would have been surprising to see research based techniques (such as the Reddy method) *adopted* in industry, it was found that few, if any, were even known, let alone used. All four participants would use an in-house review of input and assumption data and ‘tinker’ with those inputs until they were either a) happy the inputs would not affect the model unreasonably, or b) plausible enough to continue without spending additional time adjusting them. With time at a premium, simulation outputs are more likely to be reported as being outside of the accepted range than corrected. Any additional effort to find, review and apply techniques from the literature was not considered to be worth the effort given the cost of creating a BEM.

“You’re tinkering with the model to adjust the parameters at the crudest level because here isn’t the time or money left in the budget... [At] a research level, if your fans were out you’d go back to your fan curves and the precise fan representations ... and fix all of that. We just don’t have the money to do that in a commercial project. You can instead say ‘well the fans are out’ and go from there.”

- Participant 1

5.4.3 Access to the right building information

Access to building information, or a lack thereof, is seen as a major hindrance to producing reliable BEMs. The practitioner's ability and willingness to obtain reliable building information is determined largely by cost and time. Building information required for BEMs is diverse, ranging from metered energy data to occupancy patterns to as-built drawings and operation manuals.

Of particular importance is the access to well formatted historical energy data and fuel data e.g. natural gas. All four participants stated that metered energy data was critical to making an initial BEM of an existing building (pre-calibration).

"Lack of metered/measured data 100%. Not just energy, but profiles and densities. Not only that but chiller efficiencies. A water based chiller could be really efficient at low loads, and then another chiller being more efficient at a higher load. But without knowing that, you could be out by 15-20%"

- Participant 4

Most of this information would be procured during an auditing process in either a type one or two audit. However, a lot of information, such as as-built information, is not found during this time. Understanding the BMS and the controls of the building was also of major concern. All of the participants felt that understanding how the BMS worked and linked into the HVAC and other systems was vital.

"They might have O&M (operation and maintenance) manuals or as-built drawings when the building was handed over, but since then they've been lost... Even in a new building there is stuff that isn't routinely documented the classic being the glazing types which no one ever writes down."

- Participant 3

As BEM requires a wide breadth of information to run, there were instances when the participants filled gaps by using assumptions based on similar cases

such as the glazing type. Given the ability to collect information is directly related to available time and cost, then the most important and relevant information will be collected first with the unimportant information often being modelled in generic terms or with assumptions based on other similar buildings.

5.4.4 Ability to capture building control

All four participants discussed the need to understand and replicate the BMS in BEMs, but more specifically the controls and algorithms that underpin those systems. There were two basic themes here; the ability to understand how the building management system operates from the stand-point of the engineer, and the ability of the simulation software to replicate these controls.

Control strategies can range from simple manual operations like opening and closing a window, to complex automated controls linked across multiple building systems.

"I think most simulation people don't understand controls. There is a grim reality that HVAC control that needs 10-15 years' experience to understand because it's about getting multiple systems to work together, so you need them mapped out in your brain so you can then think about how they work together. Rather than how they work individually."

- Participant 1

While knowing when a window is opened is difficult, simple manual changes to a building can be easily replicated in a BEM. However, it can be a lot harder for more complex automated control systems and often depends on the simulation engine software itself. Two participants highlighted this as a major concern or limitation to their models.

"The area that modelling really struggles to represent reality is in the controls of the HVAC kit. E+[EnergyPlus] is getting there with the 'if this then that' which is approaching some of the algorithms you might find in a BMS [Building Management System]. But if you start to look at complicated controls, that's where simulation really begins to struggle."

- Participant 3

“We’re not using EnergyPlus for two reasons. One is that the control representation is inferior to IES², it’s got certain blind spots.... the controls in EnergyPlus are entirely based on PID control, and the real world uses a mix of proportional and PID, and the lack of a proportional control module in EnergyPlus is actually a major problem.”

- Participant 1

5.5 Summary – Credibility gap

The credibility gap discussed by (Tupper, Franconi, Chan, Hodgins, et al., 2011) (see section 1.1) exists not only from a technical perspective, but also from an industry perspective.

All four participants were concerned about the use of the use of BEM on existing projects from a technical standpoint – it is a difficult and long process which is difficult to have confidence in. All four saw the use of BEM as a double-edged sword; while BEM can produce a sound and reliable outcome, it can also possible to produce a poor result. The lack of trust in BEM is due to a multitude of elements from a technical standpoint. Unintuitive software design, access to building information and the ability to understand and replicate a BMS are the main difficulties faced. These faults make the process less reliable in their eyes, though no less valid as an option – BEM was still seen as an excellent tool from which every project could benefit provided these issues could be addressed. This is a frustration to the practitioner, they can see the potential for wider use however they are not wholly trusting of their ability to use BEM with total confidence in its current form. These effects contribute to the limited use of BEM in industry, though they are not solely responsible for it.

The potential for embarrassment with BEM is significantly higher than with traditional empirical models. The simpler route is one much less likely to produce a poor result. In most cases clients are not asking for work which requires the precision that a BEM provides. And while this may limit the amount of work which does require the use of a BEM, the extra expense is not seen as being worth the effort – particularly if the modelling constraints persist. This has

² IES is an alternative simulation engine.

a knock-on effect; a reduction in practitioner upskilling and a disconnection with literature. The practitioner is less likely to develop the skills necessary to complete BEMs if these investments are not financially rewarded. The lack of calibration methods employed beyond a simple plausibility check was the case for all four participants. And while two participants had kept in touch with the literature, they employed almost no methods from it in their work. This represents a failure to take calibration research to industry in a usable way.

The practitioners also expressed concern at the potential for embarrassment due to errors when using BEM which could lead to liability issues. Two participants indicated times where a fault in their work was caught out which in turn soured their relationship with the client. The amount which can go wrong with a BEM compared to the traditional model represents the opportunity cost of BEM. Why would a building owner opt for the high-risk low-reward option when the low-risk low-reward option is still available? This is particularly important considering that the bulk of the work requested is what might be deemed as 'low hanging fruit' – simple savings or system optimisations. These can be assessed using traditional empirical models. The analogy can be made that BEM is a sledgehammer while traditional models are nut crackers. Sure, the sledgehammer *can* crack a nut, by why not use the nut-cracker instead? It's easier and is less risky. Occasionally a nut too big for the nut-cracker will come along and the only option is to resort to the sledgehammer. The problem is this is infrequent at best (as low as 5% according to one participant). BEM is therefore more likely to be undervalued as it is risky and often not suited to most projects. If this is the case for the expert user, it is even less likely that a less experienced practitioner is likely to use BEM.

The issue of calibration was not of immediate concern under these circumstances. However, that is not to say that the adoption of the Cory method would improve the reliability of the few BEMs which are used for existing building energy assessment.

5.6 Post-workshop Interviews

The post-workshop interview focussed on the calibration tool developed in research phase one, comparing it with the Apidae online calibrator, and

investigating its potential industry use. Emphasis was placed on identifying the barriers from both practical and technical perspectives.

It was found that none of the participants had knowledge of any tools specifically used for calibration. One did, however, mention a visualisation add-on to the IES programme. This add on does not help the modeller to calibrate, but rather displays a comparison of simulated and actual data.

Findings from the post-workshop interviews showed that the tool would be well received in industry. While not as 'polished' as the Apidae tool, the processes used in the Cory method were seen as being very useful additions to the modelling process.

5.6.1 Overall view of the Cory method

The tool was seen as usable and useful by all four participants. In their eyes it acts as a procedural extension to regular work. The most important aspect for all four participants was the ability to understand and control the calibration process. They liked the simplicity of making a change, seeing the results, and then re-evaluating the model. Indeed, they commented that it is not too far removed from their current practices of 'guess and check'.

"It's definitely a lot better than the default method of sticking your thumb in the air and hoping. You're informed about what you're changing and that you're changing it appropriately rather than operating on a hunch."

- Participant 3

The process reduces the level of uncertainty input by the user and provides them with a tool for quality assurance reporting. The user is able to present how the model was changed in a systematic way rather than relying on experience, something that the client would have to place faith in previously. This was seen by all of the participants as being a simple, learnable process that could be adopted in industry without adding significant expense.

"The thing is, because it cuts the time by so much, because it makes calibration a systematic, understandable process, then it would be done [in practice]. It takes it from a research task to an industrial process task."

- Participant 2

The introduction of the energy signature method was more successful than anticipated because it was one of the few methods beyond the simple plausibility checks the participants had seen, despite being experts in industry.

Overall the understanding the calibration tool by all four participants was informed by their understanding of building physics. The Cory method worked because it operates in familiar conditions to standard practice i.e. the process does not require knowledge in the areas of sensitivity analysis, advanced mathematics or computer science, and so is not seen as a 'black box' calculation.

"The most attractive part of it is that you've got some control over it rather than say another program where you potentially don't where it does something in the background and [then] does a bunch of different things"

- Participant 4

5.6.2 Positive aspects of the Cory method

5.6.2.1 Good user control

The user remains in control of the process at all times, and while this does come with its disadvantages (see section 5.6.3), it was seen as being the most important factor by all of the participants. Control over the process meant two things; being able to control what was changed and by how much. In contrast the Apidae tool provides some, but not total, control. The Apidae process is automated with the user providing a selection of parameters to change and range of possible values to work within. The participants felt generally, that the more control they had over the process, the more comfortable they were with the results and how those results came about.

5.6.2.2 Satisfactory to use

All four participants saw the use of the signature method as understandable in principle. That is to say that when the calibration and characteristic signatures align, they understood that the parameter they were looking at was a likely candidate for change. This was reflected in the System Usability Scores (SUS)

given in Figure 28. Three of the five SUS scored fell within the 'good' range and two fell within the 'Ok' range (Bangor et al., 2008) making this a successful result. It should be noted that as the participants are not expert evaluators this figure must be considered as indicative only.

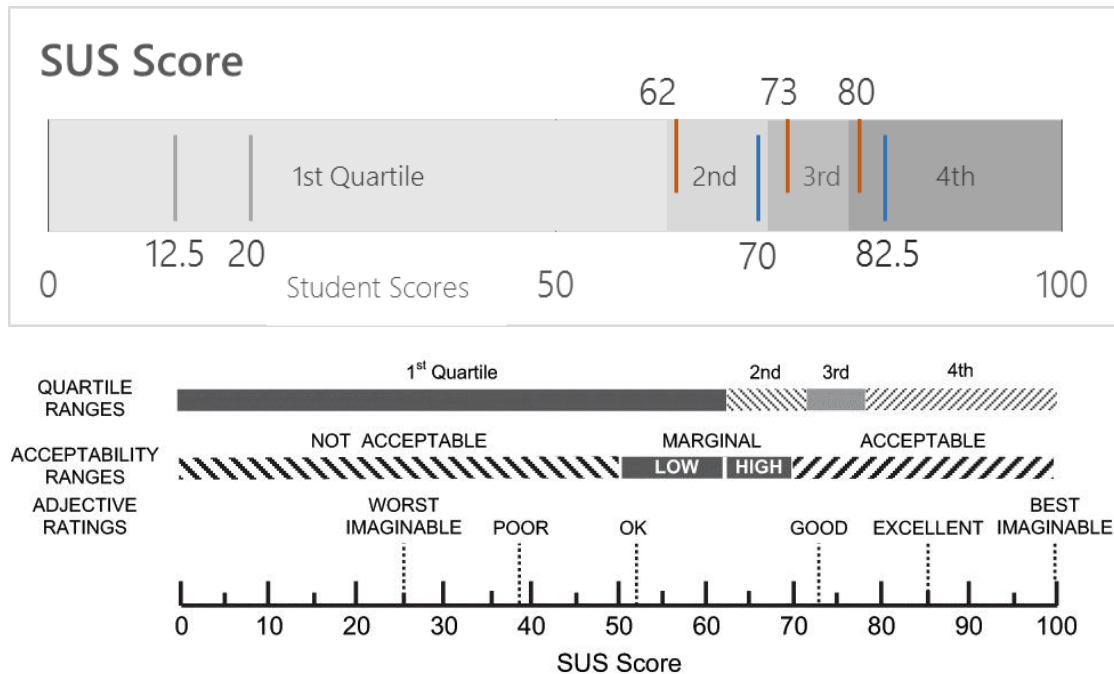


Figure 28 SUS scores and acceptability ranges for the streamlined Cory tool

The tool itself, though still limited, was able to be used easily with the participants making changes to the models in quick, easy succession. The participants all felt that with a detailed tutorial they could quickly master the process.

5.6.3 Negative aspects of the Cory method

Not all the feedback on the streamlined Cory tool was positive. It took a long time for all four participants to read, understand and make decisions from the energy signature graphs indicating that they were still problematic.

Energy signature interpretation

Three of the four participants found interpreting the energy signature graphs quite difficult. While they could read what the graph was showing, interpreting what that meant for the model was a lot harder. This was due to two things: firstly, the tool did not provide direction for what to do next in the process. Often the participants would find themselves asking if the parameter value

needed to be increased or decreased to match. This was most often found out through trial-and-error, however it was not as intuitive as they had previously thought.

"Having some hints on how to interpret the graphs ... [how about] some subtle hints?"

- Participant 3

Once the graphs were deciphered, it was difficult to say exactly how an input value should be changed. For instance, if an Energy Signature showed that the cooling setpoint was too high, the immediate response would be to lower it, however one participant continually suggested that to change one parameter would have an effect on all of the other related parameters. While this is true, when the calibration signature is reassessed, any changes to the other parameters will be presented to the user. However, this does show that the participant was aware of the interactions between different parameters and their likely impacts on the model, something that a less experienced individual may not have picked up.

5.6.3.1 The need for experience

As a continuation of the industry constraint '*Lack of building experience amongst modellers*' (see page 102) the participants all put forward that the Cory method, and any other calibration tool, ought not to be seen as a magical wand that can fix any un-calibrated model. This was particularly important for automated calibration and optimisation tools (discussed further in section 6.1) which aim to give the best answer in the least amount of time possible. While the process was an improvement to current practice, it should still be used with caution and vigilance.

"If you've got the nous to model the building in the first place, then you've got the nous to use this tool"

—"Do you feel that it still gives enough freedom? So that a 10-year-old could do it?"

- Researcher

"No you still need the experience to be able to drive it. It's that classic thing, [if] simulation [is] too easy then anyone thinks they can use it. And if anyone thinks they can use it then you get chumps using it."

- Participant 3

This point made clear the importance of balancing usability and user experience. While the usability needs to be good enough for the engineer to be able to use the tool (and want to use it), the participants all felt that the tool needs to remain in the hands of more experienced users. For instance, a user such as an architect or junior building scientist with no training in services or mechanical engineering could see the Cory method, or any other calibration tool, as a means of 'fixing' their models, when in reality it only gives the user more ability to find out what is wrong in the first place – false positives are still a very real possibility. All four saw the tool as a means of improving the ability of the engineer to quality assure the model, but that the process must be guided by experience.

5.6.4 Contrasting manual and automated approaches

The differences, both positive and negative, between the automated and manual approaches are described in this section. In contrasting the manual Cory method and automated Apidae tools against each other, a lot of discussion emerged about the degree of automation and optimisation that should be used in calibration. The Cory method is an almost entirely manual approach – the user needs decides what gets changed and by how much. In contrast the Apidae tool is more automated optimising the values of each parameter chosen by the user.

The Apidae tool was seen quite differently to the Cory method. Despite being far more 'polished' and user friendly, it was seen as something to be cautious of in practice, largely because it is an automated process.

5.6.4.1 Perceived advantages of the automatic approach

The Apidae method was seen as far more refined and attractive for use in current practice largely due to its more refined and intuitive interface and limited learning curve. While a usability assessment was not carried out on this tool, all four participants commented that it would be significantly easier to

introduce into practice than the Cory method as the user has significantly less to do and less to learn.

The limited amount of work required by the user to get a result is due to the 'hands free' nature of the Apidae method. The user does not need to be present while the calculation algorithm runs on a server off-site. All that is required of them is to pick the parameters that are (in their judgment) uncertain, pick a range of possible values for these parameters and then leave the process to figure a result within the bounds set. As a consequence of this, the user has no requirement to learn or change the parameter values themselves, meaning the process can be operated by anyone provided they choose the right parameters to work within.

Given the current constraints, as outlined in section 5.3, all four participants believed that the level of automation used by the Apidae calibrator was acceptable and indeed very attractive as it cuts the amount of time and working down by a significant margin. This does, however, present other issues highlighted by the participants.

5.6.4.2 Perceived disadvantages of the automatic approach

The biggest concern with the Apidae calibrator was that the input parameters still needed to be picked by the user. This piqued distrust in how the Apidae calibrator could be used by inexperienced modellers to calibrate models incorrectly leading to false-positive outcomes. This is illustrated in one comment:

"someone from the outside has the point of view of "are you kidding me? There's a tool I can you to optimise it to make it match!?" from their perspective they're not going to care about the science, they're just going to want it to match."

- Participant 4

Presuming the optimisation algorithm behind the Apidae calibrator results in accurate outcomes, the problem with the process is that the user sets the boundaries for the algorithm to work within. This level of control was not seen as a positive as it gifted inexperienced practitioners the ability to come to false-

positives with the aid of a reputable tool (meaning they could justify their answers by providing proof despite possibly being incorrect). As the input parameters selected for optimisation are still heuristic, the level of experience informing those selections is still paramount.

“Well you’re still picking your variables so the fact that you’re picking your variables is putting a lot of user control in there. And one of the challenges I found with a lot of simulators is that they don’t understand the variables very well. So they will pick their pet variables, and so the methodology needs to lead rather than lag on that point. It actually needs to either have a comprehensive set of variables that forces the simulator to look at, or has some other methodology to make sure that it’s not limited by the imagination of the simulator. And a lot of simulators tend not to have very big imaginations. Typically, because they’re trying to get the job done on time and on budget”

- Participant 1

The difference in approach between the manual Cory method and the automated Apidae tool can be seen here. The Cory method looks to confirm or disconfirm the suspicions of the modeller by identifying correlated signatures while the Apidae tool looks to identify which combination of the selected parameters is likely to give the best result. Both methods still rely on the modeller to select the right parameters. The manual approach was seen as being at an advantage here because it still requires the modeller to exhaust the range of input parameters available in order to get to the right result whereas the automatic approach will give the user the best result only from the range selected.

This brings forward two points; firstly, it assumes that the manual approach will lead to a better outcome as the modeller must be more diligent with the first step in the process (selecting the right inputs). Thus, if given enough time, it will eventually come to the right result. However, this raises an issue perhaps solved by the second point. That is, if enough parameters are selected, and the ranges set wide enough, the automatic approach will come inevitably come to a ‘satisfactory’ answer, and in doing so, may save the practitioner time in exchange for computing time.

This does, however, place inherent trust in the optimisation algorithms and automated processes in the Apidae calibrator. All four participants held an immediate sense of scepticism towards the Apidae calibrator citing it as a 'black box'. However, this misgiving might be cleared if the algorithm underpinning the Apidae optimisation tool were known to them.

5.6.5 Summary – comparing the two tools

Overall none of the practitioners were particularly familiar with calibration methods present in the literature, largely as a result of the disconnect between industry need and research focus. The manual Cory method was seen as being the preferred *technique* for calibrating, that is, the one they would place more trust in because they could understand each step in the process. However, the automated Apidae tool would be more likely to be adopted in industry because of its quick and easy to use nature. Though this is for practical reasons, not for accuracy reasons.

The ideal approach would in fact be a balance between the control of the manual method and the computational power of the automated method.

6 Discussion

6.1 Research question one

Is a lack of a fast and usable calibration process inhibiting the use of BEM in industry?

While it was the intention of the first research question to understand and compare how current BEM modelling and calibration practices in industry compare with the Cory method, it was found that BEM is not used by the participants for the vast majority of existing building retrofits.

Two global themes were found to underpin this finding, industry constraints and modelling constraints. However, the adoption of BEM as a tool for use in industry is limited firstly and foremostly by industry constraints. A value gap is present where, building owners have less incentive to undertake more expensive whole-building retrofits over small-scale optimisations at little up-front cost. BEM was seen as the 'last resort' option by the participants due to its complexity and expense. This reduces its uptake in industry as a means of assessment for existing building retrofits. Three factors contribute to this; firstly, traditional empirical tools are seen as more than adequate for small-scale optimisations. BEM is not well suited to this type of work as it is too costly.

Secondly, the uncertainties in using BEM for existing buildings limits uptake. As described by the participants, they would rather 'play it safe' with a low-risk, low-reward strategy given the uncertainty in outcome of BEM, particularly when lacking key input variables.

Thirdly, the resulting opportunity cost makes BEM an unattractive option, especially without incentives.

6.1.1 Preferred use of static empirical models to dynamic models

The preferred use of empirical models and traditional spreadsheet calculations is entirely based on the simplistic nature of the work that is being requested of practitioners by building owners. As demonstrated by Alajmi (2012), simple savings of under 10% could be found at little or no cost to the building owner in simple walk-through audits, equivalent to a type one audit as per NZS3598:2014 (Standards New Zealand, 2014)). It was made clear that energy auditing and deeper energy efficiency investigations for existing buildings simply are not a common occurrence for practitioners, at least in part due to a lack of knowledge in the building owners.

However, government funding is made available to businesses who wish to undertake the first steps for improving their buildings. In New Zealand EECA offer up to 40% of the cost for energy auditing, systems optimisation and performance monitoring and targeting (EECA, 2016a, 2016b, 2016c). These funds are subsidies only, with the building owner still required to provide the remaining 60% of the cost. The purpose of this is to stimulate growth and awareness, and the uptake of basic ECMs amongst building owners.

While this is not applicable outside of New Zealand, it does serve to illustrate possible reasons for the increased use of empirical modelling for energy efficiency assessments, and thus the answers given in the interviews. For instance, one participant remarked to a client:

"...instead of spending 80k, [we'll] spend 20k and we'll do a desktop study where we use the set of generic model and all you're doing is changing variables like orientation, insulation etc. and then you then generate a list of variable that impact on overheating so you then essentially have a spreadsheet tool which saves the client money but isn't necessarily as accurate"

- Participant 4

Shapiro (2011) and Olgyay and Seruto (2010), found building energy audits are not evaluating the majority of potential retrofit solutions, leading to limited

savings. Shapiro found cases where savings of under 10% were presented but a well-executed project could find more than 40% savings. Ultimately it is cost, followed distantly by fitness for purpose that drives the use or non-use of a BEM.

6.1.2 Dealing with uncertainty

Uncertainty is present in a lot of different areas of building energy modelling, but it is generally associated with a lack of access to building information, and the impact of modeller assumptions. Despite many efforts, the impact of uncertainty is difficult to calculate in building energy modelling. Reddy et al.,(2007a) notes not only does the number of possible combinations of uncertain parameters increase exponentially (e.g. 3 uncertain parameters with 4 possible values is $4^3 = 64$ combinations, 4 parameters with four possible values $4^4 = 256$ combinations), but identifying how what impact these unknown parameters has on the model can greatly affect its outcome.

Berkeley (2013) discusses the importance of understanding the impact of modeller assumptions and the sensitivity of buildings in different scenarios, suggesting that the parameters considered to have the most significant impact on the model may be variable with context. So not only does this problem become exponential when uncertain parameters are added, but the importance of these parameters are likely to be heuristically defined by the modeller. Correctly estimating the uncertainty in the model is therefore determined in-part by the amount of time available to the practitioner and their experience.

6.1.3 Experience vs inexperience

The experience of the modellers plays an important role in the quality of the simulation. Determining how the building should be modelled and how unknowns should be accounted for are decisions left to the experience of the modeller. Also, modellers in practice are required to be skilled, however, there is limited work available to them meaning there are also career obstacles in addition to a lack of work in BEM.

Given that time is often limited, heuristically defined parameters are therefore likely to be based on experience meaning that for inexperienced practitioners,

the potential for embarrassment is high. This point re-iterates the importance of domain experience in the modelling process discussed by Ibarra and Reinhart (2009). This point is (again) summarised by one participant:

"[A problem for] ...a lot of simulators is that they don't understand the variables very well. So they will pick their pet variables and so the methodology needs to lead rather than lag on that point it actually needs to either have a comprehensive set of variables that forces the simulator to look at, or has some other methodology to make sure that it's not limited by the imagination of the simulator. And a lot of simulators tend not to have very big imaginations. Typically, because they're trying to get the job done on time and on budget."

- Participant 1

The potential for embarrassment due to savings not matching predictions may in-part be due to a lack of an available calibration procedure to help modellers with uncertain parameters. Furthermore, as mentioned in section 5.6, no participant was aware of a tool specifically used for calibration. These issues stem from the second major theme; modelling constraints (see section 5.4).

6.1.4 Balancing risk and reward

The answers to research question one may be summarised as the balancing of risk and reward; or more succinctly, the opportunity cost. Given the modelling constraints identified by the participants, why would an industry professional, experienced or not, undertake BEM when the risks associated with its use are possibly larger than the rewards?

Essentially the modelling constraints are tied directly to opportunity cost and trade-offs. What is the practitioner willing to sacrifice in a project? Research has shown it is feasible to calibrate BEMs to a high level (Coakley, 2014; Raftery, 2011; T. A. Reddy et al., 2007; Robertson et al., 2013 etc.), but these efforts are resource intensive often taking thousands of simulations and requiring precise building information and are often made with near-complete access to building data. Furthermore, the construction industry under-values BEM in the eyes of the interview participants. This limits the amount of time and thus the quality of outcome that can be achieved making the process extremely difficult with little

guidance from BEM guidelines and protocols (see section 2.7) or tools available to assist in this process.

6.2 Research Question two

Would the introduction of the Cory method improve the outcomes of energy modelling in industry?

The evidence suggests, yes, the adoption of the Cory method would improve the outcomes of BEM in industry. Though it would not necessarily increase the *number* of BEMs being used, it would improve the *quality* of the model. The low number of BEMs being used in energy efficiency retrofit projects is not just due to the lack of a calibration process, but is rather linked to a set of reasons leading to a value and credibility gap for BEMs.

Despite their expertise in the area, all four participants appeared to be mainly using simple plausibility checks and trial and error to calibrate their models, similar to that of Carroll and Hitchcock (1993). This process is vulnerable to all of the uncertainty issues stated in section 2.3.3.1, and is still time consuming and tedious. While the participants all believed that the Cory process was firstly more systematic than any approach they had used before, and secondly helped inform the user of which parameters to change, the workshop process still revealed some useful discussion points.

6.2.1 Automated vs manual approaches

The use of a manual approach like the Cory method was met with favour as the user was in control of the process through each step, even if this was more tedious than the mostly automated Apidae tool.

Berkeley (2013; 2015) suggests that the user cannot be taken out of the process entirely, meaning that fully automated approaches are not necessarily the best option as often there are decisions that must be made by the modeller. However, this must be balanced with the modeller being guided so as not to make arbitrary and impactful decisions. This is an issue with the control the user has over how the parameter values are selected and the modelled values. As stated in section 6.1.2, this raises the issue of dimensionality i.e. the more uncertain parameters there are, and the wider the

range of possible values for these parameters, the wider the number of total possible combinations there may be. This means that any arbitrary range of parameters may lead to a non-unique solution. Given enough parameters and wide enough ranges, it is possible to 'calibrate' or match the model to a non-unique solution (Carroll & Hitchcock, 1993). Thus any automated approach with an inexperienced modeller may produce a false-positive.

Following Berkeley's (2015) recommendations, a number of actions would help to alleviate this issue; firstly, providing a choice of standardised uncertainty profiles (ranges of possible values for common parameters bounded by confidence intervals). Secondly, incorporate a 'status bar' for each parameter to help users understand which parameters are likely to have a significant impact on the results. Thirdly improve modeller education on the impact of decision making and acknowledge bias.

These findings assume the continuation of a manual approach, where large sensitivity analyses are undertaken on other generic or templated models. However, the increase in computer power and commercially available sensitivity analysis tools may allow the modeller to undertake these assessments themselves.

6.3 Research question three

Which aspects and features of calibration tools require development in order to increase industry uptake of BEM quality assurance?

A wide range of features and debate were generated by this question. The participants were interested in the improvement of automated approaches such as optimisation. However, it became clear through the interview coding and analysis process that the issue related more to creating a market for the use of BEM rather than dealing solely with calibration

6.3.1 Future considerations for calibration tools

The final portion of the interviews looked to future development of BEM calibration tools. This is speculative, and conversation in the interviews was generally isolated to a number of 'wish list' items that would improve the outcomes of BEM calibration for the modeller. As such, these are 'findings' or

'results' as such, but rather as open discussion about what features industry practitioners would like to see in future tools to help overcome some of the constraints limiting the use of BEM in practice.

As for the use of calibration tools to improve model believability, the method developed by Cory was widely appreciated as a major improvement to model quality for all practitioners. However, there are major areas which could be improved: automation, visualisation, inter-operability between simulation packages, and the ability to perform parametric sensitivity analyses were the main areas highlighted.

As for other tools, the idea of optimisation and automation is important, though not without the guiding hand of the modeller. All participants advocated for the need to have control over the process, otherwise they tend to get worried by the 'black box' mantra of automated tools such as Apidae. However, the introduction of a self-checking mechanism in Apidae or other future automated tools would help alleviate much of that doubt.

The ability to quantify uncertainties and perform sensitivity analyses are also important factors, however they all rely on the front-end package being intuitive and clear for the user (Berkeley et al., 2015). Quick feedback and reporting are also important aspects which then allow for faster analysis. Also, having models of the existing building would be incredibly helpful to the modeller as they would not have to spend so much time on the mundane but high-risk parts of the modelling procedure.

Automation

A lack of automation was identified as being a major inhibitor to the use of BEM by the four participants. Improved automation would reduce time spent on mundane tasks such as data handling and manipulation. Adding to the level of automation is the use of large scale batch simulations decreasing wait time through the use of cloud computing

Optimisation

Optimise models for a better level of calibration through the use of optimisation tools such as the Apidae calibrator while still allowing the modeller control of what gets optimised. This helps to keep the process clear for the

user. It also reduces the degree to which the optimisation can be called a 'black box'.

Visualisation

Improvements in the way data is presented to the user will help them to quickly identify areas of concern. Providing the user to quickly navigate through the model results will allow them to further isolate incorrect areas.

Interoperability

Increasing the interoperability of calibration tools between BEM simulation software such as IES, EnergyPlus, ESP-r etc. will allow for wide-scale uptake.

6.3.2 Replication of building control

One of the most intriguing findings from the interview process was that all four participants expressed a hesitancy toward modelling HVAC and other system controls in simulation software. One participant even remarked that some software developers did not look to develop the area further:

"I had a conversation with a leading researcher from a respected US institution and said that 'you really need to get the controls right in your simulation program' to which he replied 'no one understands controls so why should we model them correctly?'"

- Participant 1

This represents two failures; firstly, building modeller need to understand the complexities of control systems, and secondly, these systems need to be able to be replicated accurately within simulation software.

Again the first issue here is dealt with largely in the findings (see section 5.4.4). Experience is required to understand how these systems work together and how the building is controlled. For example, imagine a building scenario with two chillers, one that runs efficiently at low loads, and one that runs efficiently at high loads. Rather than having a simple on/off control for a two-chiller cooling system (e.g. both on at 8am, both off at 5pm), a more complex system might be to turn on the first chiller from 8am at high output-load, say 80%, with the aim of cooling the building at the start of the working day. Then later

as the outdoor temperature drops, the first chiller is turned off and the second chiller, which is more efficient at low output-loads is turned on 30%. The BMS synchronises the two systems to balance out and operate more efficiently.

The second issue however, is not dealt with well in the literature. While the understanding of building controls is beyond the scope of this thesis, some light can be shed on the difficulties faced. (Bannister, Thomas, & Lowndes, 2011) compared control systems across different simulation software. They noted four underlying issues:

- Programming languages between simulation software and control systems are quite different, and little guidance is given to help the modeller to reconcile the two.
- Documentation for simulation software is often weak in the area of building control representation (i.e. replicating the control systems in the BEM). This leaves users unsure of the extent to which their simulation software can adequately represent the controls.
- Understanding of HVAC and controls is weak in the simulation community leading to weak representations in the models.
- There are significant differences between how controls are modelled and how they actually function in the real world. Oftentimes simulation control provides an exact calculated response when a change in the control occurs, real life lags behind this significantly.

These points were found by four experienced users of simulation software who were also considered expert in building control, meaning the effects were likely to be worse for less experienced modellers. Significant gaps were found in the areas of chiller representation and supply air temperature, fan and pump control, and the staging of central plant start-up. Findings relating to representing air supply temperature were noted by (Zibin, 2014) who attempted to provide a 'bottom-up' means of calibrating BEMs using BMS trend data.

Further to this is the question of whether simulation software can actually model these control systems accurately at all. Wetter (2012) developed a method of coupling different simulation software (BEM software and dedicated control algorithm software) together to increase functionality, in this case BMS

controls with EnergyPlus via a middleware programme. Though the co-simulation between the two software was possible, it was noted that doing so for a single BEM is not practical or easy, particularly for those unskilled in programming. This further illustrates the gap between control and BEM simulation programming languages.

Whether these constraints have a large outcome on the accuracy of a BEM is as yet unknown, though it is likely to have a significant impact. Building control is largely up to the practitioner to observe, characterise (that is, map out the links between systems), and finally replicate in the model. None of these elements are particularly well documented in literature, though the findings of (Berkeley, 2013) showed that uncertainty around HVAC representation was the most impactful element of a BEM. It follows that the control systems which operate the HVAC plant also need to be modelled with care.

6.4 Validity of results and discussion of the research method

The method used in this research was limiting, but enlightening. While there were several practical limitations, such as the lack of available expert participants, the research method did help to answer the research questions. However, the validity of these results is still difficult to determine.

The validity of the results is usually dependent on two things; the replicability of the results and the size of the sample being used. This would be appropriate for quantitative research, or quantitative analysis of qualitative results. However, this research uses four in-depth interviews with purposively selected experts panning a total of 3-4 hours per interview and workshop. From the size of the population alone it cannot rely on the comfort of generalisability or the weight of a wide range of samples. In-fact the use of the word 'sample' is not wholly appropriate here as the participants were not randomly selected, they were sought.

(Crouch & McKenzie, 2006) discuss that small, exploratory studies, such as this, are conceptually generative. They look to indicate findings rather than conclude facts in the research area. They go on to say that exploratory research such as this does not *have* to stand-up to the same level of speculation as other quantitatively bound research. Instead, it must focus on

rigour in procedure. While this does not amount to validity, they argue that a number of small in-depth interviews can elicit a deep understanding of views, rather than an exhaustive range.

(Golafshani, 2003) presents the case that 'validity' in qualitative studies cannot be pinned down to a single fixed concept. It is generally used as a term to differentiate 'good' and 'bad' research based on the ideas of quality, trustworthiness and rigour which remain defensible. This research remains defensible by being transparent in its process and thorough in its analysis (see section 4.6). However, that does not mean it is not without its limitations.

This research is clearly limited in what could be feasibly gathered, analysed and reported in the time provided given the difficulties in the student workshops. As time was limited, it looked to gather rich information from a very niche group of experts, making data saturation a difficult task. Compounding this is the exploratory and inductive nature of it. To remedy this fault, the structure and composition of the workshop and interview process was designed to gather and link as much information together as is feasible.

7 Conclusions

This research looked to explore the area of Building Energy Modelling and calibration techniques used in industry. Preliminary investigations of the literature and through personal correspondence with EMANZ, EECA, BRANZ and the NZGBC suggested that the problem statement below may be true.

The lack of a usable calibration tools is inhibiting the use of BEM in industry

Thematic analysis of four expert participants in a series of in-depth semi-structured interviews and workshops indicates that yes, the problem statement is valid, though there are other reasons for the lack of uptake in BEM use and calibration tools. The other main reasons are that a credibility gap exists because parameter uncertainty and modeller experience are still major problems when undertaking BEMs. To help alleviate this, calibration tool design needs to focus on keeping the user in control of the process, while still providing enough automation to reduce the cost of calibration.

7.1 A credibility gap exists from the perspective of the modeller

Given the relatively small market and the lack of incentives for building owners to improve the energy efficiency of their buildings, work tends to be limited to simple single-system optimisations. The lack of whole-building retrofits in turn does not require a large modelling industry, which limits the need for modeller upskilling and good connections with current New Zealand and international research.

This puts the practitioner in a difficult position. There is little to no incentive to learn how to model existing buildings, as evidenced by the current small size of

the industry, but there is also a lack of guidance and tools available to them. All four participants believed that every project could benefit in some form from the use of BEM, however currently none required it. While the desire is there (as far as can be seen amongst a sample of just four experts), the undertaking is simply too risky at present. As discussed in section 5.5, a credibility gap exists where the risks of BEM are not balanced by the benefits of using BEM. The risk for professional liability, embarrassment and small fees mean that it is not seen as a first-choice option for existing building energy efficiency retrofits.

7.2 Uncertainty is a key factor

One of the barriers is that modellers are often inexperienced in either their understanding of buildings or their systems, leading to uncertainty in the construction of BEM. This is compounded by the difficulty in getting access to building data required for a BEM and hence leads to uncertainty in the ability to characterise and replicate the HVAC systems in existing buildings. This leads to uncertainty in the modelling process. The amount and breadth of experience a modeller has is likely to have a strong effect on their ability to accurately create a BEM, and limit the uncertainty.

7.3 The need for a semi-automated process

In the process of comparing the two processes; the manual Cory method and the automated Apidae tool, all four participants recognised the need for a quality assured tool that was fast and easy to use. However, there is a balance to be struck between these two aspects.

While Cory method was better understood by the participants, there was difficulty in interpreting the graphs and selecting the input parameters to be chosen. The Apidae tool provided a fast but poorly understood process for calibrating. The ability to select each parameter being tested and identify a range of likely values is of the utmost importance. While it is impossible in complex models to exhaust all possible input parameters and their values, the use of automated sensitivity analysis could help to find the parameters with the greatest impact.

However, scepticism of fully automated processes remains a key issue amongst the four participants interviewed. Findings from this research indicate that the process needs to be controlled by the practitioner from beginning to end, guided by experience and informed decision making. A manual checking process like the Cory method could be used to quality assure the outputs of automated calibration techniques.

Overall, the adoption of the Cory tool would be expected to improve the outcomes of BEM in practice.

7.4 Overcoming barriers to BEM calibration uptake

Increase the value-case for BEM in New Zealand

As found in the literature review (see chapter2) the introduction of the NABERS rating scheme has greatly improved the value and quality of the commercial building stock in Australia. While the scheme has been adopted in New Zealand, uptake is still slow. The value case for whole building energy retrofits needs to be incentivised and increased over simple system optimisations. These have the ability to increase savings to decrease GHG and improve the quality of the building stock.

Introduction of more guidelines and protocols outlining how to calibrate and recognise uncertainties

As noted in section 2.7, the current state of industry guidelines and protocols for existing buildings is limited. An increase in the availability of descriptive guidelines and calibration methods is needed to improve the BEM credibility gap.

Research into the impacts of uncertainty quantification specific to the New Zealand climate and building stock are also needed. This will help to decrease the level of modelling uncertainty for practitioners. Writers of simulation guidelines and researchers must co-operate so that the findings of research are transferred to practice.

Increase in education, training and incentives for practitioners

Currently very few practitioners in New Zealand are capable of performing complex BEMs for existing buildings. This is evidenced by the small population of professionals shortlisted to take part in this research.

Education, training and incentives are needed to help close the credibility gap. However, this is predicated on an increase in the value of BEMs in New Zealand. An accreditation scheme for modellers may help to increase trustworthiness of BEM, although the size of the NZ market suggests a specialist qualification is unlikely to be self-supporting. The modeller is currently constrained by the perceived lack of value of BEM resulting in their own lack of upskilling in the area and keeping connected to literature. The lack of industry knowledge is a large issue, but can be managed with education and training.

Introduce guides on different calibration techniques available and their advantages and disadvantages

BEM quality assurance techniques are not sophisticated or commonplace in firms. Major gaps in understanding, disconnection with literature and the application of research to the real world are lagging rather than leading at present. Educating practitioners about specific calibration metrics, techniques and frequently encountered obstacles is necessary to empower them to take control of the process and to suggest BEM as a legitimate means of energy saving prediction.

7.5 Future research

The BEM and commercial building industries would benefit greatly if research was further developed into usable and workable tools. As a result of this work, the following areas are proposed as candidates for further research:

- A strict comparison using real building data between calibration methods described by Coakley et al. (2014) in a similar vein to that of (Robertson et al., 2013). This may help to establish if the calibration techniques developed in industry are equal, or if they even produce the same results. A systematic sensitivity analysis of both automated and

manual systems could help determine if the same uncertain parameters are identified by each method.

- Undertaking this research with a wider participant sample. This research relied on expert interviews to get a more in-depth understanding of how BEM practices are employed in New Zealand industry. Undertaking a comparative analysis between expert and novice (e.g. recent graduate) modellers could explore the findings of Berkeley (2013)
- Additionally, undertaking the same approach as Berkeley (2013), but on the ability of professional participants to undertake a calibration process on the same building could help explore the decision making process and how this effects the outcomes of systematic calibration approaches.
- Determining the impact of misrepresentation of building controls on the results of BEM. This area requires the expertise of someone familiar with both energy modelling and control systems to help determine how big an impact incorrectly modelling control has, and to see if the level of calibration would change as a result of it.

8 References

- Alajmi, A. (2012). Energy audit of an educational building in a hot summer climate. *Energy and Buildings*, 47, 122–130. <https://doi.org/10.1016/j.enbuild.2011.11.033>
- Al-Homoud, M. S. (2001). Computer-aided building energy analysis techniques. *Building and Environment*, 36(4), 421–433. [https://doi.org/10.1016/S0360-1323\(00\)00026-3](https://doi.org/10.1016/S0360-1323(00)00026-3)
- Amitrano, L., Isaacs, N., Saville-Smith, K., Donn, M., Camilleri, M., Pollard, A., ... Cory, S. (2014). *BEES part1:Final Report* (Building Energy End-use Study No. SR297/1). Judgeford: BRANZ. Retrieved from http://www.branz.co.nz/cms_show_download.php?id=182d3ae296d4d92be905477582ae2453c0671d58
- Ascione, F., Bianco, N., De Masi, R. F., De'Rossi, F., & Vanoli, G. P. (2015). Energy retrofit of an educational building in the ancient center of Benevento. Feasibility study of energy savings and respect of the historical value. *Energy and Buildings*, 95, 172–183. <https://doi.org/10.1016/j.enbuild.2014.10.072>
- ASHRAE. (2002). *ASHRAE guideline 14 2002: Measurement of Energy and Demand Saving*. Atlanta: ASHRAE. Retrieved from https://gaia.lbl.gov/people/ryin/public/Ashrae_guideline14-2002_Measurement%20of%20Energy%20and%20Demand%20Saving%20.pdf
- ASHRAE. (2006). ANSI/ASHRAE/IES Standard 100-2006 Energy Efficiency in Existing Buildings. ASHRAE. Retrieved from <https://www.ashrae.org/resources--publications/bookstore/standard-100>

- ASHRAE. (2013). ANSI/ASHRAE/IES Standard 90.1-2013 -- Energy Standard for Buildings Except Low-Rise Residential Buildings. American Society Heating Refrigeration Air conditioning Engineers. Retrieved from <https://www.ashrae.org/resources--publications/bookstore/standard-90-1>
- ASHRAE. (2016, March). ASHRAE Standard 209P Energy Simulation Aided Design for buildings except Low-rise Residential Buildings (Under Public review). American Society Heating Refrigeration Air conditioning Engineers.
- Assistant Secretary for Public Affairs. (2013, September 6). System Usability Scale (SUS). Retrieved 10 August 2015, from <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>
- Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative Research*, 1(3), 385–405.
- Augenbroe, G. (2002). Trends in building simulation. *Building and Environment*, 37(8), 891–902.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6), 574–594.
- Bannister, P. (2005). The ABGR validation protocol for computer simulations. Retrieved from http://www.inive.org/members_area/medias/pdf/inive/ibpsa/bs05_0033_40.pdf
- Bannister, P. (2011). NABERS: Lessons from 12 Years of Performance Based Ratings in Australia. Exergy Australia. Retrieved from [file:///C:/Users/user/Downloads/43%20NABERS%2012%20years%20paper%20-%20final%20\(1\).pdf](file:///C:/Users/user/Downloads/43%20NABERS%2012%20years%20paper%20-%20final%20(1).pdf)
- Bannister, P., Thomas, P. C., & Lowndes, P. (2011). Representation of HVAC control in common simulation packages. In *Proceedings of the 12th IBPSA Building Simulation Conference* (pp. 70–77). Sydney. Retrieved from http://www.ibpsa.org/proceedings/BS2011/P_1138.pdf
- Bensouda, N. (2004, November 15). *Extending and formalizing the energy signature method for calibrating simulations and illustrating with application for three*

- California climates*. Texas A&M University. Retrieved from <https://repository.tamu.edu/handle/1969.1/1080>
- Berkeley, P. (2013, January 1). *Method for Testing and Classifying the Effect of the Modeler on Building Energy Simulation Results* (PhD). University of California. Retrieved from <http://escholarship.org/uc/item/5dn3r716>
- Berkeley, P., Haves, P., & Kolderup, E. (2014). Impact of Modeller Decisions on Simulation Results (pp. 203–210). Presented at the ASHRAE/IBPSA - USA Building Simulation, Atlanta. Retrieved from https://www.ashrae.net/File%20Library/docLib/Events/ASHRAE-IPBSA-USA/Presentations/26_Berkeley.pdf
- Berkeley, P., Haves, P., & Kolderup, E. (2015). The Effect of Modeller Decisions on Simulation Uncertainty: Some Implications for User Interface Design. In *BS2015*. Hyderabad: IBPSA. Retrieved from <http://www.ibpsa.org/proceedings/BS2015/p2739.pdf>
- Bernard, H. R. (2011). *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman Altamira. Retrieved from <https://books.google.co.nz/books?hl=en&lr=&id=WhKYqATAySwC&oi=fnd&pg=PR1&dq=Research+Methods+in+Anthropology:+Qualitative+and+quantitative+methods&ots=6zxhAyPWwG&sig=ghCowkPuvA97wNVaOyP-hu3P2Q0>
- Bernard, H. R., & Ryan, G. W. (2009). *Analyzing qualitative data: Systematic approaches*. SAGE publications. Retrieved from <https://books.google.co.nz/books?hl=en&lr=&id=dzwXBAAQBAJ&oi=fnd&pg=PR17&dq=analyzing+qualitative+data+bernard&ots=RKtU2HxY1U&sig=sDTgzqrlidXuJ-pLx01cb7OYxniU>
- Bogner, A., Littig, B., & Menz, W. (2009). *Interviewing experts*. Palgrave Macmillan Basingstoke. Retrieved from <http://library.wur.nl/WebQuery/clc/2060884>
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.

- Bradley, D. E., Kummert, M., & McDowell, T. P. (2004). Experiences with and interpretation of standard test methods of building energy analysis tools. Retrieved from <http://strathprints.strath.ac.uk/6582/>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4–7.
- Carabott, M. (1989). Building simulation reconciliation using empirical data. In *Proceedings of Building Simulation* (Vol. 89, pp. 23–24).
- Carroll, W. L., & Hitchcock, R. J. (1993). Tuning simulated building descriptions to match actual utility data: methods and implementation. *ASHRAE Transactions-American Society of Heating Refrigerating Airconditioning Engin*, 99(2), 928–934.
- CIBSE. (2015, November). AM11 Building Performance Modelling 2015. Chartered Institution of Building Services Engineers. Retrieved from <http://www.cibse.org/knowledge/cibse-am/am11-building-performance-modelling-new-2015>
- Clarke, J. A. (2001). *Energy simulation in building design*. Routledge. Retrieved from <http://books.google.co.nz/books?hl=en&lr=&id=WH0VCiF8jkoC&oi=fnd&pg=PR3&dq=J.+Clarke+Energy+Simulation+in+Building+Design&ots=c8ZAf5vjty&sig=LE3snFMoADTJqA6dmSvFqiYiFQM>
- Clarke, J. A., Strachan, P. A., & Pernot, C. (1993). An approach to the calibration of building energy simulation models. *TRANSACTIONS-AMERICAN SOCIETY OF HEATING REFRIGERATING AND AIR CONDITIONING ENGINEERS*, 99, 917–917.
- Coakley, D. (2014). Calibration of detailed building energy simulation models to measured data using uncertainty analysis.
- Coakley, D., Raftery, P., & Keane, M. (2014). A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, 37, 123–141. <https://doi.org/10.1016/j.rser.2014.05.007>

- Cory, S. (2015). *An Exploration of the Feasibility of Converting the New Zealand Commercial Building Stock to be Net Zero Energy* (PhD). Victoria University of Wellington, Wellington.
- Cory, S., Gates, A., & Donn, M. (2011). The creation of Generic Energy Simulation Models Which Represent Typical Commercial Buildings and their Calibration Against Real Data (pp. 1473–1480). Presented at the 12th Conference of International Building Performance Simulation Association, Sydney: Victoria University of Wellington. Retrieved from http://ibpsa.org/proceedings/BS2011/P_1492.pdf
- Crawely, D., Hand, J., Kummert, M., & Griffith, B. (2008). Contrasting the Capabilities of Building Energy Performance Simulation Programs. *Building and Environment*, 43(4), 661–673.
- Crouch, M., & McKenzie, H. (2006). The logic of small samples in interview-based qualitative research. *Social Science Information*, 45(4), 483–499. <https://doi.org/10.1177/0539018406069584>
- Daly, D., Cooper, P., & Ma, Z. (2014). Understanding the risks and uncertainties introduced by common assumptions in energy simulations for Australian commercial buildings. *Energy and Buildings*, 75, 382–393. <https://doi.org/10.1016/j.enbuild.2014.02.028>
- Diamond, S. C., & Hunn, B. D. (1981). *Comparison of DOE-2 computer program simulations to metered data for seven commercial buildings*. Los Alamos Scientific Lab., NM (USA). Retrieved from <http://www.osti.gov/scitech/biblio/7064521>
- EECA. (2012). NABERSNZ. Retrieved 26 April 2015, from <http://www.nabersnz.govt.nz/about-nabersnz>
- EECA. (2015a). Commercial buildings programme. Retrieved 29 April 2015, from <http://www.eecabusiness.govt.nz/services-and-funding/commercial-buildings>
- EECA. (2015b). Energy use in your building. Retrieved 24 June 2015, from <http://www.eecabusiness.govt.nz/how-to-be-energy-efficient/managing-premises>
- EECA. (2016a, April). Energy audits. Retrieved 30 May 2016, from <https://www.eecabusiness.govt.nz/funding-and-support/energy-audits/>

- EECA. (2016b, April). Monitoring and targeting. Retrieved 30 May 2016, from <https://www.eecabusiness.govt.nz/funding-and-support/monitoring-and-targeting/>
- EECA. (2016c, April). Systems optimisation. Retrieved 30 May 2016, from <https://www.eecabusiness.govt.nz/funding-and-support/systems-optimisation/>
- EMANZ. (2015). What is an Energy Audit? Retrieved 24 June 2015, from <http://www.emanz.org.nz/energy-management-audits/what-energy-audit>
- Fabrizio, E., & Monetti, V. (2015). Methodologies and Advancements in the Calibration of Building Energy Models. *Energies*, 8(4), 2548–2574. <https://doi.org/10.3390/en8042548>
- Few, S. (2006). *Information dashboard design*. O'Reilly. Retrieved from https://www.thali.ch/files/Shop/Documents/018161_Chapter_1_Clarifying_The_Vision.pdf
- Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: strategies for qualitative research. Retrieved from <http://library.wur.nl/WebQuery/clc/437460>
- Glaser, B., & Strauss, A. (1967). Applying grounded theory. *The Discovery of Grounded Theory: Strategies of Qualitative Research*. Hawthorne, NY: Aldine Publishing Company, 237–51.
- Goddard, W., & Melville, S. (2004). *Research Methodology: An Introduction*. Juta and Company Ltd.
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597–606.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of Qualitative Research*, 2(163–194), 105.
- Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. *Field Methods*, 18(1), 59–82. <https://doi.org/10.1177/1525822X05279903>

- Guyon, G. (1997). Role of the model user in results obtained from simulation software program. In *Proc. Building Simulation* (Vol. 97). Retrieved from http://www.inive.org/members_area/medias/pdf/Inive%5CIBPSA%5CUFSC609.pdf
- Heidell, J. A., & Taylor, Z. T. (1985). *Comparison of empirically measured end-use metered data with DOE 2. 1 simulation*. Pacific Northwest Lab., Richland, WA (USA). Retrieved from <http://www.osti.gov/scitech/biblio/5746470>
- Henninger, R. H., Witte, M. J., & Crawley, D. B. (2003). Experience testing Energy Plus with the IEA HVAC BESTEST E100 E200 Series (pp. 467–474). Presented at the Eighth International IBPSA Conference, Eindhoven, Netherlands: GARD Analytics, Inc. Retrieved from http://apps1.eere.energy.gov/buildings/energyplus/pdfs/bibliography/bs03_henninger_0467_474.pdf
- Hills, A. J. (2014). *Visualising the Invisible: Displaying Building Resource-use Benchmarks in a 3D City Visualisation* (Masters thesis). Victoria University of Wellington, Wellington. Retrieved from <http://researcharchive.vuw.ac.nz/handle/10063/4396>
- Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), 79–102.
- Hornberger, G. M., & Spear, R. C. (1981). Approach to the preliminary analysis of environmental systems. *J. Environ. Manage.; (United States)*, 12(1). Retrieved from <http://www.osti.gov/scitech/biblio/6396608>
- Ibarra, D., & Reinhart, C. F. (2009). Daylight factor simulations—how close do simulation beginners ‘really’ get? In *Building Simulation* (Vol. 196, p. 203). Retrieved from http://web.mit.edu/tito_/www/Publications/BS09_DaylightingNovices.pdf
- International Organization for Standardization. (2014). ISO 50002:2014 Energy audits - requirements with guidance for use. International Organization for Standardization. Retrieved from <https://www.iso.org/obp/ui/#iso:std:iso:50002:ed-1:v1:en>
- International Standards Authority. (1998). ISO 9241-11:1998 - Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability.

- International Standards Authority. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883
- Investment Property Databank Ltd. (2013). IPD Australia Green Investment Property Index: June 2013 Headline Results. Investment Property Databank Ltd. Retrieved from <http://www.nabers.gov.au/public/WebPages/DocumentHandler.ashx?docType=3&id=98&attId=0>
- IPCC. (2007). Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (R. K. Pachauri, A. Reisinger, & L. A. Meyer, Eds.). IPCC. Retrieved from http://www.researchgate.net/profile/Hans_Otto_Poertner2/publication/272784739_Synthesis_Report_IPCC_AR5/links/54edbe2c0cf272fc6d224447.pdf
- IPMVP Committee. (2001). *International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings, Volume I*. Colorado: National Renewable Energy Lab., Golden, CO (US). Retrieved from <http://www.osti.gov/scitech/biblio/776003>
- Jaffe, A. B., & Stavins, R. N. (1994). The energy-efficiency gap - What does it mean? *Energy Policy*, 22(10), 804–810. [https://doi.org/10.1016/0301-4215\(94\)90138-4](https://doi.org/10.1016/0301-4215(94)90138-4)
- Johannessen, G. H. J., & Hornbæk, K. (2014). Must evaluation methods be about usability? Devising and assessing the utility inspection method. *Behaviour & Information Technology*, 33(2), 195–206. <https://doi.org/10.1080/0144929X.2012.751708>
- Kandil, A.-E., & Love, J. A. (2014). Signature analysis calibration of a school energy model using hourly data. *Journal of Building Performance Simulation*, 7(5), 326–345. <https://doi.org/10.1080/19401493.2013.838608>
- Lewis, C., Polson, P. G., Wharton, C., & Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 235–242). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=97279>

- Liu, M., Wei, G., & Claridge, D. E. (1998). Calibrating AHU models using whole building cooling and heating energy consumption data. *Proc. of ACEEE 1998 Summer Study on Energy Efficiency in Buildings*, 3, 229–241.
- Ma, Z., Cooper, P., Daly, D., & Ledo, L. (2012). Existing building retrofits: Methodology and state-of-the-art. *Energy and Buildings*, 55, 889–902. <https://doi.org/10.1016/j.enbuild.2012.08.018>
- Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? A review of qualitative interviews in IS research. *Journal of Computer Information Systems*, 54(1), 11–22.
- Newell, G., MacFarlane, J., & Kok, N. (2011). Building Better Returns. Australian Property Institute. Retrieved from http://www.api.org.au/assets/media_library/000/000/219/original.pdf
- Nexant Inc. (2008). *M&V Guidelines: Measurement and Verification for Federal Energy Projects Version 3.0*. Boulder Colorado: U.S. Department of Energy. Retrieved from www1.eere.energy.gov/femp/.
- Nielsen, J. (1994). *Usability engineering*. Elsevier. Retrieved from <https://books.google.co.nz/books?hl=en&lr=&id=DBOowF7LqIQC&oi=fnd&pg=PP1&dq=Jakob+Nielsen+1994&ots=Bk38YOMQyV&sig=vWmF8uZyAdmiFzIf6BrVVgOCVkc>
- NSW Office of Environment and Heritage. (2011). Energy Guide to Building Energy Estimation. Office of Environment and Heritage. Retrieved from <http://www.nabers.gov.au/public/WebPages/DocumentHandler.ashx?docType=2&id=165&attId=0>
- Office of Environment and Heritage. (2014a). Environmental Benefit. Retrieved 26 April 2015, from <http://www.nabers.gov.au/AnnualReport/environmental-benefit.html>
- Office of Environment and Heritage. (2014b). The National Australian Built Environment Rating System (NABERS) | NSW Environment & Heritage. Retrieved 5 October 2014, from <http://www.environment.nsw.gov.au/business/nabers.htm>

- Olgyay, V., & Seruto, C. (2010). Whole-building retrofits: A gateway to climate stabilization. *ASHRAE Transactions*, 116(2), 1–8.
- O'Reilly, M., & Parker, N. (2013). 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research*, 13(2), 190–197. <https://doi.org/10.1177/1468794112446106>
- Popescu, D., Bienert, S., Schützenhofer, C., & Boazu, R. (2012). Impact of energy efficiency measures on the economic value of buildings. *Applied Energy*, 89(1), 454–463. <https://doi.org/10.1016/j.apenergy.2011.08.015>
- Raftery, P. (2011, January). *Calibrated whole building energy simulation: An evidence-based methodology* (Thesis). Retrieved from <http://aran.library.nuigalway.ie/xmlui/handle/10379/1795>
- Raftery, P., Keane, M., & Costa, A. (2009). Calibration of a detailed simulation model to energy monitoring system data: a methodology and case study. In *Eleventh International IBPSA Conference*. Citeseer. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.299.1209&rep=rep1&type=pdf>
- Reddy, A. T. (2006). Literature review on calibration of building energy simulation programs: Uses, problems, procedures, uncertainty, and tools. *ASHRAE Transactions*, 226–240.
- Reddy, T. A., Maor, I., & Panjapornpon, C. (2007a). Calibrating Detailed Building Energy Simulation Programs with Measured Data—Part I: General Methodology (RP-1051). *HVAC&R Research*, 13(2), 221–241. <https://doi.org/10.1080/10789669.2007.10390952>
- Reddy, T. A., Maor, I., & Panjapornpon, C. (2007b). Calibrating Detailed Building Energy Simulation Programs with Measured Data—Part II: Application to Three Case Study Office Buildings (RP-1051). *HVAC&R Research*, 13(2), 243–265. <https://doi.org/10.1080/10789669.2007.10390953>

- Robertson, J., Polly, B., & Collis, J. (2013). Evaluation of automated model calibration techniques for residential building energy simulation. *National Renewable Energy Laboratory*. Retrieved from <http://www.nrel.gov/docs/fy13osti/60127.pdf>
- Roth, K. W., Westphalen, D., Feng, M. Y., Llana, P., & Quartararo, L. (2005). *Energy impact of commercial building controls and performance diagnostics: market characterization, energy impact of building faults and energy savings potential*.
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage. Retrieved from <https://books.google.co.nz/books?hl=en&lr=&id=jh1iCgAAQBAJ&oi=fnd&pg=PT13&dq=the+coding+manual+for+saldana&ots=OuBwYnZdpN&sig=V2uEtJN1fWQmo6tILTp1jrSkvc>
- Seidler, J. (1974). On Using Informants: A Technique for Collecting Quantitative Data and Controlling Measurement Error in Organization Analysis. *American Sociological Review*, 39(6), 816–831. <https://doi.org/10.2307/2094155>
- Shackel, B., & Richardson, S. J. (1991). *Human Factors for Informatics Usability*. Cambridge University Press.
- Shapiro, I. (2011). 10 Common problems in energy audits. *ASHRAE Journal*, 53(2), 26–31.
- Socolow, R. H., & Sondregger, R. C. (1976). *The Twin Rivers program on energy conservation in housing: Highlights and conclusions* (No. 32) (pp. 207–242). Princeton New Jersey: Princeton: Center for Environmental Studies. Retrieved from <http://www.sciencedirect.com/science/article/pii/0378778878900038>
- Standards New Zealand. (2000). AS/NZS 3598:2000. SAI Global Limited.
- Standards New Zealand. (2014). AS/NZS 3598.1:2014 Energy Audits Part1: Commercial Buildings. SAI Global Limited. Retrieved from <https://shop.standards.govt.nz/catalog/3598.1:2014%28AS|NZS%29/scope>
- Thom, H. C. S. (1954). The rational relationship between heating degree days and temperature. *Monthly Weather Review*, 82(1), 1–6. [https://doi.org/10.1175/1520-0493\(1954\)082<0001:TRRBHD>2.0.CO;2](https://doi.org/10.1175/1520-0493(1954)082<0001:TRRBHD>2.0.CO;2)

- Tupper, K., Franconi, E., Chan, C., Fluhrer, C., Jenkins, M., & Hodgins, S. (2011). Pre-Read for BEM innovation summit. *Retrieved from Rocky Mountain Institute Website: [http://www.rmi.org/content/files/Summit_PreRead_Apr-19-2011_\(2\).pdf](http://www.rmi.org/content/files/Summit_PreRead_Apr-19-2011_(2).pdf)*
- Tupper, K., Franconi, E., Chan, C., Hodgins, S., Buys, A., & Jenkins, M. (2011). Building energy modeling: industry-wide issues and potential solutions. In *Proceedings of the 12th Conference of International Building Performance Simulation Association*. Retrieved from http://ibpsa.org/proceedings/BS2011/P_1317.pdf
- United Nations. (2015, December). Adoption of the Paris Agreement: Proposal by the president. United Nations. Retrieved from <https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf>
- U.S. Department Of Energy. (2015). Building Technologies Office: Building Energy Software Tools Directory. Retrieved 29 June 2015, from http://apps1.eere.energy.gov/buildings/tools_directory/
- Waltz, J. P. (2000). *Computerized building energy simulation handbook*. CRC Press. Retrieved from https://books.google.co.nz/books?hl=en&lr=&id=HzcbrR_2JX8C&oi=fnd&pg=PR5&dq=Waltz+james+computerized+building+energy+simulation+handbook&ots=aeO46z383C&sig=vlGJ1L3qDhC6b_u58brCSZIHagg
- Wei, G., Liu, M., & Claridge, D. E. (1998). Signatures of heating and cooling energy consumption for typical AHUs.
- Wetter, M. (2012). Co-Simulation of Building Energy and Control Systems with the Building Controls Virtual Test Bed. *Journal of Building Performance Simulation*, 4, 3, 185-203. Retrieved from <http://escholarship.org/uc/item/7cp597jf>
- Wilkinson, S. (1998). Focus group methodology: A review. *International Journal of Social Research Methodology*, 1(3), 181–203.
- Zhao, H., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6), 3586–3592. <https://doi.org/10.1016/j.rser.2012.02.049>

Zibin, N. (2014, December 18). *A Bottom-Up Method to Calibrate Building Energy Models Using Building Automation System (BAS) Trend Data* (masters). Concordia University. Retrieved from <http://spectrum.library.concordia.ca/979581/>

9 Appendices

9.1 Appendix A - Interview guide

9.1.1.1 Interview Guide - Pre workshop

Background and practice

1. Please tell me a little about your background and role as it relates to commercial buildings and energy. Company name & position
2. Would you clarify for me your qualifications as they relate to building energy use?
3. With respect to buildings and how they use energy, would you say you tend to work more with new or existing buildings?
 - a. What proportion of your time would you say is spent looking at existing building energy use?
 - b. What type of building do you mostly work with?
Energy Efficiency Assessments

Energy efficiency assessments

4. What type of audit is common? Type 1, 2 or 3 (or not standardised)
 - a. In your opinion, which audit type would allow you to collect sufficient information to produce a Building Energy Model?
5. Please describe for me your standard process for calculating and recommending Energy Conservation Measures for a large commercial building.
How do you Quality Assure this process, so that the energy savings and other benefits are likely to occur?

Computer Modelling and software

6. What proportion of your projects require Building Energy Modelling?
7. What proportion of your projects would you suggest could benefit from building energy modelling?
 - a. During a project, how much time is spent creating a Building Energy Model?

8. Briefly, could you please describe the general process you use to model existing buildings?
9. How detailed would you consider a typical model needs to be? Does the level of detail change, and if so when and why?
10. In your experience, what are the major limitations to energy simulation or modelling of buildings?
11. What quality assurance tests do you undertake when modelling? How do you check your model is reliable or valid?

Calibration

12. To you, what does it mean to calibrate a Building Energy Model?
 - a. How do you determine whether or not a model is calibrated?
13. ASHRAE guideline 14:2002 provide a tolerance level beyond which a model cannot be considered calibrated. How would you usually go about correcting a model that fell outside of this range?
14. For what purposes might it be necessary to calibrate a model?

9.1.1.2 Interview Guide - Post-workshop

There are three key sections of questioning – usability, modeller bias and verification. Each tool will have the same set of questions.

Tool Impressions

1. What are your impressions of the tool itself?
 - a. What do you feel were its strengths?
 - b. What do you feel were its weaknesses?
2. Which specific elements were confusing or difficult to understand?
 - a. Through the use of the tool, how easily were you able to identify poor calibration, and what, if any, were these?
3. What usability issues could be improved? *E.g. increased Automation, clarity, visual display.*

The role of the user

1. Do you feel that the adoption of this tool would improve the reliability of your modelling?
2. Do you feel that the modeller's bias and ability to unduly tinker with the model are sufficiently limited with this process?
 - a. Do you feel that the process still provides sufficient freedom for the modeller?

Tool uses in current practice

1. Are there areas where you feel model calibration is too limited should **not** be used in building energy assessment?

2. Can you suggest where the calibration process be most useful?
3. What are the benefits or limitations in the use of templates in modelling energy saving opportunities in real buildings?
4. Do you feel that uptake of the calibration process would help you to make more informed decisions when suggesting energy saving opportunities?

Other tools for calibration

1. Are you aware of any other tools that look to aid modellers in creating calibrated models?
2. What are the barriers that you feel might prevent uptake of these types of tools?
3. What aspects of these tools will need to change/be improved in order to increase uptake?

Requirements for uptake

1. What are the current barriers of use against Building Energy Model calibration in Industry?
2. What should be the main focus for the uptake of calibration tools for industry?

9.2 Appendix B – Ethic committee documents

9.2.1 Student participant information and consent sheets



An exploration of simple building energy model calibration in New Zealand Industry

INFORMATION SHEET FOR PARTICIPANTS

Thank you for your interest in this project. Please read this information before deciding whether or not to take part. If you decide to participate, thank you. If you decide not to take part, thank you for considering my request.

Who am I?

My name is Ethan Duff and I am a Masters student in *the* Building Science Programme at Victoria University of Wellington. This research project is work towards my Master's thesis

What is the aim of the project?

The research explores the usability of a Building Energy Modelling (BEM) tool for use in industry. This tool will be used to help calibrate or match energy models of buildings against their measured energy data. I.e. matching simulated data with actual measured data.

This part of the research looks to see what flaws exist within the current tool, before a more refined version is presented to industry professionals in workshops at the end of the year.

The main objective is to determine how the use of this tool might benefit industry in assessing building energy performance and for suggesting energy saving measures.

This research has been approved by the Victoria University of Wellington Human Ethics Committee [provide approval number].

How can you help?

You are invited to take part in a usability study of a Building energy tool which is taught as part of the 2nd assignment in the *BILD 423 Buildings and Energy* paper.

The research will be conducted during class time and will require you, as a participant, to be observed as you work to complete the assignment. The results of these observations will be used to streamline this process into a more user-friendly tool.

Occasional questions may be asked of you as you complete the assignment work. It is anticipated that these questions will take up to 3 minutes to answer in class. These questions are:

- *At your current stage, what aspect of the process do you find difficult to use or understand and why is this so?*
- *In what way(s) do you think this could be made easier or more user friendly?*

Additional information may be sought to ensure that your answers are clearly understood.

You can stop your participation at any time, without giving a reason. You can withdraw from the study by October 5th 2015. If you withdraw, the information you provided will be destroyed or returned to you.

What will happen to the information you give?

This research is confidential. I will not name you in any reports, and I will not include any information that would identify you. Only my supervisors and I will read the notes or transcript of the interview. The interview transcripts, summaries and any recordings will be kept securely and destroyed 1 year after the research ends.

You will only be identified in the final research as a student participant i.e. Participant A

What will the project produce?

The information from my research will be used in my Master's thesis. You will not be identified in my report. I may also use the results of my research for conference presentations, and academic reports. I will take care not to identify you in any presentation or report.

If you accept this invitation, what are your rights as a research participant?

You do not have to accept this invitation if you don't want to. If you do decide to participate, you have the right to:

- choose not to answer any question;
- withdraw from the study by October 5th 2015;
- ask any questions about the study at any time;
- read over and comment on a written summary of your interview;
- be able to read any reports of this research by emailing the researcher to request a copy.

If you have any questions or problems, who can you contact?

If you have any questions, either now or in the future, please feel free to contact either:

Student/Researcher:

Name: Ethan Duff

email: duffetha@myvuw.ac.nz

Supervisor:

Name: Nigel Isaacs

Role: Senior Lecturer

School: School of Architecture

Phone: (04) 463 9745

nigel.isaacs@vuw.ac.nz

Human Ethics Committee information

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convener: Associate Professor Susan Corbett. Email susan.corbett@vuw.ac.nz or telephone +64-4-463 5480.



An exploration of simple building energy model calibration in New Zealand Industry

CONSENT TO INTERVIEW

This consent form will be held for 5 years.

Researcher: Ethan Duff, School of Architecture and Design, Victoria University of Wellington

- I have read the Information Sheet and the project has been explained to me. My questions have been answered to my satisfaction. I understand that I can ask further questions at any time.
- I agree to take part in the research, which requires me to be observed in class, during the latter half of Assignment 2 of the BILD 423 Buildings and Energy course. I understand this will involve in-class probing questions as explained in the participant information sheet.

I understand that:

- I may withdraw from this study up to four weeks after the interview October 5th 2015, and any information that I have provided will be returned to me or destroyed.
- The information I have provided will be destroyed 1 year after the research is finished.
- Any information I provide will be kept confidential to the researcher and the supervisor. I understand that the results will be used for a Master's Thesis report and a summary of the results may be used in academic reports and/or presented at conferences.
- My name will not be used in reports, nor will any information that would identify me.
- I would like a copy of the transcript of my in-class answers: Yes ☐ No ☐
- I would like a summary of my in-class answers: Yes ☐ No ☐
- I would like to receive a copy of the final report and have added my email address below. Yes ☐ No ☐

Signature of participant: _____

Name of participant: _____

Date: _____

Contact details: _____

9.2.2 Industry participant information and consent sheets



Usability & Utility: An exploration of simple building energy model calibration in New Zealand Industry

INFORMATION SHEET FOR INDUSTRY PARTICIPANTS

Thank you for your interest in this project. Please read this information before deciding whether or not to take part. If you decide to participate, thank you. If you decide not to take part, thank you for considering my request.

Who am I?

My name is Ethan Duff and I am a Masters student in the Building Science Programme at Victoria University of Wellington. This research project is work towards my Master's thesis

What is the aim of the project?

The research explores the usability of two Building Energy Modelling (BEM) tools for use in industry. These tools are used to help calibrate or match energy models of buildings against their measured energy that is matching simulated data with actual metered data.

This part of the research requests your participation in a small workshop and interview to test the usability of these tools and to provide commentary on their usefulness and utility in practice.

The main objective of this research is to determine how the use of these tools might benefit industry in assessing building energy performance and for suggesting energy saving measures.

This research has been approved by the Victoria University of Wellington Human Ethics Committee (Ethics approval number 22390).

How can you help?

You are invited to take part in a usability study of a Building Energy Modelling tool in a small workshop held between Monday 22nd February and Friday 4th March 2016.

The research will be conducted at the School of Architecture (130 Vivian St Te Aro, Wellington) and will require you, as a participant, to be observed as you work through a series of tasks using the tool. After the workshop, you will be asked to participate in a brief interview to discuss the strengths, weaknesses and challenges of the tools. You will also be asked about the role of modelling in your professional work and your attitudes toward energy modelling in practice. This is expected to take no more than 3 ½ hours of an afternoon suitable to you.

You can stop your participation at any time, without giving a reason. You can withdraw from the study by March 31st 2016. If you withdraw, the information you provided will be destroyed or returned to you.

What will happen to the information you give?

This research is confidential. I will not name you in any reports, and I will not include any information that would identify you. Only my supervisors and I will read the notes or transcript of the interview. The interview transcripts, summaries and any recordings will be kept securely and destroyed 1 year after the research ends. You will only be identified in the final research as an industry participant e.g. Industry participant A

What will the project produce?

The information from my research will be used in my Master's thesis. I may also use the results of my research for conference presentations, and academic reports.

If you accept this invitation, what are your rights as a research participant?

You do not have to accept this invitation if you don't want to. If you do decide to participate, you have the right to:

- choose not to answer any question;
- withdraw from the study by March 31st 2016;
- ask any questions about the study at any time;
- read over and comment on a written summary of your interview;
- be able to read any reports of this research by emailing the researcher to request a copy.

If you have any questions or problems, who can you contact?

If you have any questions, either now or in the future, please feel free to contact either:

Student/Researcher:

Name: Ethan Duff
 School: School of Architecture
 Mobile: 027 740 1384
 Email: duffetha@myvuw.ac.nz

Supervisor:

Name: Nigel Isaacs
 School: School of Architecture
 Phone: (04) 463 9745
 Email nigel.isaacs@vuw.ac.nz

Human Ethics Committee information

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convener: Associate Professor Susan Corbett. Email susan.corbett@vuw.ac.nz or telephone +64-4-463 5480.



Usability & Utility: An exploration of simple building energy model calibration in New Zealand Industry

CONSENT TO INTERVIEW AND WORKSHOP

This consent form will be held for 5 years.

Researcher: Ethan Duff, School of Architecture and Design, Victoria University of Wellington

- I have read the Information Sheet and the project has been explained to me. My questions have been answered to my satisfaction. I understand that I can ask further questions at any time.
- I agree to take part in the research, which requires my participation in a small workshop and interview. I understand this will involve in-class probing questions as explained in the participant information sheet.

I understand that:

- I may withdraw from this study after March 31st 2016, and any information that I have provided will be returned to me or destroyed.
- The information I have provided will be destroyed 1 year after the research is finished.
- Any information I provide will be kept confidential to the researcher and the supervisor. I understand that the results will be used for a Master's Thesis report and a summary of the results may be used in academic reports and/or presented at conferences.
- My name will not be used in reports, nor will any information that would identify me.
- I would like a copy of the transcript of my interview answers: Yes ☐ No ☐
- I would like a summary of my interview answers: Yes ☐ No ☐
- I would like to receive a copy of the final report and have added my email address below. Yes ☐ No ☐

Signature of participant: _____

Name of participant: _____

Date: _____

Contact details: _____