Understanding the performance gap: a machine learning approach on residential buildings in Turin, Italy

Roberto Boghetti¹, ², *, Fabio Fantozzi¹, Jérôme H Kämpf², Giacomo Salvadori¹

¹ Università di Pisa, Department of Energy, Systems, Territory and Constructions Engineering, Pisa, Italy
² Idiap Research Institute, 1920 Martigny, Switzerland

*roberto.boghetti@gmail.com

Abstract. Buildings account for the highest share of primary energy usage and greenhouse gas emission in the E.U. and U.S. [1], and most of this energy is used for space and water heating. Being able to gain a broader understanding of the gap between predicted and in situ measured thermal performance of buildings may, in a lot of cases, help reducing the energy consumption and, therefore, alleviating our pressure on the environment [2]. The aim of this research is to further investigate this performance gap and to evaluate the possibility of using machine learning algorithms to effectively predict the energy demand of buildings. For this purpose, a group of residential buildings in the city of Turin, Italy, is taken as case study: an estimation of their yearly heating demand is made using different machine learning algorithms, and their results are evaluated and discussed. The research showed that the use of machine learning resulted in a performance gap in line, if not lower, with the current literature. The reasons for this outcome, as well as possible future research directions are finally discussed.

1. Introduction

The design phase of a building is a crucial step that highly influence its future behaviour and being able to make accurate predictions is vital for a conscious planning. There is, however, a well-documented mismatch between the energy demand predicted during the design phase and the as-built measured performance. This is often referred to as “performance gap” [3]. While the existence of this gap is widely accepted by the scientific community, a definitive agreement on its magnitude has not been found yet. Given the complexity of the problem and the lack of a standardization of the evaluation process, the results of different measurements are in fact often impossible to compare. Nonetheless, there is a consistent number of studies on the topic that prove that this gap is a common issue and that its magnitude can reach significant levels [4].

While no correlation can be easily established between the magnitude of the performance gap and the commonly used building parameters, the extensive research done on the topic has shed some light on the underlying factors that influence the problem. Among the main causes, the uncertainty associated with the human behaviour is often indicated as the principal driver of this discrepancy. Furthermore, when looking at the problem from an urban-scale point of view, there is often a lack of building-scale details in the available data that is usually filled with assumptions that can contribute to increase the error of the model. The use of machine learning algorithms could help in both cases as it would automatically take these uncertainties into account, removing the need for the user to create behavioural models and to make assumptions on the missing data.
The aim of this paper is to analyse the performance gap between predicted and measured yearly heating demand of buildings at urban scale, and to investigate the possible advantages of using machine learning driven models over traditional methods. While there exists a consistent body of literature on the subject, a standard methodology has not been agreed upon by the scientific community, and the research panorama is vastly inhomogeneous: for this reason, the exploration of different methods, as well as different case studies, can represent a valuable addition to the ongoing discussion.

2. The performance gap
“Performance gap” is a broad term that generally refers to a deviation between the expected and actual performances and can describe several different situations. In the case of buildings, it can be used for different aspects of their life cycle, from air quality to CO₂ emissions. In the context of this paper, the term is used in its most common acceptation, that is to denote a mismatch in the energy consumption of the buildings. In particular, the research is focused on analysing the energy used for space and water heating as these usually represent the most significant share of the total energy consumption. The heat consumption is investigated over a period of a year, and therefore the results are given as aggregated yearly demand.

3. Previous researches
Given the central importance of energy efficiency in the present days and the promising results that machine learning methods are demonstrating, an extensive research is being conducted on the topic. A comprehensive list of recent studies in this regard is given by Amasyali et al. [5]. As there is a significant number of parameters that define the problem, however, several different approaches can be taken, effectively reducing the list of comparable studies. Nonetheless, this great diversity is indispensable to understand how the different algorithms can perform under different circumstances. An important step in this direction is to use different models on the same problem and to compare their performances. Both Robinson et al. [6] and Deng et al. [7] used this approach for estimating the energy consumption of commercial buildings in the US, showing different results.

4. Methodology
A group of 386 residential buildings in the city of Turin (Piedmont, Italy) was taken as case study. The database contained 2D GIS data of the buildings as well as their year of construction, height and number of floors. As one of the goals of the research is to investigate if acceptable estimations can be obtained using easily available urban scale information, no other data regarding the buildings characteristics was gathered from other sources. Finally, the actual heating demand of each building, referred to the season 2011/2012, was provided by the energy provider.

A machine learning model was built using Python and tested on the case study. Multiple algorithms were evaluated in the process using the Scikit-learn module [8]. To avoid making wrong assumptions or using wrong statistical models for the human behaviour, only the known data was used as input, leaving the uncertainties to the model itself. The results were then evaluated, and the resulting performance gap contextualized and discussed.

5. Database overview
The database was provided by the Politecnico di Torino as a shapefile, a vector data storage format that is commonly used for geographic information systems (GIS) software. It contains data for 58970 buildings, whose geometries are stored as a 2D footprint. For the scope of this paper, however, only 386 residential buildings were considered due to the limited availability of real consumption data. The distribution of their height, footprint, measured heating demand (target variable) and period of construction is given in Figure 1. It can be noted that buildings in the sample generally have considerable dimensions: the median values of the footprint area, height and gross volume are respectively 308.51m², 21.39m and 5992m³. The buildings are also relatively old: more than 85% of them were built before
1971, with a concentration of data in the period 1961-1970. No information was provided on possible renovations.

Figure 1. Distribution plots of the input data for coverage, height, heating demand and period of construction.

6. Model overview
To test the performance of a machine learning approach, a completely automated model was written in Python. Once chosen the database, the input data is processed in order to extract meaningful features and to scale their values. The database is then split into a training and a validation set. Several algorithms are tested on the validation set, using different setup combinations for each of them. Meaningful information about their precision is finally provided.

6.1. Data preparation
During this phase, meaningful features for each building are extracted from the input database. These can be divided into three sets: building-scale, neighbourhood-scale and urban-scale features. The first set is comprised of information that solely depends on the geometry of the building, while for the other two the surrounding environment is also considered: in these cases, a circular area around the centroid of the building is taken. For the neighbourhood-scale, the radius of this area corresponds to the maximum length of the shadow that can be cast within the city at noon during the year, as this scale is used for taking into account solar gains. Furthermore, as directionality has an important impact in this case, the area is also divided into sectors of equal dimension, whose number depends on the density of the city. Regarding urban-scale features, whose characteristics were taken from the work of Carrozza et al. [9], the radius of the area was chosen to be three times bigger, in order to grasp the effects of heat islands and urban canyons. A representation of the two areas is shown in Figure 2 while a list of all the features used in the model is given in Table 1. For linear models, the distributions are then normalized to enhance performances. Finally, the features are scaled using the Robust Scaler and the database is split into a training and a validation set: the proportion used in this work was 66.6% for the training and 33.3% for the validation, resulting in respectively 258 and 128 buildings.
Table 1. List of the features used in the machine learning model.

| Description                        |
|------------------------------------|
| **Building-scale**                 |
| Aspect                             | Ratio between the sides of the bounding box of the building. |
| Bound ratio                        | Ratio between the footprint and the bounding box of the building. |
| External surface                   | Total external surface of the walls and roof. |
| Footprint area                     | Total area of the building footprint. |
| Form factor                        | Ratio between the perimeter and the area of the building footprint. |
| Gross volume                       | Total volume of all interior spaces of the building, calculated as 80% of the occupied volume. |
| Height                             | Median height of the building. |
| Number of floors                   | Number of floors of the building. |
| Orientation                        | Orientation of the main façade of the building expressed in degrees, counting counter clockwise from north. |
| Perimeter                          | The perimeter of the footprint. |
| Period of construction             | Period of construction of the building. |
| **Neighbourhood-scale**            |
| Shadowed portion n                 | Ratio between the height of the highest shadow that can be cast from the direction n on the façade of the building and the height of the building. The division of the area is shown in Figure 2. In this case, the area was divided into 20 sectors. |
| **Urban-scale**                    |
| Average building height (ABH)      | Average height of the buildings in the urban area. |
| Aspect ratio (AR)                  | Ratio between the height of the building and the mean width of roads. |
| Building coverage ratio (BCR)      | Ratio between the urban area and the total surface that is occupied by the footprints of buildings. |

Figure 2. Areas of influence for the neighbourhood-scale (a) and urban-scale (b) features.
7. Results
Given the impossibility to know \textit{a priori} which machine learning algorithm will perform better [10], different models were chosen, tuned and tested. The best results that every algorithm achieved in terms of mean absolute percentage error (MAPE) are shown in Table 2.

Table 2. List of algorithms and their precision in terms of MAPE.

| Algorithm                          | MAPE  | Algorithm          | MAPE  |
|------------------------------------|-------|--------------------|-------|
| Linear regression                  | 44.27%| Gradient boosting  | 32.33%|
| Ridge regressor                    | 43.85%| Random forest      | 31.95%|
| Elastic Net                        | 43.68%| KNeighbors regressor| 34.60%|
| Lasso                              | 44.27%| Bagging regressor  | 36.87%|
| Support vector regression          | 27.32%| Extra trees regressor| 35.09%|

The model that scored the highest precision is Support Vector Regression (SVR) with RBF kernel, \( C=1.0 \) and \( \text{epsilon}=0.1 \). The MAPE on the validation set was 27.32\%, with a peak error of 101.89\%. The correlation between measured (x-axis) and calculated (y-axis) heating consumption values is shown in Figure 3, along with the results of a simulation (uncalibrated) on the same data using the simulation software CitySim [11].

![Figure 3](image)

\textbf{Figure 3.} Precision of the model (left) and of CitySim (uncalibrated, right) for comparison.

A stronger correlation is noticeable in the first graph, indicating that the machine learning model estimations are generally closer to the actual energy demands than those made with CitySim, where the performance gap is wider.

7.1. Features importance

![Figure 4](image)

\textbf{Figure 4.} Features weights according to the Random forest algorithm.
To evaluate the importance of the input features, the weights assigned by the Random forest algorithm were investigated. As shown in Figure 4, the urban-scale features are all within the 10 most influential predictors, while regarding the shadowed portions the southern sectors are the most important for the algorithm.

8. Discussion and conclusion
While the case study was not representative enough to generalize the results, the research raised several interesting points. The error of the model, even with little input information, was in line with the literature values of the performance gap. It is likely that this approach could have mitigated some of the uncertainties that are proven to be the main causes for the performance gap, especially those that are associated with the human behaviour. In this regard, it is interesting to note that these buildings where the model had the worst performances are all located in the same neighbourhood, Crocetta, one of the most prestigious and central areas of the city. The higher income of the families that live in that neighbourhood can be among the possible reasons for this discrepancy, as it might be associated with a wider difference between user behaviours, and, therefore, could be a meaningful feature to consider if available.

9. Possible research directions
As both the problem and the presented model have a high degree of complexity, further researches on the topic can be focused on several different aspects. A valuable approach, for example, could be focusing on improving the model in order to gain new insights on the performance gap. In this regard, a deeper study on the impact of different features on the final result could be an important step to better understand their connection with the gap. Furthermore, finding new meaningful features, especially those related with urban morphology, and quantifying their impact on the heating demand of buildings, can also help improving the precision of other models. An optimization of the presented model is also possible through a more conscious split of the database, where the testing buildings are picked in a controlled way, with the goal of creating a more representative set from which the algorithm can learn.

References
[1] Cao X, Dai X and Liu J 2016 Energy and buildings 128 pp 198-213
[2] Acaravci A and Ozturk I 2010 Energy 35 pp 5412-20
[3] Menezes A C, Cripps A, Bouchlaghem D and Buswell R 2012 Applied energy 97 pp 355-64
Van Dronkelaar C, Dowson M, Burman E, Spataru C and Mumovic D 2016 Frontiers in Mechanical Engineering 1 p 17
Norford L K, Socollow R H, Hsieh E S and Spadaro G V 1994 Energy and buildings 21 pp 121-31
De Wilde P 2014 Automation in Construction 41 pp 40-9
[4] Shi X, Si B, Zhao J, Tian Z, Wang C, Jin X and Zhou X 2019 Sustainability 11 p 937
[5] Amasyali K and El-Gohary N M 2018 Renewable and Sustainable Energy Reviews 81 pp 1192-205
[6] Robinson C, Dilkina B, Hubbs J, Zhang W, Guhathakurta S, Brown M A and Pendyala R M 2017 Applied energy 208 pp 889-904
[7] Deng H, Fannon D and Eckelman M J 2018 Energy and Buildings 163 pp 34-43
[8] Pedregosa et al JMLR 12 pp 2825-30
[9] Carozza M, Mutani G, Coccolo S and Kaempf J H 2017 Proc. of the 3rd IBPSA-Italy Conf
[10] Wolpert D H and Macready W G 1997 IEEE transactions on evolutionary computation 1 pp 67-82
[11] Robinson D, Haldi F, Leroux P, Perez D, Rasheed A and Wilke U 2009 InProceedings of the Eleventh International IBPSA Conference 2009 pp 1083-90