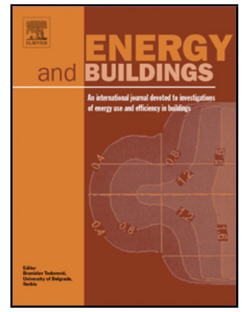
Title	The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings
Authors	Gallagher, Colm V.;Bruton, Ken;Leahy, Kevin;O'Sullivan, Dominic T. J.
Publication date	2017-10-13
Original Citation	Gallagher, C. V., Bruton, K., Leahy, K. and O'Sullivan, D. T. J. (2018) 'The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings', Energy and Buildings, 158, pp. 647-655. doi: 10.1016/j.enbuild.2017.10.041
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://doi.org/10.1016/j.enbuild.2017.10.041
Rights	https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2025-04-17 15:52:41
Item downloaded from	https://hdl.handle.net/10468/5887

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Accepted Manuscript



Title: The Suitability of Machine Learning to Minimise Uncertainty in the Measurement and Verification of Energy Savings

Author: Colm V. Gallagher Ken Bruton Kevin Leahy Dominic T.J. O'Sullivan

PII: S0378-7788(17)31851-0
DOI: https://doi.org/doi:10.1016/j.enbuild.2017.10.041
Reference: ENB 8061

To appear in: *ENB*

Received date: 25-5-2017
Revised date: 24-8-2017
Accepted date: 8-10-2017

Please cite this article as: Colm V. Gallagher, Ken Bruton, Kevin Leahy, Dominic T.J. O'Sullivan, The Suitability of Machine Learning to Minimise Uncertainty in the Measurement and Verification of Energy Savings, *<![CDATA[Energy & Buildings]]>* (2017), https://doi.org/10.1016/j.enbuild.2017.10.041

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Highlights: "The Suitability of Machine Learning to Minimise Uncertainty in the Measurement and Verification of Energy Savings"

Colm V. Gallagher, Ken Bruton, Kevin Leahy, Dominic T.J. O'Sullivan

August 23, 2017

- The suitability of machine learning algorithms to improve the measurement and verification of energy savings in industrial buildings is presented.
- Six individual modelling algorithms are applied and their prediction accuracy was validated in the context of a case study.
- Machine learning was found to reduce error by 51.1% compared to an assumed typical approach.
- A higher measurement frequency does not always result in reduced uncertainty in savings quantified.
- The use of machine learning under missing baseline data conditions is shown to be advantageous.

Accepted Manuscript

ACCEPTED MANUSCRIPT

The Suitability of Machine Learning to Minimise Uncertainty in the Measurement and Verification of Energy Savings

Colm V. Gallagher[a,b,*], Ken Bruton[b,c], Kevin Leahy[a,b], Dominic T.J. O'Sullivan[a,b]

[a] *Intelligent Efficiency Research Group, School of Engineering, University College Cork, Cork, Ireland*
[b] *MaREI Centre, Environmental Research Institute, University College Cork, Cork, Ireland*
[c] *Department of Mechanical, Biomedical and Manufacturing Engineering, Cork Institute of Technology, Cork, Ireland*

Abstract

Accurate energy modelling is a critical step in the measurement and verification (M&V) of energy savings, as a model for consumption in the baseline period is required. Machine learning (ML) algorithms offer an alternative approach to train these models with data-driven techniques. Industrial buildings offer the most challenging environment for the completion of M&V due to their complex energy systems. This paper investigates the novel use of ML algorithms for M&V of energy savings in industrial buildings. This approach enables the extension of the traditional project boundary also. The ML techniques applied consist of bi-variable and multi-variable ordinary least squares regression, decision trees, k-nearest neighbours, artificial neural networks and support vector machines. The prediction performances of the models are validated in the context of a biomedical manufacturing facility to find the optimal model parameters.

Results show that models constructed using ML algorithms are more accurate than the conventional approach. A 51.09% reduction in error was achieved using the optimal model algorithm and parameters. The use of a higher measurement frequency reduced the spread of error across the six models. However, further analysis proved the use of more granular data did not always benefit model performance. Results of the sensitivity analysis showed the proposed ML approach to be beneficial in circumstances where missing baseline data limits the model training period length.

Keywords: Measurement and verification, energy efficiency, machine learning, energy modelling, uncertainty analysis, energy performance

1. Introduction

An energy conservation measure (ECM) is implemented to reduce consumption in energy systems. The term ECM encompasses a wide range of measures and is used to refer to any energy performance improvement project. In recent years, the measurement and verification (M&V) of energy savings has received increased focus due to measures imposed by energy policy worldwide. Improving efficiency across all elements of energy systems is being utilised as an essential tool to achieve policy targets. The European Union has issued the Energy Efficiency Directive to ensure member states achieve individual improvements in energy efficiency [1]. Accurate and reliable estimation of energy savings from a wide range of ECMs are needed to cumulatively ensure the effective implementation of the Directive.

To quantify the savings resulting from an ECM, the energy consumption in the reporting period, or post-ECM, must be compared to what the consumption would have been had the ECM not been implemented. This is known as the adjusted baseline. Hence, the post-ECM consumption must be normalised to pre-ECM conditions. Engineering or statistical methods are typically employed to construct a baseline model capable of performing this normalisation. The accuracy with which the energy savings can be quantified is reliant upon the level of uncertainty that exists. Thus, uncertainty analysis is a necessary step in reliably estimating the energy savings, as an estimation of energy savings alone is insufficient to validate an ECM. A quantifiable measure of uncertainty must also be provided to give an indication of the savings estimation accuracy.

The three sources of uncertainty in M&V are sampling, modelling and metering. This body of research is concerned with minimising the modelling uncertainty that exists in projects. Estimating the uncertainty in M&V provides a deeper insight into the energy savings and supports the decision making process in developing baseline consumption models [2]. Both the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) and the Efficiency Valuation Organization (EVO) provide extensive methodologies to quantify all three sources of uncertainty in any project [3, 4]. A statistical methodology to evaluate the predictive accuracy of building energy baseline models has also been developed and applied [5]. The application of this evaluation procedure reviewed five energy baseline models and highlighted the potential of each to minimise modelling uncertainty in whole-building cases. However, it was found that all the models re-

*Corresponding Author
Email address: c.v.gallagher@umail.ucc.ie (Colm V. Gallagher)
URL: www.ucc.ie/en/ierg (Colm V. Gallagher)

ACCEPTED MANUSCRIPT

viewed performed poorly when the energy consumption of the building varied in ways that are not predictable from the outdoor air temperature or the time of week. This is a common occurrence in industrial buildings containing complex energy systems in which many variables are affecting consumption. The energy consumption in these applications is most often a function of a number of independent variables which frequently change over time.

In 2015, industry accounted for 25.3% of final energy consumption in the EU [6] and 21.1% in Ireland [7]. ECMs are being used to reduce this energy consumption that industrial activities are responsible for. However, the increased scrutiny of M&V in these applications has highlighted the shortcomings in accurately verifying the energy savings and associated uncertainty. M&V of energy savings in the residential and commercial sectors is in a more mature state, with the methods used in each intertwined with one another. Research in this field is currently progressing towards the concept of M&V 2.0, which differs from traditional M&V as it uses large data sets and automated advanced analytics to streamline and scale the process [8]. The topic of industrial applications of M&V has been largely neglected due to its dissimilarity with residential and commercial applications. As a result of this, it is more difficult to perform M&V and the lack of accurate verification impedes efforts to maximise industrial energy efficiency [9].

Outside air temperature and production levels are often given as common examples of independent variables that can be used to develop baseline models. In machine learning, these variables are called features. For many energy systems in industrial buildings, the features that impact energy consumption most significantly can be identified using knowledge of engineering first principles. M&V practitioners are often satisfied to develop baseline models using these most prominent features as they can be employed to achieve reasonable levels of accuracy. This results in many relationships between energy consumers not being analysed, but would the inclusion of these features significantly improve the accuracy of energy savings verification? The size of these relationships relative to the more prominent features make them difficult to infer and utilise for the purposes of model development. This paper investigates the ability of machine learning to verify energy savings in industrial buildings, with the objective of minimising uncertainty to deliver accuracy and precision. The proposed approach allows the less prominent features to be included in the analysis, as it enables a wider project boundary be employed and more efficient processing of data.

2. Research Questions

A number of previous studies have investigated the suitability of machine learning for modelling baseline energy consumption in end-use residential and commercial applications. These are reviewed in detail in Section 3. There are many commonalities between energy modelling in residential and commercial buildings. However, industrial buildings operate quite differently with complex, multi-faceted energy systems. The deficiency of research on M&V in the industrial sector means critical questions remain unanswered, the following of which are addressed in this paper:

1. Does a wider boundary of analysis aid the reduction of uncertainty in M&V?
2. Can machine learning be utilised to improve the prediction accuracy for M&V in industrial buildings?
3. How does missing baseline data affect the ability to accurately perform M&V?
4. Can optimal modelling parameters be identified for all use cases?

3. Related Work

3.1. Guidance Documentation

EVO developed the most widely used and recognised M&V protocol: the International Performance Measurement and Verification Protocol (IPMVP). IPMVP defines four distinct methodologies that can be applied to estimate the energy savings for a given project [10]. Option A and B focus on isolating the project boundary to the ECM only. If this is not possible, Opton C allows you to use a whole-facility approach, although this is only advisable if the savings are greater than 10% of the total site energy use. Option D is a calibrated simulation approach often used in situations when there is no baseline data available. An initial estimate of potential savings is used to guide this decision making process. Guidance is provided to identify parameters but it does not provide a rigid calculation process to follow [11]. ASHRAE Guideline 14 also provides three distinct methodologies for calculating energy savings. These are akin to that of IPMVP. Both EVO and ASHRAE define acceptable levels of uncertainty as when the savings are larger than twice the standard error of the baseline value [3, 4]. ASHRAE also stipulate that maximum levels of uncertainty must be calculated based on annual savings only. Other M&V protocols include the guidelines published by the U.S. Department of Energy and the California Energy Evaluation Protocols. Typically, M&V costs are 1% to 5% of total project costs using IPMVP Option A and 3% to 10% for verification using Option B [12]. Minimising these costs, while maintaining accuracy, is critical to maximising the benefits of an ECM.

3.2. Machine Learning in M&V

Machine learning has been extensively applied to model energy consumption across residential and commercial applications in the field of energy engineering. M&V practices in industrial buildings are dissimilar from the more developed residential and commercial sectors. Hence, there remains significant scope to improve the accuracy with which industrial energy savings are verified. In this section, the research published to date is reviewed to depict the state of maturity in each sector.

3.2.1. Residential & Commercial Applications

In the residential buildings sector, the available data for developing baseline models is often restricted to whole building consumption, outside air temperature and occupancy. It is important to note that despite occupancy being highly correlated

Accepted Manuscript

ACCEPTED MANUSCRIPT

with energy use, it does not always significantly improve the accuracy of the baseline model [13]. The most widely recognised technique for modelling building energy consumption is with the use of physics-based models. Lü et al. [14] developed a simple and accurate model for building energy consumption using minimal physical parameters.

Data-driven modelling approaches are less well established, but can be implemented successfully. Ekici and Aksoy [15] utilised the orientation, insulation thickness and transparency ratio to develop a back propagation artificial neural network (ANN) capable of predicting heating energy consumption with 94.8-98.5% accuracy. Similarly, Yu et al. [16] constructed a building energy demand predictive model based on the decision tree method with an accuracy of 92% on a test data set. Catalina et al. [17] propose a multiple regression model to predict heating energy demand, based on the building global heat loss coefficient, the south equivalent surface, and the difference in the indoor set point temperature and the sol-air temperature. The use of these model features is a novel approach that could be employed more frequently in M&V. Swan and Ugursal [18] reviewed both top-down and bottom-up modelling of the end-use energy consumption in the residential sector and concluded that the continuous development of alternative energy sources and technologies, coupled with the focus being placed on efficiency, has created a requirement for bottom-up models. Additionally, Dounis [19] demonstrated that these techniques may play an important role in conserving energy in buildings. There is a well established precedent for energy modelling in residential applications of M&V and this ensures low levels of uncertainty are achievable in everyday projects.

The largest portion of published research in the field of M&V focuses on commercial buildings. There is a wide-range of studies proving the suitability of ANNs for modelling commercial buildings energy consumption [15, 20–25]. The ASHRAE Great Energy Predictor Shoot-out identified ANNs as the most accurate method of modelling a building's energy use [26, 27]. In the second ASHRAE Great Energy Predictor Shoot-out, hourly whole-building data was used by four competitors to model the energy consumption of commercial buildings. A machine learning approach employing ANNs won the competition, although a statistical regression method was found to perform almost as well as the ANNs [24]. Additional machine learning algorithms such as support vector machines (SVM) are also proven to be accurate in modelling the energy consumption of buildings in the commercial sector [21, 28–30]. Neto and Fiorelli [22] present a comparison between physics based modelling and ANNs for forecasting energy consumption and highlight the requirement for training data as a hindrance to using the ANN, despite it performing as well as the physics based model. This requirement is common across all machine learning algorithms and the performance of the models often depend on the quantity of training data available. More recently research has been conducted on the use of first-principle, or physics-based, modelling to train machine learning models and this offers further potential for the utilisation of machine learning in M&V [31]. Granderson et al. [32] applied novel M&V 2.0 modelling approaches to commercial buildings and showed that interval data acquired from advanced metering infrastructure offers significant potential for scaling the adoption of M&V using a whole-building approach.

3.2.2. Industrial Applications

The lack of research investigating energy modelling in industrial M&V applications, coupled with the success achieved in residential and commercial applications, are strong indicators of the potential advancements that are possible. In contrast, industrial buildings contain multiple factors that affect the more complex consumption and the savings realised are often small relative to the total facility consumption. The M&V methods across all applications hold many commonalities; however, without specific methodologies that address the requirements of each case, the accuracy of energy savings estimation is restricted. The available information stored within the vast quantities of data that are common in industrial facilities offers a powerful opportunity to advance the subject area.

Kelly Kissock and Eger [9] present a methodology for measuring whole-facility industrial energy savings that accounts for weather and production. The method can use sub-metered data or whole-plant utility billing data. The purpose of the methodology is to extract information about savings from the data set; however, this is limited by the quality of the data set itself. The use of monthly data was noted as a significant limiting factor in analysis of ECMs on different time-scales. This research highlights the potential benefits of employing granular energy data for the purposes of M&V. Rossi and Velázquez present a prescriptive methodology for performing M&V on CHP plants in industrial buildings [33]. These methodologies are beneficial to progressing the field of M&V, although they are only applicable under specific conditions. Research has established that machine learning-based energy modelling approaches are capable of performing better than traditional approaches, while requiring significantly less input data from the end user [30, 34]. The question remains as to how these techniques can be integrated into the process of M&V to minimise uncertainty, while not increasing costs.

3.3. Optimal Model Parameters

The model training period and the measurement frequency of the data are variable and have been found to significantly influence prediction performance. Cho et al. [35] examined the effect of measurement frequency on the performance of temperature dependent regression models. The training data varied from 1 day to 3 months with the relative error for predicting annual energy consumption being 100% and 6% respectively. Jain et al. [36] analysed the impact of measurement frequency varying from 10-minute to daily measurements and found that hourly data was the most appropriate for multi-family residential applications. Zhao and Magoulès [34] conducted a comprehensive review of simplified engineering methods, statistical methods and artificial intelligence methods for the prediction of energy consumption in buildings. This review concluded that it is difficult to identify any one best performing model without complete comparison under the same circumstances. Hence,

ACCEPTED MANUSCRIPT

the optimal training period length and data measurement frequency for each individual case must be investigated within the context of the unique project conditions. This paper performs this analysis in a biomedical manufacturing facility and investigates the impact training period characteristics have on model performance.

4. Methodology

As outlined in Section 2, the objective of this analysis is to assess the suitability of machine learning for improving M&V in industrial buildings. As concluded by Zhao and Magoulès [34], each model must be compared under the same circumstances for a complete analysis. Hence, a biomedical manufacturing facility was chosen as a test-bed for this proposed approach. The site has an approximate footprint of 4.2 acres and has over 1000 employees. It comprises processes such as casting, milling, grinding and packaging. As a result, it utilises a significant quantity of energy in both machines directly, as well as in the preparation and conditioning of clean-rooms to enable medical device manufacture. The site was deemed applicable as no ECM was performed during the period of analysis. This allowed for complete evaluation of the model prediction performance; something that would not be possible had the consumption changed as a result of an ECM. The building characteristics (e.g. envelope, materials, openings) did not change during the period of analysis, hence, these factors were not considered in the baseline energy model.

The compressed air electricity consumption at the site was selected to be modelled for this investigation. ECMs are very commonly performed on compressed air systems as they can achieve savings of 20-50% [37]. In this case study, the total electricity consumption of the compressors was metered, but there was no flow meter on the compressed air header. To carry out M&V on this system, there are two clear options. One solution is to install a meter to measure compressed air flow. This would be useful for quantifying savings on the generation side of the system, although it would increase costs and delay the project as baseline data would need to be gathered. The alternative solution is to model the compressed air electricity consumption on it's relationships with other energy consumers within the facility. This approach requires the construction of a model of baseline consumption; thus, it was deemed an ideal case study for conducting the analysis.

There were 24 months of data available for this study. The data set was split to hold out 12 months of data for training the models and the remaining 12 months stored separately to be used as a testing data set for model evaluation on unseen data. These would be representative of baseline (pre-ECM) and reporting (post-ECM) periods in a practical application. It is important to note that although an 80/20 split of training to test data is more common in machine learning, a 12 month testing period is representative of real world M&V applications. This approach has been developed and applied previously by Granderson et al. [5] and is effective in simulating the conditions necessary for comparison of model performance. Table 1 contains a summary of the data available to be used in the modelling process. The input variables employed contain a mix of both building and process related energy consumers. These are classified in Table 1 and ensures that both building and process related energy consumption is accounted for in the models, thus providing an accurate representation of the system operation in the baseline period. All analysis was carried out using the open source programming language R.

4.1. Algorithms

Five prominent machine learning algorithms were selected to solve this problem; multi-variable linear regression, decision tree regression, k-nearest neighbours, artificial neural networks and support vector machines. There are a wide range of algorithms that could be applied for this type of analysis; however, these five were selected based on previous success in the field in the published literature reviewed in Sect. 3. For comparative purposes, it was decided that an ordinary least squares regression model constructed using outside air temperature and production electricity was a reasonable assumption of a typical approach taken by M&V practitioners.

In machine learning, the term hyper-parameter is used to distinguish from standard model parameters. Standard model parameters are learned in the model training process. However, hyper-parameters cannot be directly learned from the regular training process. These parameters convey properties of the model such as its complexity and the speed of learning. The optimised value of each hyper-parameter was found by performing a grid search on possible values and using 10-fold cross-validation on the training data to determine the best performing model. 10-fold cross-validation was deemed an appropriate means to estimate prediction error based on published research [38, 39]. It also prevents the modelling algorithms from over-fitting to the training data. In addition to this, the ANN weight decay hyper-parameter is used to prevent over-fitting.

The optimised hyper-parameter values are specific to each individual application and thus, allow the methodology to be adaptable and customisable to the properties of any given data set. Descriptions of the machine learning algorithms applied, the hyper-parameters associated with each and the notation used throughout this paper can be found in Table 2.

4.2. Performance Metrics

The models trained using the baseline period data were applied to the testing data to evaluate prediction performance. The coefficient of variation of the root mean square error (CV(RMSE)) is a measure of the variability between the actual and predicted values. It is calculated by dividing the root mean square error by the average energy consumption [3]. The CV(RMSE) is a metric used to quantify modelling error in both ASHRAE Guideline 14 and IPMVP. The equation for the metric is provided in Equation 1, where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the average of the actual value, and n is the total number of predictions in the period of analysis. In the ASHRAE Great Energy Predictor Shoot-out II, CV(RMSE)

Accepted Manuscript

ACCEPTED MANUSCRIPT

Variable	Description	Average Value	Type	Service
Compressed Air	Total electricity consumption of four air compressors.	363.3 kW	Dependent	Process
Chilled Water	Total electricity consumption of chilled water generation system.	142.7 kW	Predictor	All
Heating	Electrical heating load.	36.4 kW	Predictor	All
Cooling Tower Water Pumps	Electricity consumption of cooling tower water pumps.	5.9 kW	Predictor	All
Dust Extraction	Total electricity consumption of dust extraction system.	87.6 kW	Predictor	Process
Grid Electricity	Quantity of electricity imported from the national grid. On-site generation services the remainder of the load.	1747 kW	Predictor	All
Production HVAC	Electricity consumption of HVAC servicing production floor area.	82.21 kW	Predictor	Process
Non-Production HVAC	Electricity consumption of HVAC servicing all non-production areas.	28.3 kW	Predictor	Building
Production	Production equipment electricity consumption.	1355 kW	Predictor	Process
Outside Air Temperature	Outside air temperature measured in degrees Celsius.	15 °Celsius	Predictor	Building
Operation	Status of operation in the facility (1 = In-production, 0 = On-standby).	-	Predictor	Process

Table 1: Summary of variables included in the available dataset.

Algorithm	Description	No. of Features	Hyper-parameters	Grid Search	Notation
Bi-variable Linear Regression	An ordinary least squares approach assumed to be representative of typical M&V practice. Production electricity consumption and outside air temperature are the features employed.	2	Intercept	True/False	Bi-Lin
Multi-variable Linear Regression	A more detailed ordinary least squares model constructed using 9 additional features from the available data set.	11	Intercept	True/False	Multi-Lin
Decision Tree Regression	Models in the form of a tree structure with decision nodes. The topmost node in a tree represents the best predictor.	11	Maximum tree depth	d_{max} = 1:10	Tree
k-Nearest Neighbours	Non-parametric model where the input consists of the k closest training examples in the feature space. The output is the average of the values of its k-nearest neighbours.	11	Maximum no. of neighbours Distance Kernel	k_{max} = 1:10 d = 1:5 *kernel* = rectangular, triangular	k-NN
Artificial Neural Networks	Non-linear statistical model. It is a two-stage regression model typically represented by a network diagram. A single hidden layer feed-forward neural network was developed in each instance.	11	No. of hidden units Maximum no. of iterations Threshold Weight decay	*size* = 1:11 it_{max} = 500,000 t = 0.01 d = (0.5,0.1,0.01,0.001)	N-net
Support Vector Machines	Non-parametric technique reliant on kernel functions. Examples are represented as points in space with a clear gap separating mapping categories.	11	Kernel Cost	*kernel* = linear c = (0.25,0.5,1,10)	SVM

Table 2: Description of machine learning algorithms employed in the analysis.

Accepted Manuscript

ACCEPTED MANUSCRIPT

was the primary metric employed to determine overall model ranking.

$$CV(RMSE) = \frac{\sqrt{(1/n)\sum_i^n (y_i - \hat{y}_i)^2}}{\bar{y}} * 100 \quad (1)$$

In the same study, normalised mean bias error (NMBE) was the secondary metric used to support the evaluation process [24]. The mean bias error is an indication of overall bias in a regression model and is calculated using the formula in Equation 2. It quantifies the tendency of a model to over or underestimate across a series of values. This metric is independent of time-scale so care must be taken as overall positive bias error can cancel out negative bias. In contrast, the CV(RMSE) does not suffer from this problem.

$$NMBE = \frac{(1/n)\sum_i^n (y_i - \hat{y}_i)}{\bar{y}} * 100 \quad (2)$$

The median of the absolute relative error (med(absRTE)) is a useful metric to understand the typical error in the prediction of total energy consumption over the testing period. The metric is similar to the mean absolute percent error, but uses the median to overcome the sensitivity of the mean to extreme values. Equation 3 contains the formula for calculating the median of the absolute relative error.

$$med(absRTE) = median(\frac{abs(y_i - \hat{y}_i)}{y_i}) \quad (3)$$

4.3. Model Uncertainty

The procedure for calculating the uncertainty introduced by the baseline model is explicitly defined by both IPMVP and ASHRAE Guideline 14. In both cases, the CV(RMSE) is used, along with other measures, to compute the uncertainty associated with the model. The formulae used to calculate this uncertainty varies between IPVMP (*Statistics and Uncertainty for IPMVP*) and ASHRAE Guideline 14. However, the CV(RMSE) is common to both cases, with all other equation parameters being independent of the model constructed. These parameters include t-statistics, sample size and the number of independent variables. As CV(RMSE) is the only parameter effected by the model performance, it is important to focus on minimising it to achieve the objectives stated in Sect. 2.

In a practical application using IPMVP approaches, the CV(RMSE) is calculated by applying the baseline model to the pre-ECM dataset (i.e. applying the model to the data used to train it). This approach is very susceptible to over-fitting the model to the training data. To overcome this issue, a building in which no ECM has been implemented was chosen as a test site. A truer measure of performance can be found by applying the baseline model to the testing data. This model validation procedure allows for direct comparison of the adjusted baseline, calculated by the model, and the measured post-ECM consumption. This approach has previously been developed by Granderson et al. [5].

5. Results and Discussion

5.1. Potential of Additional Model Features

A notable characteristic of the analysis is the use of additional model features that would otherwise be overlooked. Sect. 1 describes the relevance of these features and the typical approach taken by M&V practitioners. It was deemed that a typical approach would use production electricity consumption and outside air temperature as the predictors. This assumption is based on correlation analysis and engineering first principles, which are commonly employed techniques in M&V.

Analysis was carried out to assess the value in employing 9 additional features in the model construction process. Hence, the second approach employed all 11 features that were available in the facilities data set. Baseline energy models were developed using an ordinary least squares regression algorithm for both the traditional (2 model features) and proposed (11 model features) approaches. 12 months of training data and a selection of measurement frequencies were used in the analysis. The performance of each model was evaluated using 12 months of unseen testing data.

Fig. 1 illustrates the results of this analysis. For daily, hourly and quarter-hour measurement frequencies, it was found that the models developed using all 11 model features outperform those constructed using the more traditional approach. This is not the case when less granular weekly and monthly data is employed. The more straightforward approach performs best in these instances. It is the hypothesis that the more complex model is too reliant on the additional variables, the detail of which is lost at these measurement frequencies. The three best performing models across the spectrum of temporal granularities are those developed using all 11 model features. These would ordinarily not be employed for this analysis using current methodologies. The use of these additional features expands the boundary of analysis; thus, offering a novel and more accurate means of achieving the objectives of M&V. The best performing model overall uses all 11 features and a 15-minute measurement frequency. In comparison to the most accurate traditional model, CV(RMSE) and NMBE are reduced by 15.9% and 75.6% respectively.

5.2. Harnessing the Power of Additional Features

The findings in Sect. 5.1 identified the potential benefits of expanding the scope of analysis beyond the currently employed techniques. Further analysis investigated the ability of different algorithms to improve the accuracy in estimating the adjusted baseline, beyond that of the ordinary least squares model constructed using the 11 model features. The training period was held constant at 12 months with monthly, weekly, daily and quarter-hourly measurement frequencies reviewed. The traditional, bi-variable approach used previously was again employed for comparative purposes, with all other modelling algorithms utilising all 11 features.

Fig. 2 contains a graphical representation of the model performance for each set of project parameters. Comprehensive performance metrics are included in Table 3. The use of a higher measurement frequency does not always improve the

ACCEPTED MANUSCRIPT

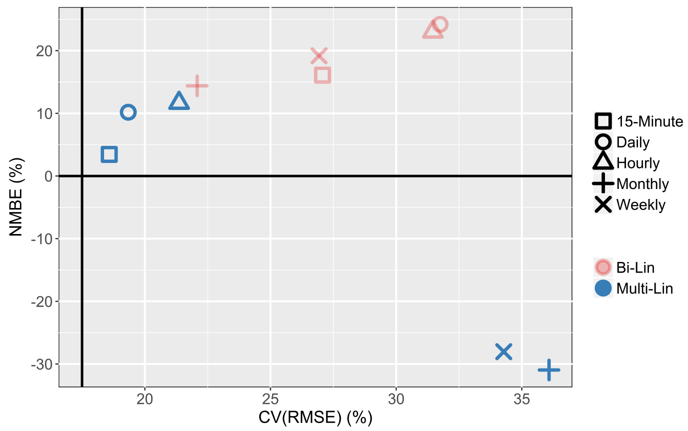


Figure 1: Assessing the value of additional model features for different temporal granularities.

performance of each model; however, it does reduce the spread of error between all 6 models. Monthly data provided the most accurate model across the analysis, a feed-forward ANN with -3.24% NMBE and 10.8% CV(RMSE). In contrast, the ordinary least squares, decision tree and SVM models, constructed using all 11 features, performed with unacceptable levels of accuracy at this measurement frequency. The k-nearest neighbours model constructed using weekly interval data predicted the adjusted baseline with next best accuracy. The performance of the ANN and k-nearest neighbours models deteriorates significantly as measurement frequency increases. In contrast to this, the performance of the multi-variable ordinary least squares model becomes significantly more accurate as the measurement frequency increases. SVMs become more prominent as the measurement frequency increases also. Decision trees perform poorly across all conditions. The algorithm was unable to sufficiently construct a model at lower measurement frequencies, possibly as it is too simplistic a means of modelling this complex system. The performance of the ANNs are erratic, with the greatest accuracy achieved at lower measurement frequencies.

A comprehensive and complete analysis of all approaches under the same set of operating conditions enables an accurate comparative review to be carried out. There is no clear most appropriate modelling algorithm across all four measurement frequencies. Therefore a conclusion cannot be drawn on the most accurate machine learning algorithm for modelling baseline energy in M&V, although there is a clear best performing model for each measurement frequency. Despite this, only one model is required for the purposes of any M&V project. The optimal model is the ANN that uses 11 model features and monthly interval data. This highlights the need to conduct this type of analysis in each case as individual project requirements and characteristics will influence model performance.

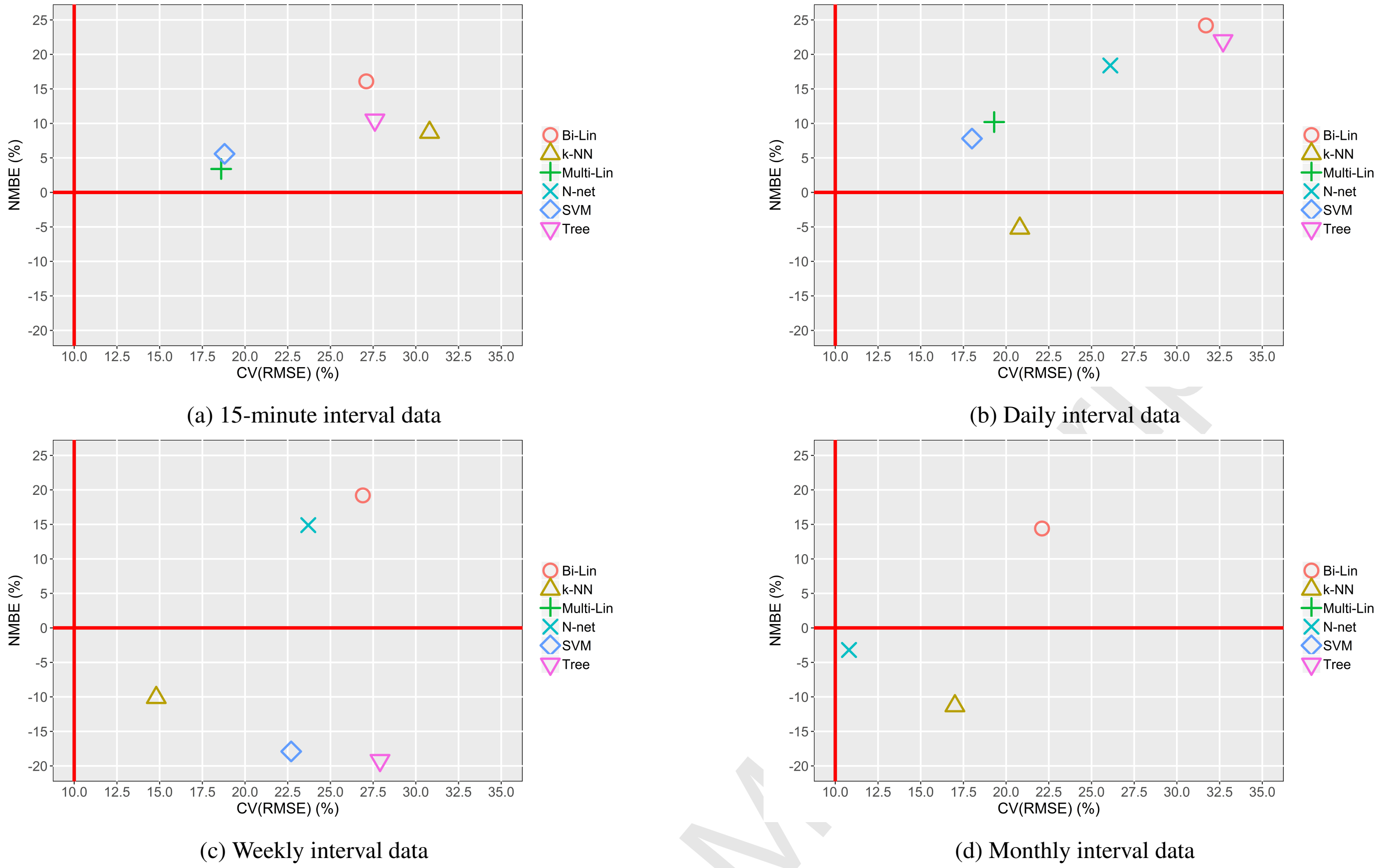
5.3. *Sensitivity Analysis: Training Period Length*

In Sect. 5.1 and 5.2, a 50:50 ratio of training to testing data was applied. Although this is uncommon in most machine learning applications, it is representative of typical M&V cases in which only 12 months of baseline data are available and the adjusted baseline must be predicted for a 12 month reporting period. The algorithms and approach proposed thus far have been proven to be suitable for improving the accuracy of M&V. However, the sensitivity of this accuracy to the quantity of training data available offers further scope to evolve M&V practices in industrial applications. Data availability is an ever present constraint to many M&V practitioners. To adhere to IPMVP, backfilling cannot be carried out in the baseline period. A lack of data in the baseline period is often the single biggest hindrance to completing accurate M&V. Additional metering infrastructure is usually installed to overcome this issue, but this increases project costs and delays implementation as baseline data must be gathered.

To simulate the conditions of missing data, the best performing algorithms from Sect. 5.2 were applied to construct models based on limited training data. The best performing model using the proposed approach employed the ANN algorithm and a monthly measurement frequency. The algorithm hyperparameters were optimised using 10-fold cross validation, resulting in a hidden layer with 11 units and a weight decay of 0.5. For comparative purposes, the best performing model developed using the traditional approach was also brought forward for analysis. This also used a monthly measurement frequency, while employing the ordinary least squares linear regression algorithm and just two predictor variables. Both models were evaluated using 3, 6, 9 and 12 months testing data. For reduced training periods, the most recent period of data was considered in each case. The practicalities and requirements of M&V limit its accuracy and hence, these unconventional training to testing ratios need to be investigated to fully understand the limitations of the approach.

The sensitivity of the bi-variable model is illustrated in Fig. 3. It is clear that the length of training period directly improves prediction accuracy in every case. This analysis shows the dependency of the traditional approach on the availability of baseline data to train the model. This restricts the potential applications of M&V. The results of the sensitivity analysis conducted on the ANN are included in Fig. 4 and offer a more intriguing insight into the potential of the proposed approach. The models constructed using shorter training periods are capable of performing adequately with respect to those constructed using longer training periods. It is very common in M&V that models are required to predict the adjusted baseline for a 12 month period. This is akin to that of a 12 month testing period in this analysis. For these conditions, the model constructed using 6 months training data performs with a CV(RMSE) of 10.8%, while the model constructed using 11 months training data results in a CV(RMSE) of 10.7%. The prediction accuracy achieved, using almost half the quantity of training data, highlights the potential of the proposed machine learning-based approach to be applicable to projects with limited data available. This pattern in performance is common across the 6, 9 and 12 month testing data sets. Across all testing data sets, it is clear that acceptable performance, relative to that of a 12 month training period, can be achieved using 6 months of training data or more. The results for the testing set containing 3 months of data show promising results, with a CV(RMSE) of

ACCEPTED MANUSCRIPT



(a) 15-minute interval data

(b) Daily interval data

(c) Weekly interval data

(d) Monthly interval data

Figure 2: The performance of each algorithm using 12 months training data and 12 months testing data.

	Bi-Variable Lin. Reg.	Multi-Variable Lin. Reg.	Decision Tree	k-Nearest Neighbours	Neural Network	Support Vect. Machine
Monthly						
NMBE	14.39	-30.95	-	-11.32	-3.24	-33.26
CV(RMSE)	22.08	36.08	-	17.00	10.80	41.46
med(absRTE)	8.41	27.7	-	14.2	6.89	43.3
Weekly						
NMBE	19.19	-28.03	-19.23	-10.13	14.93	-17.86
CV(RMSE)	26.93	34.28	27.93	14.82	23.67	22.72
med(absRTE)	21.58	28.36	18.89	8.06	12.74	17.58
Daily						
NMBE	24.20	10.17	22.01	-5.23	18.40	7.85
CV(RMSE)	31.75	19.34	32.73	20.81	26.15	18.04
med(absRTE)	27.17	13.65	29.75	12.45	21.95	13.24
Quarter-Hourly						
NMBE	16.10	3.45	10.50	8.70	17.43	5.62
CV(RMSE)	27.07	18.58	27.63	30.82	36.33	18.80
med(absRTE)	19.64	12.32	22.06	20.97	23.09	13.09

Table 3: Performance of each model developed using 12 months training data and evaluated using 12 months testing data with varying measurement frequency.

Accepted Manuscript

just 3.7% achieved using 4 months of training data. This pushes the limitations of traditional techniques to a wider spectrum of applications, while also improving prediction performance and hence, minimising uncertainty.

6. Conclusions

Machine learning was found to be an excellent means of minimising uncertainty in industrial applications of M&V. The suitability of five different machine learning techniques were

ACCEPTED MANUSCRIPT

Figure 3: Sensitivity analysis of bi-variable ordinary least squares linear regression model performance to quantity of training and testing data.

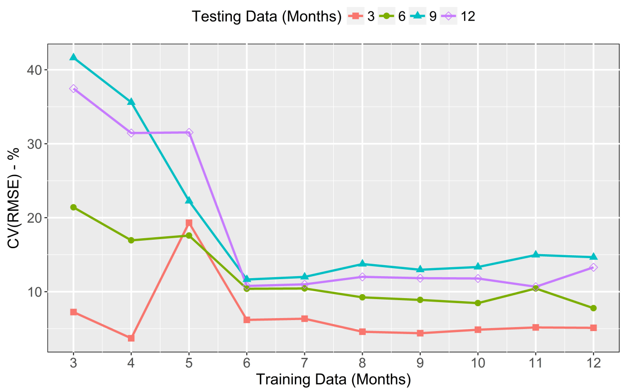


Figure 4: Sensitivity analysis of ANN model performance to quantity of training and testing data.

examined with respect to an assumed typical approach. This analysis was carried out in the context of a biomedical manufacturing facility, as it has already been proven that operating conditions must be kept constant for each technique to enable a complete investigation [34]. The results identify the potential performance improvements that are achievable by extending the boundary of analysis and incorporating additional independent variables into the model construction process. Machine learning was used as a tool to extract the knowledge contained within the data for these variables and construct models of the baseline energy consumption with varying degrees of success. The use of data-driven modelling enables a dynamic and flexible approach be taken to a wide range of projects.

Section 5.1 highlights the accuracy improvements that can be achieved by employing additional features in the analysis, i.e. extending the typical project boundary. The prediction accuracy was improved when the measurement frequency was daily or higher. The CV(RMSE) and NMBE were reduced by 15.9% and 75.6% repectively, when the best performing model constructed using al 11 features is compared to that of the more traditional, bi-variable approach. This initial analysis showed promise and hence, the same methodology was applied with four other modelling algorithms in an attempt to further improve the prediction accuracy. An exhaustive methodology was applied to construct the models for varying measurement frequency. This was necessary to identify the optimal modelling algorithm and parameters. The most accurate model was a single layer feed-forward neural network trained using monthly data. The CV(RMSE), NMBE and med(absRTE) for this model were evaluated to be -3.24%, 10.8% and 7% respectively. This represents a further 41.9% reduction in CV(RMSE) compared to that of the best performing model in the earlier analysis presented in Sect. 5.1. In addition to this, it is important to note that the spread of model error reduced as the measurement frequency increased. This is advantageous in developing consistently accurate models, as opposed to the sporadic performance at lower measurement frequencies.

The best performing models for the traditional and proposed approaches were brought forward to Sect. 5.3. For the proposed approach, this was the ANN constructed with monthly data and for the traditional M&V approach, this was the ordinary least squares regression model that contained just two predictor variables and used a monthly measurement frequency. The sensitivity of each model to the quantity of training and testing data available was investigated. It was found that the model constructed using the traditional approach was highly dependent on the length of the baseline period. Performance degraded across all testing data sets when the training period was reduced. In contrast to this, the ANN models were found to perform significantly better. Sufficient accuracy was achievable in all cases for training periods greater than 6 months. A CV(RMSE) of 10.8% was achieved with a 6 month training period and a 12 month testing period. This highlights the potential benefits of the proposed approach in overcoming the limitations of traditional M&V in industrial buildings. Interestingly, a 4 month training period and a 3 month testing period resulted in CV(RMSE) of 3.7%.

7. Future Work

It has been proven that machine learning is a suitable tool that can be used to minimise uncertainty in M&V. Future work will focus on formalising a methodology that will incorporate these machine learning techniques into standard M&V practice, without putting a strain on resources. The suitability of the approach will be further validated through the use of additional case studies. This will also aid the strengthening of the research field. Beyond this, there is an opportunity to progress M&V in industrial buildings to the same level as that of residential and commercial applications. Utilising the proposed techniques for the purposes of automated advanced analytics, commonly referred to as M&V 2.0, is an objective the authors intend on achieving. This area will be the main focus of future work by the research team.

Acknowledgements

This research is based upon works supported by Science Foundation Ireland under Grant No. 12/RC/2302.

Accepted Manuscript

ACCEPTED MANUSCRIPT

References

[1] European Parliament and Council, Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency (2012). doi:10.3000/19770677.L_2012.315.eng.

[2] T. Walter, P. N. Price, M. D. Sohn, Uncertainty estimation improves energy measurement and verification procedures, Applied Energy 130 (2014) 230–236. doi:10.1016/j.apenergy.2014.05.030.

[3] American Society of Heating Refrigerating and Air-Conditioning Engineers, Guideline 14 - Measurement of Energy, Demand, and Water Savings (2014).

[4] Efficiency Valuation Organization, Statistics and Uncertainty for IPMVP, Tech. Rep. June, Efficiency Valuation Organization (2014).

[5] J. Granderson, P. N. Price, Development and application of a statistical methodology to evaluate the predictive accuracy of building energy baseline models, Energy 66 (1) (2014) 981–990. doi:10.1016/j.energy.2014.01.074.
URL http://dx.doi.org/10.1016/j.energy.2014.01.074

[6] Eurostat, Europe in figures - Eurostat yearbook 2016 (2016).
URL http://ec.europa.eu/eurostat/tgm/refreshTableAction.do?tab=table{&}plugin=1{&}pcode=tsdpc320{&}language=en

[7] M. Howley, M. Holland, Energy in Ireland 1990 2015 (2016) 88.

[8] J. Granderson, S. Touzani, S. Fernandes, C. Taylor, Application of automated measurement and verification to utility energy efficiency program data, Energy and Buildings 142 (2017) 191–199. doi:10.1016/j.enbuild.2017.02.040.
URL http://www.sciencedirect.com/science/article/pii/S0378778817300294

[9] J. Kelly Kissock, C. Eger, Measuring industrial energy savings, Applied Energy 85 (5) (2008) 347–361. doi:10.1016/j.apenergy.2007.06.020.

[10] Efficiency Valuation Organization, Core Concepts, Tech. Rep. October, Efficiency Valuation Organization (2016). doi:10.1073/pnas.1317033110.

[11] S. Ginestet, D. Marchio, Retro and on-going commissioning tool applied to an existing building: Operability and results of IPMVP, Energy 35 (4) (2010) 1717–1723. doi:10.1016/j.energy.2009.12.024.
URL http://www.sciencedirect.com/science/article/pii/S0360544209005441

[12] T. Jayaweera, F. Stern, D. M. Violette, R. Baumgartner, The Uniform Methods Project : Methods for Determining Energy Efficiency Savings for Specific Measures The Uniform Methods Project, Tech. Rep. April, National Renewable Energy Laboratory, US DOE (2013).
URL http://www.nrel.gov/docs/fy13osti/53827.pdf

[13] X. Liang, T. Hong, G. Q. Shen, Improving the accuracy of energy baseline models for commercial buildings with occupancy data, Applied Energy 179 (2016) 247–260. doi:10.1016/j.apenergy.2016.06.141.

[14] X. Lü, T. Lu, C. J. Kibert, M. Viljanen, A novel dynamic modeling approach for predicting building energy performance, Applied Energy 114 (2014) 91–103. doi:10.1016/j.apenergy.2013.08.093.

[15] B. B. Ekici, U. T. Aksoy, Prediction of building energy consumption by using artificial neural networks, Advances in Engineering Software 40 (5) (2009) 356–362. doi:10.1016/j.advengsoft.2008.05.003.

[16] Z. Yu, F. Haghighat, B. C. Fung, H. Yoshino, A decision tree method for building energy demand modeling, Energy and Buildings 42 (10) (2010) 1637–1646. doi:10.1016/j.enbuild.2010.04.006.

[17] T. Catalina, V. Iordache, B. Caracaleanu, Multiple regression model for fast prediction of the heating energy demand, Energy and Buildings 57 (2013) 302–312. doi:10.1016/j.enbuild.2012.11.010.

[18] L. G. Swan, V. I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13 (8) (2009) 1819–1835. doi:10.1016/j.rser.2008.09.033.

[19] A. I. Dounis, Artificial intelligence for energy conservation in buildings, Advances in Building Energy Research 4 (1) (2010) 267–299. doi:10.3763/aber.2009.0408.
URL http://www.tandfonline.com/doi/abs/10.3763/aber.2009.0408

[20] S. Wong, K. K. Wan, T. N. Lam, Artificial neural networks for energy analysis of office buildings with daylighting, Applied Energy 87 (2) (2010) 551–557. doi:10.1016/j.apenergy.2009.06.028.

[21] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks, Energy Conversion and Management 50 (1) (2009) 90–96. doi:10.1016/j.enconman.2008.08.033.

[22] A. H. Neto, F. A. S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy and Buildings 40 (12) (2008) 2169–2176. doi:10.1016/j.enbuild.2008.06.013.

[23] S. Karatasou, M. Santamouris, V. Geros, Modeling and predicting building's energy use with artificial neural networks: Methods and results, Energy and Buildings 38 (8) (2006) 949–958. doi:10.1016/j.enbuild.2005.11.005.

[24] J. S. Haberl, S. Thamilseran, Great Energy Predictor Shootout II: Measuring retrofit savings - overview and discussion of results, ASHRAE Transactions 102 (2) (1996) 419–435.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-0030315652{&}partnerID=tZOtx3y1

[25] J. F. Kreider, D. E. Claridge, P. Curtiss, R. Dodier, J. S. Haberl, M. Krarti, Building Energy Use Prediction and System Identification Using Recurrent Neural Networks, Journal of Solar Energy Engineering 117 (3) (1995) 161. doi:10.1115/1.2847757.

[26] J. Kreider, J. Haberl, Predicting hourly building energy use: The great energy predictor shootout – Overview and discussion of results, in: 1994 American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE) annual meeting, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States), 1994.

[27] J. Kreider, J. Haberl, Predicting hourly building energy usage: The results of the 1993 ASHRAE Great Energy Predictor Shootout identify the most accurate method for making hourly energy use predictions, ASHRAE Journal (1994) 72–81.

[28] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Applying support vector machine to predict hourly cooling load in the building, Applied Energy 86 (10) (2009) 2249–2256. doi:10.1016/j.apenergy.2008.11.035.

[29] B. Dong, C. Cao, S. E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy and Buildings 37 (5) (2005) 545–553. doi:10.1016/j.enbuild.2004.09.009.

[30] R. E. Edwards, J. New, L. E. Parker, Predicting future hourly residential electrical consumption: A machine learning case study, Energy and Buildings 49 (2012) 591–603. doi:10.1016/j.enbuild.2012.03.010.

[31] P. de Wilde, C. Martinez-Ortiz, D. Pearson, I. Beynon, M. Beck, N. Barlow, Building simulation approaches for the training of automated data analysis tools in building energy management, Advanced Engineering Informatics 27 (4) (2013) 457–465. doi:10.1016/j.aei.2013.05.001.

[32] J. Granderson, S. Touzani, C. Custodio, M. D. Sohn, D. Jump, S. Fernandes, Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings, Applied Energy 173 (2016) 296–308. doi:10.1016/j.apenergy.2016.04.049.

[33] F. Rossi, D. Velázquez, A methodology for energy savings verification in industry with application for a CHP (combined heat and power) plant, Energy 89 (2015) 528–544. doi:10.1016/j.energy.2015.06.016.

[34] H. X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews 16 (6) (2012) 3586–3592. doi:10.1016/j.rser.2012.02.049.

[35] S.-H. Cho, W.-T. Kim, C.-S. Tae, M. Zaheeruddin, Effect of length of measurement period on accuracy of predicted annual heating energy consumption of buildings, Energy Conversion and Management 45 (18) (2004) 2867–2878. doi:10.1016/j.enconman.2003.12.017.

[36] R. K. Jain, K. M. Smith, P. J. Culligan, J. E. Taylor, Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy, Applied Energy 123 (2014) 168–178. doi:10.1016/j.apenergy.2014.02.057.

[37] R. Saidur, N. A. Rahim, M. Hasanuzzaman, A review on compressed-air energy use and energy savings, Renewable and Sustainable Energy Reviews 14 (4) (2010) 1135–1153. doi:10.1016/j.rser.2009.11.013.

[38] Y. Zhang, Y. Yang, Cross-validation for selecting a model selection procedure, Journal of Econometrics 187 (1) (2015) 95–112. doi:10.1016/j.jeconom.2015.02.006.

Accepted Manuscript

ACCEPTED MANUSCRIPT

URL http://www.sciencedirect.com/science/article/pii/S0304407615000305
[39] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: International Joint Conference on Artificial Intelligence, Vol. 14, 1995, pp. 1137–1143. arXiv:arXiv:1011.1669v3, doi:10.1067/mod.2000.109031.

Accepted Manuscript