

# Using influencing factors and Multilayer Perceptrons for energy demand prediction

Kitty Boersma  
University of Twente  
P.O. Box 217, 7500AE Enschede  
The Netherlands  
k.boersma@student.utwente.nl

## ABSTRACT

Energy demand is rising, exhibiting more and more fluctuations, and smart grids need to be able to adjust accordingly. Therefore, an accurate way of predicting the energy consumption of a household is needed. In this research, the Pearson Correlation Coefficient is used to determine the effects of using internal and external influencing factors that influence the energy consumption of a household. These internal and external influencing factors are taken into account and are combined with existing and experimental knowledge about Multilayer Perceptrons. Next to that, two data resolutions are compared. The study found that using a 1-hour data resolution produces a more accurate prediction. Additionally, by using influencing factors, a possible manner of improving the accuracy of energy prediction is found. By these means, the research aims to aid future research on this topic.

## Keywords

Energy prediction, Multilayer Perceptron, Pearson Correlation Coefficient, Influencing features, Deep learning

## 1. INTRODUCTION

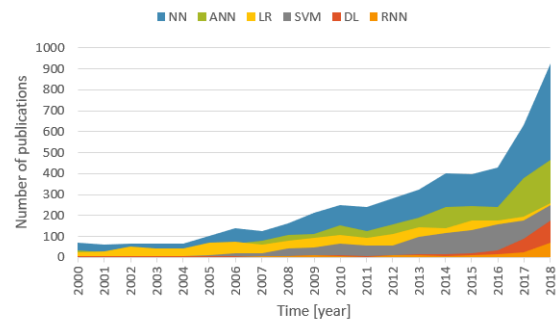
Nowadays, an ever-increasing amount of energy is consumed by residential buildings worldwide. On average, they consume about 40% of the global primary energy and this rate grows with 1.5% per year within Europe alone [21]. Consequently, the growth of urbanization and electricity demands asks for new requirements for future power grids. To satisfy the demands, power grids need to be able to predict, learn, schedule and monitor local energy production and consumption [14]. Additionally, to improve the flow of energy, energy predictions over various time horizons are needed when connecting residential buildings to future smart grids [15].

Energy consumption is difficult to predict, due to uncertainty of fluctuations. Fluctuations might be caused by the complexity of a building's energy producing and consuming technologies, or by unpredictable consumer behaviour. Other influencing factors can be found outside the physical building, such as the price of energy or the weather. Demand Response (DR) or Demand Side Man-

agement (DSM) programs can help keep fluctuations in energy use as low as possible. Modeling and predicting energy consumption can aid Demand Response or Demand Side Management programs.

Energy usage can be modeled as a time series, a value that changes over time, and can be predicted using many different methods. Predicting the value of this type of time series is challenging, given its highly non-linear character. Over the course of time, many different methods have been used to predict energy usage. Energy demand forecasting was also extensively pursued in the literature, mostly by applying various time series and machine learning methods. Some of these methods find their origin in the field of mathematics, such as Linear Regression (LR) [6] or ARIMA[5]. Other methods have a statistical background. These methods include Hidden Markov Models (HMM)[15] or Factorial Hidden Markov Models (FHMM)[10].

Deep learning was used for energy prediction from 2014 onward. At that moment, methods such as Conditional Restricted Boltzmann Machine (CRBM) [15] and factored Conditional Restricted Boltzmann Machine (FCRBM) [16] were introduced. Next to that, Long Short-Term Memory (LSTM) was used for building energy prediction in [23]. Recently, Artificial Neural Networks (ANNs) [9] and Support Vector Machines (SVMs) [4] became popular choices for the forecasting of energy consumption.



**Figure 1. A summary of the Scopus-indexed publications focusing on energy prediction since the beginning of the 21st century until now (i.e. 2000-2018)**

To provide a broader perspective, in Figure 1 an overview of the evolution of machine learning methods applied to energy prediction problems since the beginning of this century is presented. Deep learning is a relatively new concept and has clear advantages over traditional machine learning

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

31<sup>st</sup> Twente Student Conference on IT June. 30<sup>th</sup>, 2019, Enschede, The Netherlands.

Copyright 2019, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

methods [20]. In [16], several deep learning methods were found to be successful for energy prediction. Following that, in [20] a detailed study concerning Multilayer Perceptrons (MLPs) was conducted, proving that MLP could drastically improve the accuracy of building energy prediction. However, the optimal usage of MLP for this problem has not yet been found. Therefore, this research will try to improve MLP to increase the accuracy of building energy prediction.

## 2. RESEARCH QUESTION

This paper addresses the research question listed below. The main question can only be answered when the three subquestions have been answered.

RQ Can MLP be improved in such a way that it increases the accuracy of energy prediction?

RQ1 What are the benefits of using MLP when compared to other machine learning models?

RQ2 How can MLP be improved for energy prediction?

RQ3 What are the results of the new found method?

## 3. RELATED WORK

Machine learning has been studied for many years. Deep learning, however, was only introduced in 2006 by Bengio, Hinton and Le Cun [12]. The first mention of deep learning as a solution for energy prediction was in 2016 in [15]. Also in 2016, Long Short-Term Memory (LSTM) was applied to energy prediction of buildings in [13], as well as studies concerning building energy prediction using Conditional Restricted Boltzmann Machines (CRBM) and Factored Conditional Restricted Boltzmann Machines (FCRBM) in [17] and [16].

In 2017, the use of MLPs for energy prediction is first compared with the most commonly used machine learning methods, such as Support Vector Machines, Gaussian Processes, Regression Trees, Ensemble Boosting and Linear Regression in [20]. It was concluded that MLPs present better prediction accuracy, with higher accuracy in terms of RMSE and NRMSE, and therefore outperform these traditional machine learning methods in accurate and reliable prediction outcomes. However, there still exists a challenge in finding the optimal parameters for the MLP model. Next to that, the use of MLP in combination with influencing factors (i.e. features that influence the total energy usage of a building) in building energy prediction is mentioned in [18]. Although this study concludes Deep Belief Networks are a more accurate prediction method when presented with influencing factors, MLPs produce promising results over all.

This research uses the same approach, as it will apply MLP to predict energy used in buildings. The difference is that it will build upon the existing research to improve the accuracy of the predictions done by MLP. To the authors knowledge, making use of influencing internal and external factors while optimizing MLP through several parameters has not yet been researched with regards to energy prediction. A clear understanding of the use of MLP for building energy prediction might be able to satisfy the future demands of energy grids and inspire further research in this topic.

## 4. BACKGROUND

### 4.1 Supervised learning

In the field of machine learning, supervised learning is the task of learning a function that maps a given input to an output, based on example input-output pairs. The algorithm learns a function from so-called 'labeled training data', which consists of a set of training examples. The algorithm analyzes the training data and produces a function, which can then be used for the following training examples.

Ideally, the algorithm can correctly determine the function and is able to optimally predict unseen examples. This requires the supervised learning algorithm to generalize the training data to be able to adapt to unseen situations in the most "reasonable" manner.

### 4.2 Deep learning

Deep learning are representation-learning methods with multiple layers of abstraction. Based on the structure of the brain (like neural networks), a deep learning model consists of a multi-layer, interconnected network of neurons, each layer transforming the data to a higher, more abstract representation. Using sufficient layers, very complex functions can be learned.

In other words, deep learning allows for computational models that are composed of multiple processing layers to represent data with multiple levels of abstraction. The key aspect of deep learning is that the function or task of each layer is learned from the data, and thus not designed by human engineers. In 2006, deep learning was already capable of solving problems that the best of the artificial intelligence community could not crack [12]. Moreover, it turned out that deep learning is very good at discovering intricate structures in high dimensional data.

### 4.3 Multilayer Perceptron

MLPs [22] were first introduced in the 1980s as a machine learning solution for speech and image recognition, translation software, etc. However, Support Vector Machines (SVMs) introduced strong competition in the 1990s, since they were simpler and more effective. Since deep learning is gaining popularity, MLPs have found a renewed popularity.

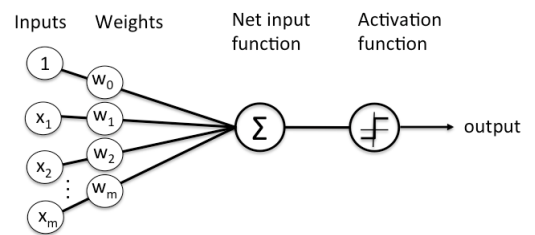


Figure 2. A perceptron

MLPs are made up of Perceptrons; a single neuron model from which large neural networks are derived (see Figure 2). MLPs consist of an input layer that uses neurons to represent the input data, an output layer that uses neurons to represent the output data, and an arbitrary number of hidden layers that use neurons to automatically discover features of the input data (see Figure 3). The layers of an MLP are connected consecutively, and any two consecutive layers are fully connected.

Each connection between two neurons is defined by a weight. This weight determines how significant the value is that is passed over the connection, by multiplying the value by the weight. Next to that, each neuron has an activation function that sums all the incoming values and creates an output value for the neuron. The passing of values through the network in a forward motion like this is called feed forward. Using this method, an MLP learns to model the correlation between the inputs and outputs.

Next to that, MLPs use back-propagation to re-calculate and update the weights used in the network. This allows the model to learn to become more accurate. When using supervised learning, the model can compare the predicted output to the expected output and use the error between the two to update its weights. This is done using an optimization function (Section 5.3.2). MLPs with one hidden layer are able to approximate any continuous function. In other words, it was proven that MLPs are universal function approximators [11]. This means that they can be used to model any kind of regression model.

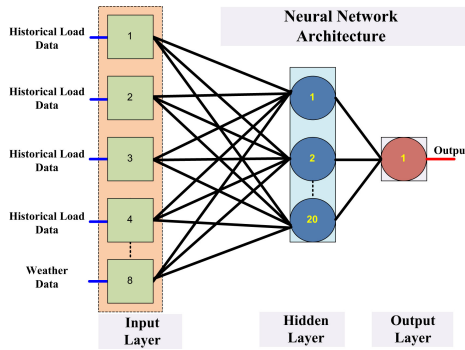


Figure 3. An example of the structure of an MLP

## 5. METHOD

### 5.1 Pecan Street Dataset

The Pecan Street [8] dataset is used, as it is the largest source of disaggregated customer energy data. Pecan street is located in Austin Texas and is part of research center on energy and water usage, amongst others, spanning multiple years. The Pecan Street database provides ac-

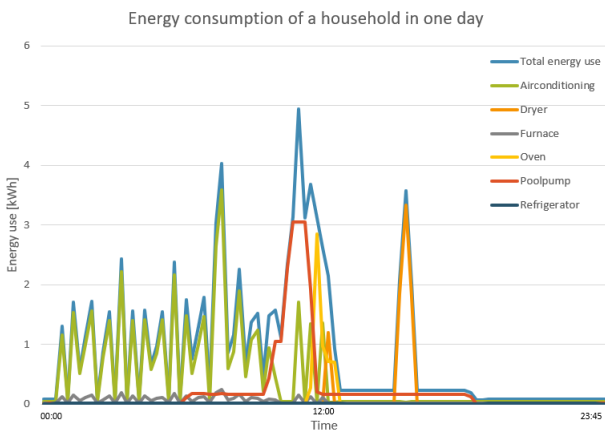


Figure 4. The energy consumption on January 7th, 2018, of the household used in this experiment, including individual appliances, with a resolution of 15 minutes.

cess to the energy usage of hundreds of individual households at one-hour, fifteen-minute and one-minute intervals recorded over several years. Next to that, the dataset provides both the total energy consumption of a household and the energy consumption of a single appliance (e.g. electric vehicle, dishwasher, etc.) or circuit (e.g. a combination of lights, fans and wall outlets) in kWh. An example of the energy consumption of a household, including the energy used by individual appliances, is displayed in Figure 4.

The Pecan Street dataset also provides data about external factors, such as the weather, energy price alerts and surveys. This research uses the data about the weather; specifically the temperature, the apparent temperature and the wind speed.

#### 5.1.1 Data

The total energy usage of an individual household over the course of several weeks is used as training data. Six weeks of data are used for the training of the model and one week is used for the testing of the model. Both data with 1 hour and 15 minute resolution is used, in order to compare the results. Next to that, feature selection (see Section 5.2) will determine which specific features are used as influencing factors.

### 5.2 Feature selection

Using a week of data as training data, the model uses 168 data points per feature with a 1-hour resolution and 672 data points per feature with a 15-minute resolution. The particular household that is used in this research has 6 internal influencing features and, additionally, four external influencing features can be used for training. Using all features can lead to a long processing time and a low efficiency of the model. Feature selection is used to select the features that influence the total energy use the most. These features are then used in the MLP. Using influencing features might increase the accuracy of the model and, at the same time, it reduces the dimensionality of the data. Dimensionality reduction is a process in which an  $n$ -dimensional vector is represented as an  $m$ -dimensional vector with  $m \ll n$ . Several approaches can be used for dimensionality reduction, like Principal Component Analysis (PCA) or Pearson Correlation Coefficient (PCC). This study uses PCC, since it was successfully used in [18] and [20].

The PCC that is defined later in Section 5.4 is used to identify influencing factors. Influencing factors are defined as factors that greatly contribute to the total energy usage (i.e. a circuit or appliance that uses much of the total energy used at that moment). This is done by identifying the influencing factors that have the highest PCC value with respect to the total energy usage.

#### 5.2.1 Internal features

Internal influencing features are single appliances or circuits that greatly influence the total energy consumption. To determine which features are influencing features, the PCC between the features and the total energy consumption is determined. If the PCC is higher than a set threshold, the features is deemed an influencing feature. The influencing feature is used as input for the MLP.

#### 5.2.2 External features

External features are features like the weather, energy prices, etc. In other words, they are features that are outside of a consumer's influence. To determine if external features influence the total energy consumption, the PCC between the feature and the total energy consumption is

calculated. Like internal influencing features, an external influencing feature is used as input for the MLP.

### 5.3 MLP and the back propagation algorithm

In Section 4.3, MLP was shortly explained. The supervised learning problem of MLP can be solved with the back-propagation algorithm. The back-propagation algorithm consists of two steps; the forward pass and the backward pass. The models internal learning parameters are used to compute the output based on the input of the model. The properties of MLP are analyzed and used to try and improve the method. The number of hidden layers and the number of neurons per hidden layer will vary, as it is part of the research.

#### 5.3.1 The forward pass - activation function

Each neuron uses an activation function to determine its output. A Rectified Linear Unit (ReLU) and its variations allow for faster and more effective training for deep neural architectures when compared to the sigmoid function or similar activation functions [19].

#### 5.3.2 Backward pass with SGD

In the second step, partial derivatives of the cost function with respect to the different parameters are propagated back through the network. An optimization function helps minimizing the error in the output. This study uses Stochastic Gradient Descent (SGD), as it was used many times before, for example in [20] and [18]. SGD is given by

$$\theta = \gamma\theta - \alpha\nabla_{\theta}J(\theta) \quad (1)$$

where  $\theta$  represents the weights of the connections in the model,  $\gamma$  represents the weight decay and  $\alpha$  the learning rate. Furthermore,  $\nabla_{\theta}J(\theta)$  represents the gradient descent, where  $J(\theta)$  is the error function (i.e. Mean Squared Error) and  $\nabla$  takes the partial derivative of the error function for each weight. The whole process is iterated until the weights have converged.

### 5.4 Metrics used for accuracy assessment

To evaluate the prediction method, various metrics are used. These metrics evaluate the error between the predicted output and the measured values. The root mean-square error (RMSE), given by the following, is used to display the error

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (2)$$

where  $N$  is the number of data samples,  $y_i$  is the input data and  $\hat{y}_i$  is the expected output data. The RMSE is then normalized to transform the error into a percentage. The NRMSE is given by

$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{(y_{max} - y_{min})} \cdot 100 \quad (3)$$

Furthermore, the Pearson Correlation Coefficient (PCC) is used to evaluate the similarities between  $y_i$  and  $\hat{y}_i$ . When a high positive correlation occurs the PCC approaches 1, while PCC approaches  $-1$  when a high negative correlation occurs. If there is little to no correlation, PCC approaches 0. The difference between the current  $y$  and

its related mean  $\mu_y$  is determined and multiplied by the difference between  $\hat{y}$  and its mean  $\mu_{\hat{y}}$ .  $\mathbb{E}$  indicates that the expected value of this multiplication is taken. The expected value is an indication of the long-term average value of repetitions of the same experiment. Next, the expected value  $\mathbb{E}$  is divided by a multiplication of the standard deviation of both  $y$  and  $\hat{y}$ . The PCC is given by the following

$$PCC = \frac{\mathbb{E}[(y - \mu_y)(\hat{y} - \mu_{\hat{y}})]}{\sigma_y \cdot \sigma_{\hat{y}}} \quad (4)$$

### 5.5 Implementation details - Libraries

The MLP model is created by TensorFlow [7], an open source framework developed by Google that is used to implement and train custom neural networks. Next to that, the Keras Deep Learning library [1] and Pandas[3], a data structures and data analysis tool for Python, are used. Pandas has build-in functionalities for time series. Furthermore, NumPy [2], a Python package for scientific computing, is used. The code used in this research can be found on Git<sup>1</sup>.

## 6. EXPERIMENT AND RESULTS

The data set obtained from the Pecan Street database was complete; there were no missing values or time stamps. The energy data of an individual household was extracted over eight weeks, leading to a total of 1344 data samples per factor for a 1-hour resolution and 5376 data samples for a 15-minute resolution. The data has a mean value of 1.33 kWh and a standard deviation of 1.71 kWh.

Table 1. List of Scenario's

		Time horizon	Resolution
Scenario 1	Energy data	1 day	1 hour
Scenario 2	Energy data	1 day	15 minute
Scenario 3	Energy data + internal factors	1 day	1 hour
Scenario 4	Energy data + internal factors	1 day	15 minute
Scenario 5	Energy data + external factors	1 day	1 hour
Scenario 6	Energy data + internal factors + external factors	1 day	1 hour

### 6.1 Range of experiments

Two main aspects are considered essential in order to define six different scenario's, namely the resolution and the use of influencing factors. Figure 1 shows a list of scenario's that were conducted to test the model, with and without internal and external factors, and with two resolutions. These were chosen with a 1-hour and a 15-minute resolution, as to evaluate the effect on accuracy. Moreover, Scenario 1 and Scenario 2 are used as a benchmark for Scenario's 3 until 6.

Scenario 1 and 2 look at the prediction capacity of MLP using just the energy data (i.e. total energy use) over a 1-hour and a 15-minute resolution. Scenario 3 and 4 use the energy data and the internal influencing factors, both over a 1-hour and 15-minute interval. Furthermore, Scenario 5 uses energy data and external factors over a 1-hour interval, and Scenario 6 uses energy data and both internal and external influencing factors over a 1-hour interval. Unfortunately, the data used for the external factors is

<sup>1</sup><https://gitlab.com/snippets/1869498>

only available in a 1-hour interval, and could not be extrapolated to create accurate data to be used as external influencing factors with a 15-minute resolution.

Table 2 refers to the results of the different scenario's. The number of hidden layers, the number of nodes and the number of epochs were changed during the experiments. Highlighted in blue are the best results for each scenario.

## 6.2 Building energy prediction

Scenario 1 and 2 (Table 2) seem to imply that the MLP creates the most accurate prediction with 4 hidden layers and at least 500 nodes per layer when no influencing factors are taken into account. Increasing the number of hidden layers or nodes in these scenario's does not increase the accuracy further.

### 6.2.1 Resolution

One of the goals of this experiment is to look at the difference in accuracy when using different resolutions for a prediction. Scenario 1 (1-hour resolution) and Scenario 2 (15-minute resolution) can be compared and Scenario 4 (1-hour resolution) and Scenario 5 (15-minute resolution) can be compared when looking for the most accurate prediction method.

When comparing the above-mentioned scenario's, it immediately stands out that the scenario's using data with a 1-hour resolution produce more accurate results when looking at the RMSE and the NRMSE. Interestingly, also the PCC value is much lower when using a 15-minute resolution.

### 6.2.2 Zero-correction

It stood out that the predicted energy values tend to dip below zero sometimes when working with data with a 1-hour resolution. Because it was assumed that the energy usage cannot be negative (i.e. we do not take into account energy sources that a household may have that provide energy to the grid), the following was done: if a value is predicted to be below 0, it is set to 0. The results of this zero-correction can be found in Table 2.

Applying a zero-correction to the prediction seems to increase the accuracy of the prediction. While using zero-correction, the best predictions for Scenario 1 are done by 3 hidden layers with 500 nodes per layer. Scenario 3 and 4, on the other hand, seem to make the best predictions using 3 or 4 hidden layers with 500 nodes per layer. An example of Scenario 4 with 3 hidden layers of 500 nodes per layer with zero-correction can be found in Figure 5.

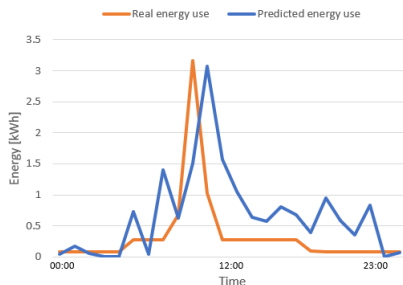


Figure 5. Example of a prediction of 24 hours of energy use by the MLP, with a 1-hour resolution and internal influencing factors

## 6.3 Building energy prediction and feature selection

The PCC that was proposed in 5.2 and defined in 5.4 is used to identify internal and external factors that potentially influence the energy demand profiles. Table 4 displays the statistical characteristics of each external factor, along with the PCC value of the different factors with respect to the total energy consumption. Additionally, Table 3 displays the same for the internal factors. Next to that, Figure 6 displays a correlation matrix between all the factors considered in the analysis.

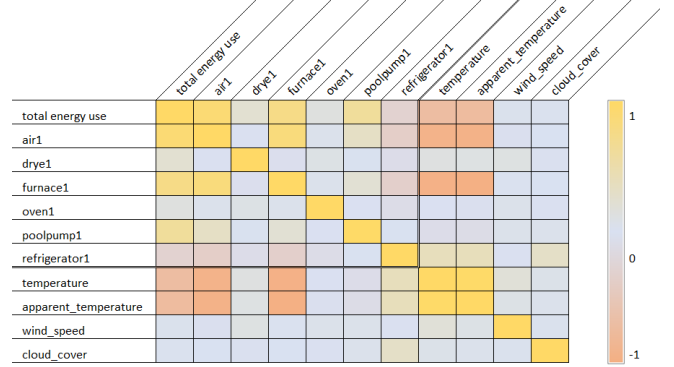


Figure 6. The PCC between multiple internal and external features obtained from the Pecan Street Database, using data of a 1-hour resolution

### 6.3.1 Internal features

Table 3 displays a high correlation for the internal factors *Furnace* and *Poolpump* (i.e. a PCC value above 0.5), and a very high correlation for the factor *Airconditioning*. These internal factors are deemed internal influencing features, as they seem to have a big effect of the total energy use. The factors *Dryer*, *Oven* and *Refrigerator* are disregarded, as their PCC value is close to 0.

The results of using the internal influencing features in the MLP model can be found in Table 2.

Looking at Scenario 1 and 4, it becomes clear that the results of Scenario 4 are only slightly better than the results of Scenario 1. The improvement in accuracy between these scenario's is not enough to draw a solid conclusion about the effects of using internal factors. However, the improvement in accuracy between Scenario 1 and Scenario 4 with zero-correction seems to be more significant. When looking at Scenario 2 and Scenario 5, both with a 15-minute resolution, there seems to be no improvement in accuracy whatsoever.

### 6.3.2 External features

Table 4 displays a low to average negative relation between the total energy use and the *Temperature* and between the total energy use and the *Apparent temperature*. This would imply that the total energy use decreases when the temperature increases. Also, both the *Temperature* and the *Apparent temperature* have a negative correlation with the *Airconditioning* and the *Furnace* (see Figure 6), implying that these external factors might influence the total energy use indirectly through their influencing on the internal influencing factors. The effects in accuracy of using one or both the external features in the MLP model can be found in Table 2.

Looking at Scenario 1 and Scenario 3, taking external factors (i.e. the *Apparent temperature* into account makes the MLP's predictions less accurate. It is only when using

Table 2. Day-ahead building energy prediction results, with accuracy expressed in terms of RMSE, NRMSE and PCC for experiments using data with one hour resolution (Scenario 1, 3 and 4) and 15 minutes resolution (Scenario 2 and 5). Number of neurons is counted per layer.

Scenario 3 (+ zero-correction) uses the influencing factor *Apparent temperature*, Scenario 4 (+ zero-correction) uses the influencing factors *Air conditioning*, *Furnace* and *Poolpump* and Scenario 5 uses the influencing factors *Air conditioning* and *Furnace*.

	RMSE	NRMSE	PCC	Hidden[#]	Number of neurons	Number of epochs	Number of factors
Scenario 1	1,02	16,10	0,60	1	500	200	
	1,00	15,87	0,59	2	500	200	
	0,98	15,47	0,67	3	500	200	
	0,97	15,42	0,69	4	500	100	
	1,02	16,13	0,66	5	500	100	
Scenario 2	1,84	21,14	0,32	1	500	100	
	1,62	18,52	0,37	2	500	100	
	1,62	18,55	0,44	4	2000	100	
	1,57	18,04	0,43	4	500	200	
	1,58	18,17	0,45	6	1000	100	
Scenario 3	0,95	14,98	0,60	2	1000	100	3
	0,94	14,94	0,61	3	500	100	3
	0,99	15,80	0,59	3	1000	100	3
	0,96	15,11	0,65	4	500	100	3
	1,01	15,93	0,67	5	500	100	3
Scenario 4	1,76	20,17	0,36	3	500	100	2
	1,63	18,74	0,44	4	500	100	2
	1,61	18,49	0,44	5	500	100	2
	1,61	18,49	0,46	6	1000	100	2
	1,60	18,36	0,48	8	500	100	2
Scenario 5	1,03	16,40	0,63	3	500	100	1
	1,08	17,12	0,61	3	1000	100	1
	1,08	17,06	0,64	4	500	100	1
	1,03	16,40	0,63	3	500	100	1
Scenario 6	1,04	16,52	0,60	3	500	100	4
	0,97	15,31	0,65	4	500	100	4
	1,05	16,64	0,65	5	500	100	4
	1,00	15,87	0,65	6	500	100	4
Scenario 1 + zero-correction	0,93	14,73	0,68	3	500	100	
	0,98	15,50	0,69	4	500	100	
	0,99	15,65	0,70	5	500	100	
Scenario 3 + zero-correction	0,95	14,98	0,52	2	500	100	3
	0,99	15,60	0,64	2	1000	100	3
	0,97	15,34	0,63	3	500	100	3
	0,95	15,07	0,65	4	500	100	3
	0,91	14,33	0,67	5	500	100	3
Scenario 5 + zero-correction	1,00	15,96	0,61	3	500	100	1
	0,96	15,21	0,67	4	500	100	1
	0,94	14,89	0,66	5	500	100	1
Scenario 6 + zero-correction	0,99	15,61	0,57	3	500	100	1
	0,99	15,81	0,67	4	500	100	4
	0,95	15,07	0,68	5	500	100	4
	0,99	15,81	0,65	6	500	100	4

**Table 3. Table containing statistical characteristics of some internal factors and the PCC value with respect to the total energy use using data of a 1-hour resolution**

Influencing internal factors	Min	Max	Mean	Std.dev	PCC
Energy [kWh]	0.04	8.56	1.33	1.71	1
Airconditioning [kWh]	0.0	4.91	0.80	1.20	0.900
Dryer [kWh]	0.0	2.71	0.11	0.39	0.239
Furnace [kWh]	0.01	1.58	0.10	0.16	0.769
Oven [kWh]	0.0	2.26	0.02	0.14	0.148
Poolpump [kWh]	0.01	3.32	0.39	0.63	0.622
Refrigerator [kWh]	0.01	0.07	0.01	0.01	-0.173

**Table 4. Table containing statistical characteristics of some external factors and the PCC value with respect to the total energy use using data of a 1-hour resolution**

Influencing external factors	Min	Max	Mean	Std.dev	PCC
Energy [kWh]	0.038	8.562	1.334	1.714	1
Temperature [F]	16.41	79.76	52.73	12.87	-0.408
Apparent temp. [F]	7.56	81.62	51.49	14.85	-0.427
Wind speed	0.05	17.83	5.97	3.50	0.048
Cloud cover	0.00	1.00	0.54	0.71	0.29

zero-correction that an improvement in accuracy can be seen.

## 7. CONCLUSION AND DISCUSSION

This paper presented the use of MLP for energy prediction of a household in combination with the use of internal and external influencing factors. It looked to improve the accuracy of the predictions using the influencing features and two different time horizons. Also, it proposed zero-correction as a way to improve accuracy in this specific case of energy prediction.

We see an outstanding difference in RMSE, NRMSE and PCC between Scenario 2 and 4, and the other scenario's. These results show that using data with a 15-minute resolution in this type of MLP produces less accurate results in terms of RMSE, NRMSE and PCC than using data with a 1-hour resolution. This could be caused by the fact that data with a 15-minute resolution has more data points, and has therefore a more accurate result than data with a 1-hour resolution. Next to that, it could be the case that data with a 15-minute resolution has overfitted more quickly, and has therefore reduced the accuracy. Moreover, Scenario 2 produces slightly better results in terms of RMSE and NRMSE when compared to Scenario 4, while Scenario 4 produces a slightly better PCC than Scenario 2.

Next to that, making use of influencing factors in the MLP seems to increase the accuracy of the prediction slightly at best, as can be seen when comparing Scenario 1, Scenario 3, Scenario 5 and Scenario 6, both with and without zero-correction. When looking at the scenario's without zero-correction, it stands out that only Scenario 3 than Scenario 1 in terms of RMSE and NRMSE. Both Scenario 5 and 6 seem to show little to no improvement when looking at the same metrics. In terms of PCC, there is no improvement in accuracy when comparing the scenario's. It is hard to draw any conclusions from these results, because Scenario 3, Scenario 5 and Scenario 6 tend to predict values below 0 sometimes. This decreases the accuracy immensely. Therefore, we should look at the results of these scenario's when using zero-correction.

When looking at the scenario's that use zero-correction, we

see an immediate improvement in all metrics when compared to the same scenario's without zero-correction. It seems that the accuracy increases slightly when using internal factors in the MLP, while external factors do not seem to influence the accuracy. This could be caused by the fact that the internal influencing factors influence the total energy consumption directly, so there is a strict correlation. Moreover, the correlation between the total energy consumption and the external influencing factors was lower than the correlation between the total energy consumption and the internal influencing factors, which could explain why internal influencing factors seem to have a greater influence on the accuracy. In other words, an influencing factor with a low negative correlation does not seem help the model to improve the accuracy.

## 7.1 Future work

Future work could include a research to the optimal resolution for energy prediction of buildings using MLP, as to research if a 1-hour resolution is the optimal resolution, and why certain resolutions produce a more accurate prediction than others in order to use this knowledge in possible applications. Next to that, future work could look into the effects of influencing features on energy prediction in more detail, especially to features that show a strong negative correlation with the total energy consumption, as to see if influencing features can improve predictions. Also, further research could take into account the correlation between influencing features amongst themselves, to see if that has an effect on the accuracy of a prediction. Furthermore, the benefits of using influencing factors could not only be in the accuracy of predictions, it could also improve the efficiency of the algorithm that is used. This could benefit the use of this method when applied in the real world.

## 8. REFERENCES

- [1] Keras.
- [2] Numpy.
- [3] Pandas.
- [4] M. Aydinalp-Koksal and V. I. Ugursal. Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied Energy*, 85(4):271–296, 2008.
- [5] R. A. Chinnathambi and P. Ranganathan. Investigation of forecasting methods for the hourly spot price of the day-ahead electric power markets. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 3079–3086. IEEE, 2016.
- [6] N. Fumo and M. R. Biswas. Regression analysis for prediction of residential energy consumption. *Renewable and sustainable energy reviews*, 47:332–343, 2015.
- [7] Google. Tensorflow.
- [8] P. S. Inc. Pecan street.
- [9] S. A. Kalogirou. Artificial neural networks in energy applications in buildings. *International Journal of Low-Carbon Technologies*, 1(3):201–216, 2006.
- [10] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han. Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the 2011 SIAM international conference on data mining*, pages 747–758. SIAM, 2011.
- [11] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2:359–366, 1989.

- [12] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [13] D. L. Marino, K. Amarasinghe, and M. Manic. Building energy load forecasting using Deep Neural Networks. *IECON Proceedings (Industrial Electronics Conference)*, pages 7046–7051, 2016.
- [14] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg. On-line building energy optimization using deep reinforcement learning. *IEEE Transactions on Smart Grid*, 2018.
- [15] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling. Comparison of machine learning methods for estimating energy consumption in buildings. *2014 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2014 - Conference Proceedings*, pages 1–6, 2014.
- [16] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling. Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks*, 6:91–99, 2016.
- [17] E. Mocanu, P. H. Nguyen, M. Gibescu, E. M. Larsen, and P. Pinson. Demand forecasting at low aggregation levels using Factored Conditional Restricted Boltzmann Machine. *19th Power Systems Computation Conference, PSCC 2016*, pages 1–7, 2016.
- [18] P. Mynhoff, E. Mocanu, M. G. t. I. P. Innovative, and undefined 2018. Statistical Learning versus Deep Learning: Performance Comparison for Building Energy Prediction Methods. *Researchgate.Net*, (September), 2018.
- [19] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.
- [20] N. G. Paterakis, E. Mocanu, M. Gibescu, B. Stappers, and W. van Alst. Deep learning versus traditional machine learning methods for aggregated energy demand prediction. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6. IEEE, sep 2017.
- [21] L. Pérez-Lombard, J. Ortiz, and C. Pout. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 2008.
- [22] F. Rosenblatt. *Perceptions and the theory of brain mechanisms*. Spartan books, 1962.
- [23] S. Ryu, J. Noh, and H. Kim. Deep neural network based demand side short term load forecasting. *2016 IEEE International Conference on Smart Grid Communications, SmartGridComm 2016*, pages 308–313, 2016.