Verification of Energy Efficiency Measures in Three Apartment Buildings Using Gaussian Process

Sakari Uusitalo [a]

[a] Faculty of Building services, Tampere University of Applied Sciences, Tampere, Finland, sakari.uusitalo@tuni.fi.

Abstract. The aim of this work was to examine whether the Gaussian process as a machine learning method is suitable for modelling time series data collected from buildings and whether it can be used to verify the effects of energy efficiency measures on three apartment buildings. A Gaussian process regression model was created using outdoor temperature and time information as inputs including information about the day of the week and the hour of the current day. Correspondingly, the output of the model was to estimate the hourly heating power demand corresponding to these inputs. The results provided by the created model were used as a reference point to verify the effects of energy efficiency measures taken on these residential buildings. The model was trained with 2016 hourly data. The 2017 data was used as test data to evaluate the functionality of the model. The impact assessment of the energy efficiency measures was performed with the measured data of 2019, which was compared with the results given by the model. Based on the performed modelling, it can be stated that using the Gaussian process, the need for hourly power of buildings was reasonably well modelled with even small amount of input variables. It can be assumed that the biggest uncertainty factor in the modelling is related to the domestic hot water consumption and the resulting power requirement. By measuring hot water consumption, modelling accuracy could probably be significantly improved. Based on the reviews, it could also be verified that the energy efficiency measures taken have had an impact on the peak power needs of residential buildings as well as on total energy consumption. For all three buildings, peak power needs appear to have decreased and overall energy consumption is lower than it would have been without the actions taken.

Keywords. artificial intelligence, machine learning, intelligent building, building services, energy efficiency, energy, power.
DOI: https://doi.org/10.34641/clima.2022.63

1. Introduction

In Finland energy efficiency measures in buildings are often targeted specifically at the heating system, but the challenge is to assess the impact of the actions taken as outdoor conditions and the occupancy rate of a building can vary considerably. To verify the impact of energy efficiency measures on the need for the heating power and energy consumption of a building, we need a model of the building. The model allows us to look at the situation where no measures has been taken and to compare the actual data with the results of the model.

We can use a computer program, based on dynamic model, to calculate the need for heating power at different outdoor temperatures. An example of such a simulation program is IDA ICE [1]. However, to build a model, we need a considerable amount of information about the properties affecting the dynamics of the building. For example, we need to find out the structures and areas of walls and windows, heat transfer coefficients, ventilation airflows, building use profiles and numerous other features affecting the need for heating capacity.

On the other hand, we may use a machine learning algorithm that is able to learn the dynamics of a building from the collected history data. For example, the algorithm can learn how the heating demand of a building depends on the outdoor temperature and time. In other words, we do not have to find out so much information about the building itself. It is enough that we know the outdoor temperature at any given time and the corresponding need for heating power.

This study examines how machine learning is suitable for modelling a building's heating power demand. The used algorithm is called Gaussian process, which is widely used in machine learning applications. The goal is to verify the effects of energy efficiency measures on three apartment buildings with student housing. Actual energy efficiency measures are not presented because they are secondary to work. However, they have been designed to reduce the peak power needs and overall energy consumption of the buildings. The subject of the comparison has been the hourly power needs of the buildings and the monthly and annual energy consumption. By comparing the hourly capacities, an attempt has been made to examine whether the measures taken have been able to cut the greatest power requirements of the buildings.

2. Gaussian process

In this section the basic idea of Gaussian process is introduced. For more complete definitions and descriptions please check out the book written by Rasmussen and Williams [2]. For visual introduction to Gaussian process, you should visit a web site authored by researchers from University of Konstanz [3].

2.1 Idea of a Gaussian process

Let $\mathbf{X}$ be a m $\times$ n -matrix in which each horizontal row $\mathbf{x}_\mathrm{i}$ represents one observation point. Each horizontal row item x_{ij} indicates the status of a particular property at that observation point. Each element $y_i = \mathrm{f}(\mathbf{x}_i) + \epsilon_i$ of the vertical vector $\mathbf{y}$, where ϵ_i is a measurement error, describes the observation of the phenomenon at the observation point. The purpose of the Gaussian process is to get to know the phenomenon under consideration by means of observations made at points and to find out the unknown events of the phenomenon at points $\boldsymbol{y}_*$ at points $\mathbf{X}_*$. Learning takes place by forming a conditional multidimensional normal distribution of the functions $\mathbf{f}$ (actually the values of the functions) suitable for the observations. The best estimate of the unknown events $\mathbf{y}_*$ of a phenomenon at the points $\mathbf{X}_*$ is the mean value vector $\boldsymbol{\mu}_*$ of that distribution. The aim of the process is not to find a function describing the actual phenomenon, but only the values produced by the function. [2]

In many cases, modelling tasks are more interested in the values produced by the function and their uncertainty than the function itself. [4]

2.2 Definition of a Gaussian process

According to the definition of Rasmussen and Williams [2], the Gaussian process is a collection of random variables, each of which follows a common normal distribution. The Gaussian process can be thought of as a generalization of the multidimensional normal distribution. Where a multidimensional normal distribution is a vector distribution, the Gaussian process is a distribution of functions, and the random variable is now the value of the function $\mathrm{f}(\mathbf{x})$ at the point $\mathbf{x}$. We can illustrate this with the follofing presentation:

$$\begin{pmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_n) \end{pmatrix} \sim \mathcal{N}\left(\begin{matrix} \begin{pmatrix} m(\mathbf{x}_1) \\ m(\mathbf{x}_2) \\ \vdots \\ m(\mathbf{x}_n) \end{pmatrix}, \\ \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, x_1) & \cdots & k(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix} \end{matrix} \right) \quad (1)$$

where $\mathrm{m}(\mathbf{x})$ and $\mathrm{k}(\mathbf{x}, \mathbf{x}')$ are a mean function and a covarince function (kernel). Rasmussen and Williams [2] have defined Gaussian process distribution as follows:

$$\begin{aligned} & f(\mathbf{x}) \sim \mathcal{GP}\big(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\big) \\ & m(\mathbf{x}) = E[f(\mathbf{x})] \\ & k(\mathbf{x}, \mathbf{x}') = cov[\mathbf{x}, \mathbf{x}'] \\ & = E\big[\big(f(\mathbf{x}) - m(\mathbf{x})\big)\big(f(\mathbf{x}') - m(\mathbf{x}')^{\mathrm{T}}\big)\big] \end{aligned} \quad (2)$$

2.3 Formation of a Gaussian process

In forming the Gaussian process, the covariance function plays an important role in determining the form of the function to be estimated. The covariance function looks at the distance (similarity) between two points $(\mathbf{x}, \mathbf{x}')$, connecting the different observations to each other. The selection of the appropriate covariance function plays a key role in the success of the modelling task. The most typical covariance functions and their properties are presented, for example, in [3] and [5]. Covariance functions often contain also free hyperparameters which should be optimized by an appropriate method using training data [2]. To form a Gaussian process, the covariance function must be computed between all possible points, giving three matrices: $\mathbf{K}_{\mathbf{XX}}$, $\mathbf{K}_{\mathbf{X}_*\mathbf{X}}$ and $\mathbf{K}_{\mathbf{X}_*\mathbf{X}_*}$. Each value of the function to be searched is normally distributed, so that a combined normal distribution [2]

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{K} & \mathbf{K}_*^T \\ \mathbf{K}_* & \mathbf{K}_{**} \end{pmatrix}\right) \quad (3)$$

can be formed from the training and estimated observations assuming that the mean function $m(\mathbf{x}) = 0$. The most probable values $(\boldsymbol{\mu}_*)$ of the function $f(\mathbf{x}_*) = y_*$ at the estimated points $(\mathbf{X}_*)$ can now be found by forming a conditional distribution [2]

$$p(\mathbf{y}_* \mid \mathbf{X}_*, \mathbf{X}, \mathbf{y}) \sim \mathcal{N}\begin{pmatrix} \mathbf{m}_* + \mathbf{K}_*\mathbf{K}^{-1}(\mathbf{y} - \mathbf{m}_*), \\ \mathbf{K}_{**} - \mathbf{K}_*\mathbf{K}^{-1}\mathbf{K}_*^T \end{pmatrix} \quad (4)$$

By assumption $\mathbf{m}_* = \mathbf{0}$, the best estimate of values $\mathbf{y}_*$ obtained by the function $\mathbf{f}_*$ is now a mean vector of the distribution

$$E[\mathbf{y}_*] = \boldsymbol{\mu}_* = \mathbf{K}_*\mathbf{K}^{-1}\mathbf{y} \quad (5)$$

and the uncertainty of that estimate is described by the variance of the distribution

$$var[\mathbf{y}_*] = \mathbf{K}_{**} - \mathbf{K}_*\mathbf{K}^{-1}\mathbf{K}_*^T. \quad (6)$$

Figure 1 shows an example of a situation where $x_* \in [-5,5]$ and the goal is to find out the values $\mathbf{y}_*$ obtained by the function $f(x_*)$ in that range. The left side of the figure shows the situation when no observations have been made of the function values and the right side of the figure shows the situation when six noisy observations have been obtained. The solid bold line describes the best estimate of the values obtained by the function in both cases. The dashed line in right side is the real function to be estimated and the grey area is the 95 % confidence interval obtained from the variance of the distribution.

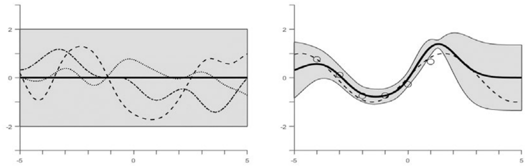

Figure 1 - Three samples of possible functions, when there are no observations about the function values and an estimated function when there is six observation points about the function values.

3. Verifying energy efficiency measures

In this section the used data and main results is presented.

3.1 Used data

The heating power demand in a building depends on the temperature difference between the indoor and outdoor temperatures. The hourly averages of the Finnish Meteorological Institute's Tampere Härmälä measuring station were used as the outdoor temperature. Wind strength can also have a large effect on heating power demand in certain situations, but no detailed information on wind strength was available for these explicit buildings, so it was decided to exclude it. The heating power required to heat domestic water depends, of course, on how much water is used and when. In principle, therefore, this may be entirely incidental. However, it is assumed that there is some correlation between the need for heating power in terms of the day of the week and the time of day. This assumes that the routines of apartment dwellers, such as washing and cooking, will remain reasonably similar. However, it is precisely the assessment of the power required to heat domestic water that involves the greatest uncertainty.

The measures taken on buildings date back to 2018 and possibly 2019. Data from 2016, 2017 and 2019 were used in the study. Data from 2016 have been used as a training data while 2017 has been used as a test year. As the systems and use of the buildings have been very similar in 2016 and 2017, the modelled data should correspond quite well to the data measured for 2017. Measures have been taken in the buildings during 2018. In 2019, the measures have had an impact throughout the year. So, when comparing the modelled and measured data for 2019, it should show the effects of the measures taken in terms of reduced peak power and energy consumption.

The outdoor temperature and time data, which included information about the day of the week and the hour of the current day, were used as inputs for the model built in the study. The output of the model was an estimate of the hourly average heating power demand corresponding to these inputs. Since time is cyclical, time information must be transformed into a form from which the used algorithm also understands its cyclical nature. Thus, the algorithm must in practice understand, for example, that the last hour of the day is close to the first hour of the following day. Therefore, the input used as a time information has been converted to cyclic using sin and cos transformations.

Before the actual construction of the model, the data were cleaned by removing e.g., clearly erroneous readings as well as moments of time that lacked relevant information such as outside temperature or heating power information. Many times, in modelling tasks like this, the data is collected from many different sources, in which case it may be incomplete in some respects and contain erroneous readings. Making data usable is often one of the biggest and most time-consuming jobs before the actual analysis and modelling work.

3.2 Building the Gaussian process model

The GPstuff function library for Matlab, made by Vehtari et al. [6], was used to build the Gaussian process model. In these studies, the addition of the squared-exponential and periodic covariance function was used. Additivity allows strong assumptions to be made about the individual components that make up the sum [5]. Other variations of the covariance function were also tested but were not found to have a significant effect on the modelling results. However, it is good to note that using the right kind of covariance function or a combination of them can have a significant impact on the success of the modelling.

3.3 Error review

Table 1 shows the RMSE values of the models for 2017 monthly. The RMSE values describe the distance between the prediction and the observation vector and cannot in themselves be used to draw conclusions about the goodness of the model. In this study, RMSE values were mainly used to define the Gaussian process model to be used when testing the effect of different covariance functions, and their combinations, on the modelling results. However, the table 1 shows that the RMSE values for building 1 are the smallest. Building 2 values are slightly higher and building 3 values are clearly higher than the other two buildings. This is natural, as the hourly outputs

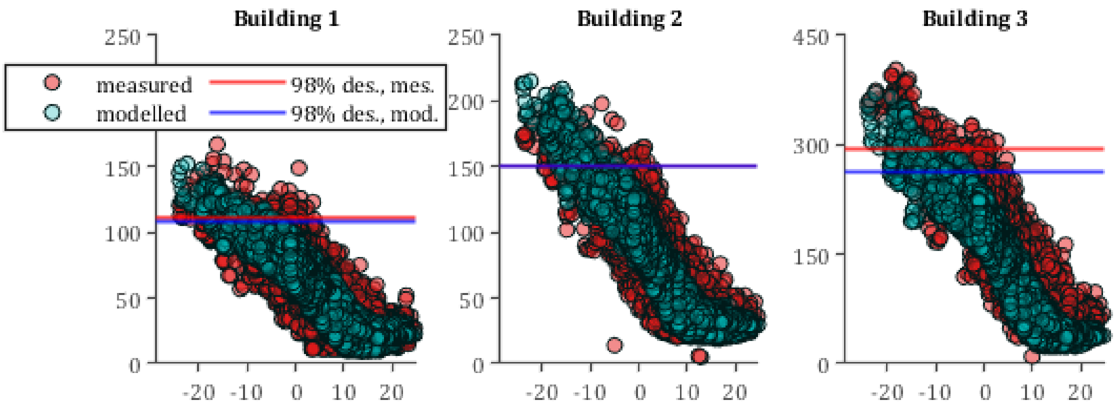

Figure 3 - Modelled and measured hourly powers of buildings from 2017.

themselves are higher for buildings 2 and 3. Building 3 also differs considerably in the actual power demand from buildings 1 and 2. Maximum power demand at building 3 is about 400 $\frac{kWh}{h}$, while buildings 1 and 2 have maximum power requirement slightly below and above 200 $\frac{kWh}{h}$. In this case, relatively equal errors in the modelling may appear as a larger absolute error in those buildings where the power range is also larger.

Table 1 - Year 2017 model RMSE-values.

	Building 1	Building 2	Building 3
jan	10.5	13.5	32.8
feb	10.6	16.0	27.9
mar	10.3	12.9	25.1
apr	12.3	16.5	27.7
may	12.4	16.5	34.5
jun	8.7	9.5	25.9
jul	6.4	7.0	19.6
aug	6.7	7.9	29.9
sep	10.6	12.1	49.2
oct	11.0	13.6	46.4
nov	11.0	17.6	40.5
dec	12.4	15.8	34.0
year	10.4	13.7	34.3

The figure 2 shows the distributions of the residual values of the 2017 models. The distribution shows that the expected values of the residual values of buildings 1 and 2 ($\mu_1 = -0.57$ and $\mu_2 = -1.50$) are close to zero, which can be considered as one of the properties of a successful regression model. Instead, the residual values of object 3 are slightly weighted to the right of zero ($\mu_3 = 22.40$), suggesting that the modeled values would appear to be smaller than the measured values.

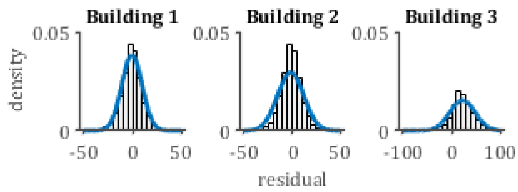

Figure 2 - Distributions of residuals between 2017 measured and modelled values.

3.4 Hourly power comparison

The figure 3 shows the modelled and measured hourly power of buildings in relation to the outdoor temperature in 2017. The figure also shows the lines indicating the level below 98% of all hourly power values. As can be seen from the figure for buildings 1 and 2, for both the modelled and measured values, the lines are very close to each other. This suggests that the modelling has been reasonably successful and in 2017 the technical systems and users of the building have performed similarly as in 2016. Instead for building 3 the line of the modelled values is lower than the measured values, so the modelled values would appear to be lower than the measured ones. This can also be clearly seen by looking at the points. The points produced by the model would appear to compress to the left edge of the point cloud of measured values.

The same comparison for 2019 is shown in figure 4. The figure shows two things. First, the image also shows how some of the actual hourly power points of the harshest frosts have dropped lower, forming their own small point cloud. In addition, the 98% limit for measured values has dropped lower in all three buildings. These findings support the claim that actions taken to the buildings have succeeded in limiting the maximum hourly powers.

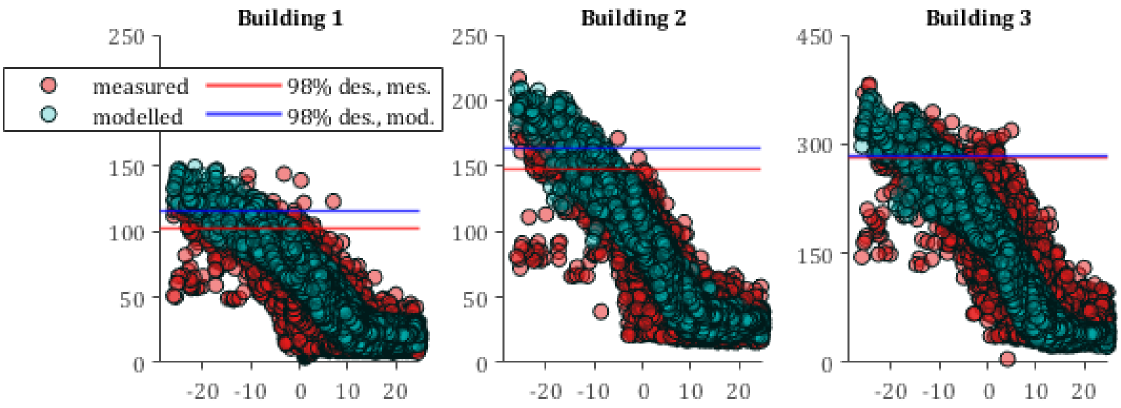

Figure 4 - Modelled and measured hourly powers of buildings from 2019.

In addition to limiting the maximum power requirement, the actions taken in the buildings also aimed at lower energy consumption. Figure 5 shows the monthly energy consumption of buildings for 2017 and Figure 6 for the whole year. It can be seen from the figures that the measured and modelled consumptions correspond quite well to each other, but small deviations occur, especially at the monthly level. For building 3, the differences are larger and follow the same regularity as when comparing hourly power needs.

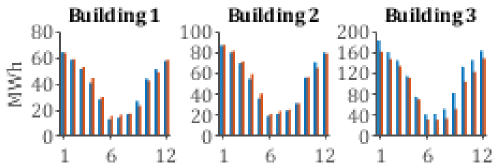

Figure 5 – Measured (blue) and estimated (orange) monthly energy consumption of buildings in 2017.

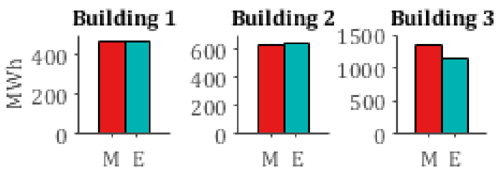

Figure 6 - Measured (M) and Estimated (E) annual consumption of buildings 2017.

Figures 7 and 8 show the same thing for 2019. Figure 7 shows that the measured consumption is lower than the modelled ones every month, except for building 3. However, for building 3 it can also be seen that the difference between the measured and modelled consumption has narrowed compared to 2017. Of course, the effect is also visible on an annual basis, and the figure 8 shows that for buildings 1 and 2, the actual total energy consumption in 2019 is significantly lower than the modelled one. For building 3, the difference between actual and modelled consumption has narrowed since 2017.

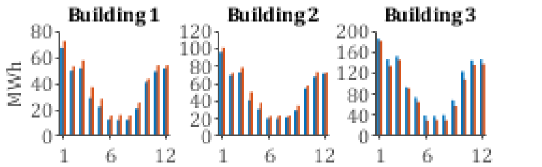

Figure 7 – Measured (blue) and estimated (orange) monthly energy consumption of buildings in 2019.

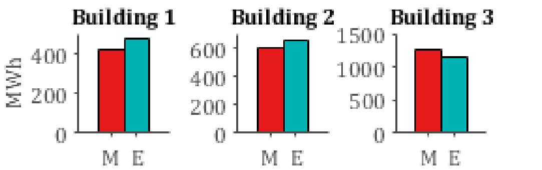

Figure 8 - Measured (M) and Estimated (E) annual consumption of buildings 2019.

4. Discussion

Based on the modelling performed, it can be stated that the method succeeded in modelling the hourly power demand of buildings with even small amount of input variables. It can be assumed that the biggest uncertainty factor in the models is related to the domestic hot water consumption and the resulting power requirement. By measuring hot water consumption, modelling accuracy could probably be significantly improved. On the other hand, information about the position of the hot water heating control valve is already often available in building automation system. Also, taking this information as one additional variable could improve the accuracy of the modelling.

It can also be said that the energy efficiency measures in buildings have had a positive effect both in limiting the maximum hourly power needs and in reducing overall energy consumption. Although the effects are clear, their absolute magnitude, especially on an hourly basis, is difficult to assess. To make hourly absolute impact assessments, the model should be more accurate, which could be achieved, for example, with the hot water consumption data described before. Regarding the annual consumption of the model, a reasonably good estimate can be made of the actual energy savings of buildings 1 and 2. In contrast, for building 3, there was more uncertainty in the modelling. This may be due, for example, to changes in the building in question that were not known and had an impact on power demand in 2017. On the other hand, modelling for this building also suggests that measures have been taken during 2018 reduced overall energy consumption compared to the situation without measures.

Machine learning can have significant applications in the analysis of building operations and the development of more intelligent control methods. In addition to the impact assessment of energy efficiency measures discussed in this work, potential applications could also include model-based predictive control, maintenance cost forecasting based on utilization rate, fault diagnosis by identifying various deviations, and clustering-based intelligent control. However, these are just a scratch on the surface of potential applications, and the construction industry also offers considerable opportunities for innovation in the use of artificial intelligence and machine learning. However, the implementation of various methods and applications still requires considerable further research and development. There are many questions that still need to be answered. What kind of information is relevant? How is the information collected and who owns it? What are the possibilities of different methods and what limitations do they have? How to get methods into the production? And many others.

5. References

[1] EQUA Simulation AB. IDA Indoor Climate and Energy URL: https://www.equa.se/en/ida-ice (referred 3.1.2022).

[2] Rasmussen, Carl Edward., and Christopher K. I. Williams. Gaussian Processes for Machine Learning. Cambridge, Mass: MIT Press, 2006. Print.

[3] J. Görtler, R. Kehlbeck ja O. Deussen. A Visual Exploration of Gaussin Processes. URL: https://www.jgoertler.com/visual-exploration-gaussian-processes/ (referred 3.1.2022).

[4] Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol: O'Reilly Media, Incorporated, 2017. Print.

[5] Duvenaud, David. Automatic model construction with Gaussian processes. Dissertation. 2014.

[6] Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen V., and Vehtari, A.. GPstuff: Bayesian Modeling with Gaussian Processes Journal of Machine Learning Research. 2013; 14(Apr):1175-1179.

Data Statement

The datasets generated during and analysed during the current study are not publicly available because the data is owned by the local energy company and the housing associations but will be available anonymously on request.