Neural network-based energy signatures for non-intrusive energy audit of buildings: Methodological approach and a real-world application

Neural Network-based Energy Signatures for Non-Intrusive Energy Audit of Buildings: Methodological Approach and a Real-World Application

Simone Eiraudo[a], Daniele Salvatore Schiera[a], Lorenzo Mascali[a], Luca Barbierato[a], Roberta Giannantonio[b], Edoardo Patti[a], Lorenzo Bottaccioli[a], Andrea Lanzini[a]

[a]*Energy Center Lab, Politecnico di Torino, name.surname@polito.it, Torino, 10138, Italy,*
[b]*TIM S.p.A, name.surname@telecomitalia.it, Milan, 20123, Italy,*

Abstract

Energy Signatures (ES) are highly informative gray-box regression models. Thanks to their simplicity and interpretability, to their data-driven approach, and to their effectiveness in describing a building response to external weather variables, they are employed for *i)* the determination of the Balance Point (BP) temperature of a building, *ii)* the ranking of the efficiency of heating or cooling systems, *iii)* the provision of building diagnostic information, and *iv)* the development of strategies for more energy-efficient buildings and the estimation of potential savings. In this work, we propose an innovative energy audit tool, based on a Feed-Forward Neural Network (NN) to determine ES from aggregated (meter-level) electric load profiles of buildings. Multiple NN-based regression models are defined for each building and compared to provide the most accurate and informative one, by considering proper fit and significance indexes. This allows the eventual existence of multiple cooling regimes to be detected. Moreover, the energy audit methodology defines and applies an innovative Key Performance Indicator (KPI), called Temperature Unstandardized Beta Weight (β^*_{Temp}), to account not only for the thermal behavior of buildings but also for the efficiency of the conditioning system and the internal heat generation. This ES approach has been applied to a dataset of electric consumption patterns from about eighty industrial buildings from a telecommunication (TLC) service provider in Italy. The useful outputs from the proposed methodology, together with its simplicity, effectiveness and applicability, are intended to support the diffused understanding of the thermal behavior of buildings and the analysis of their inefficiencies, in order to enhance energy retrofit actions and reduce consumption.

Keywords: Building Energy Efficiency, Energy Audit, Energy Signature, Neural Network

Acronyms

BP	Balance Point
COP	Coefficient of Performance
DC	Data Center
EP	Elbow Point
ES	Energy Signature
HVAC	Heating, Ventilation and Air Conditioning
KPI	Key Performance Indicator
ML	Machine Learning
MLP	Multi-Layer Perceptron
NN	Neural Network
ReLU	Rectifier Linear Unit
RMSE	Root Mean Squared Error
TLC	Telecommunication
UPS	Uninterruptible Power Supply

1. Introduction

In recent years, building energy efficiency has been among the most topical issues of the energy research community [1]. However, many obstacles hinder the pathway toward a more environmentally sustainable way of consuming energy in buildings, such as the lack of awareness about their energy behavior and its related inefficiencies. Although making a minor contribution to the final demand of the building sector as late as a few decades ago, the fast-growing power consumption from space cooling has now become a major concern. Indeed, it represents 18.5 % of the total electrical demand from buildings worldwide [2]. Despite the rise in efficiency in cooling systems and the buildings' energy footprint reduction policies, this component of their energy consumption has more than tripled over the last thirty years, and it continues to grow relentlessly.

The way forward to enhance awareness of the thermal behavior of buildings and identify their inefficiencies, such as those pertaining to cooling loads, is to provide practical tools for the application of building energy footprint analysis and evaluation. Classic physical models are widely adopted for the performance assessment of buildings [3]. These models, the so-called white-box models, enhance the analysis of cause-and-effect relationships and provide retrofit and renovation scenarios [4]. Vargas et al. [5], for instance, took advantage of the EnergyPlus simulation engine to estimate the energy-saving potentials associated with different energy efficiency actions. However, these methods require extensive information regarding the building structure, such as its thermal layout, its operation and occu-

pancy schedules, etc. Other approaches are data-driven ones, namely black-box models, which are becoming more and more popular s they offer the possibility of covering a wider spectrum of building analysis. These approaches require lower computational and modeling efforts and, in particular, they do not require information about the thermal layout of buildings. However, even without this fundamental information, they are able to provide accurate results and enhance the understanding and performance analysis of heating and cooling systems. In this context, Hwang et al. [6] employed Machine Learning (ML) to determine the most adequate time intervals for predicting the energy performance of educational buildings. The outcome of their analysis aimed to support the choice of an optimal heating and cooling control strategy.

Gassar et al. [7] highlighted the advantages and shortcomings of white and black-boxes, and stressed that hybrid approaches (i.e. gray-box models) may represent an effective alternative to analyze large dataset for building energy prediction applications. Hybrid approaches can be used to overcome the limits of the previous methods as they join the interpretability of the models with the employment of a data-driven approach, thereby enhancing the development of very accurate energy prediction models. Asaee et al., for instance, employed a hybrid approach in [8] to exploit the combined strengths of engineering modelling and a Neural Network (NN) to estimate the end-use energy consumption within buildings, including lighting, heating, and space cooling.

Among the building energy audit techniques, Energy Signatures (ES) are gray-box data-driven approaches that can be used to point out the dependence of heating or cooling energy consumption on the weather conditions [9]. ES are temperature-at-use regression models, whose results are more appropriate for buildings' energy consumption analysis with respect to those approaches focusing solely on temporal patterns. ES could provide highly informative outcomes by enhancing the extraction of the qualitative characteristics of buildings [10]. The simplest case of ES is the univariate regression model that considers only outdoor air temperature and energy consumption. Another solution that can be adopted to improve the model accuracy of ES is that of applying multivariate models which can be designed to analyze the impact of other weather variables, such as solar radiation. This approach was adopted in [11], where Tronchin et al. employed linear multivariate ES to obtain parameters from the regression model that described the thermal behavior of buildings, considering different weather variables (i.e. temperature, global solar radiation).

Furthermore, both linear and non-linear regression models can be employed in ES. In [12], a linear regression model was employed in order to analyze the total heat loss coefficients of thousands of Italian buildings. Nageler et al. [4] instead applied a non-linear ES regression model, specifically a sigmoid one, to enhance the accuracy of energy audits. They validated their model on a case study consisting of residential buildings and an office, obtaining slight deviations from the real values.

Besides their widespread usage in residential buildings, in the last few decades, ES have also been adopted for commercial and industrial purposes *i)* to rank the efficiency of buildings heating systems [12], *ii)* to produce diagnostic information, benchmarks, and control charts [13], *iii)* to characterize buildings, with the aim of planning building energy retrofit actions [14], and *iv)* to calculate the Balance Point (BP) temperature of buildings [15]. The latter represents the outdoor air temperature that determines a thermal load equal to zero, that is, the weather condition at which the building does not require heating or cooling in order to keep the indoor air temperature within a given range. In turn, computing the BP of a structure provides crucial information about the building envelope and its thermal behavior. For this reason, BP has been exploited by researchers to analyze the potential of ventilative cooling [16] and to enhance the accurate estimation of degree days in order to support effective building-energy policies [17] and to predict the energy consumption of Heating, Ventilation and Air Conditioning (HVAC) systems [18].

Most of the existing model-based approaches to calculating BP are somewhat challenging and time-consuming methodologies [13, 19]. For these reasons, in this paper, we propose a novel energy audit methodology that exploits Feed-Forward NN to accurately and automatically determine the ES of buildings from the aggregated load consumption profile. Unlike most of the works that deal with ES, our methodology allows meter-level electric consumption measurement to be used instead of the heating or cooling systems measurements, thereby avoiding the intrusive installations of sensors for the metering of disaggregated loads. On the one hand, this represents an element of complexity, concerning the accurate determination of the ES of a building. On the other hand, the non-intrusiveness of the proposed methodology represents a major advantage to apply it to real case scenarios, as many buildings are not provided with adequate sensors for disaggregated load metering purposes. Moreover, this methodology applies a robust and simple model that allows the BP and performance of a cooling system to be estimated by means of the Temperature Unstandardized Beta Weight (β^*_{Temp}), a Key Performance Indicator (KPI) that takes into account both Total Heat Loss Coefficient and the Coefficient of Performance (COP) of the cooling system. The aim of this work has been to make novel contributions to both the tools employed for ES analysis and to the proposed energy KPI. Furthermore, this methodology can be used to enhance: *i)* the detection of inefficient sites, by means of a comparison of buildings with a similar usage category, *ii)* the identification of abnormal power consumption with regard to the response of a building to outdoor weather variables, and *iii)* the estimation of the best strategies for energy efficiency of buildings, that is, providing retrofit scenarios. With respect to our previous work [20], the key contributions and novelties proposed in this manuscript can be summarized as follows:

1. A generalized form of the thermal balances and of the proposed methodological framework is presented. The proposed ES approach can be employed for all those buildings that are provided with electric air conditioners (i.e., compression chillers, heat pumps, etc). This new approach, With respect to the shortcomings of the approaches in the literature and to the aim of widespread energy audit tech-

niques, is intended to enhance the application of ES by only considering aggregated load consumption data and the outdoor temperature;
2. Several NN-based regression models are applied to the whole dataset to detect the eventual presence of multiple cooling regimes. These multiple operative regions depict different responses to the thermal load. Hence, the methodology allows the energy KPI (e.g. COP) to be estimated for each single region;
3. A sigmoid-activated perceptron is introduced to handle the eventual presence of discontinuities regarding the measured energy demand values. This is an additional feature that is able to deal with steps in the load demand of buildings in order to provide the most accurate regression model for different HVAC systems;
4. Appropriate fit and significance scores have been defined to determine the best regression model for each building, considering its accuracy and interpretability. These scores have been defined *i)* on the basis of the widely adopted coefficient of determination R^2, which provides an accuracy metric for each regression model, and *ii)* with a proper significance metric α, to focus on each single cooling regime.

The present work is structured as follows. Section 2 describes the fundamentals of ES and discusses the mathematical steps necessary to introduce the proposed data-driven ES approach. Section 3 instead describes the methodological pathway and the fundamental tools that were developed with a general perspective of application to the existing building stock. Then, Section 4 presents the investigated real-world case study, an application of the proposed data-driven ES approach, and introduces some additional hypotheses regarding the specific case study. Section 5 presents the experimental results and discusses the application of the methodology to the proposed case study. Finally, Section 6 reports the final remarks and outlines future developments.

2. A Novel Data-driven Energy Signature Approach

The analysis of the energy consumption of a building is generally tackled through two main approaches, namely *i)* time-of-use and *ii)* temperature-at-use. The main shortcoming of the time-of-use approach regards its limited usefulness in respect of thermal analysis [10]. On the other hand, a temperature-at-use approach may enlighten and more effectively identify the causes of inefficiencies in those buildings where the thermal loads make a major contribution to the total consumption. For this purpose, it could be useful to analyze the consumption pattern of a building with respect to the outdoor temperature, T_{ext}, that is, to distinguish between those consumption quotas that depend on external weather variables and independent ones, as described in Equation 1:

$$P_{TOT}(T_{ext}) = \sum P_{dep}(T_{ext}) + \sum P_{ind} \tag{1}$$

where P_{TOT} is the overall building consumption; P_{dep} are the power contributions that depend on the outdoor temperature;

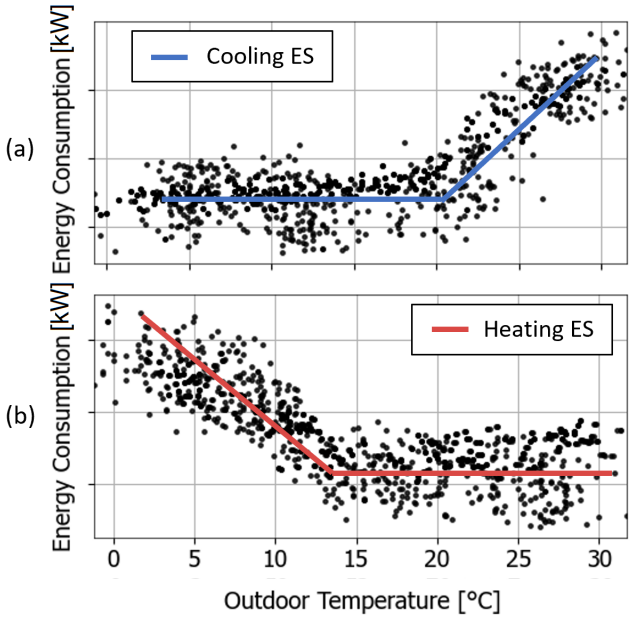

Figure 1: Two qualitative examples of Energy Signatures (ES): (a) a cooling ES, and (b) a heating ES

and, finally, P_{ind} are the power contribution that are independent of the outdoor temperature.

Equation 1 represents a univariate model, where the outdoor temperature, T_{ext} is considered as the only input variable and the total consumption, P_{TOT}, is the output. Such a univariate model is generally referred to as Energy Signature (ES) and is commonly represented by means of a scatter plot (known as the energy signature of the building) in which the mean outdoor air temperature and the mean energy consumption are displayed on the x-axis and y-axis, respectively. As shown in the qualitative examples in Figure 1, two regions can be distinguished in a common ES: *i)* an unconditioned region where the load is not affected by the outdoor temperature, and *ii)* a conditioned region, where the building consumption is affected by variations in the outdoor temperature, thereby showing a particular slope coefficient. Figure 1a depicts a typical cooling ES (blue line), which may be derived from the energy consumption values retrieved from a building that features air conditioning but not heating. Figure 1b instead shows a typical heating ES (red line), whereas the conditioned region coincides with the lower temperature sub-domain and features the highest consumption values as the lowest outdoor temperatures occur.

It is useful to introduce the thermal balance of a generic building to obtain a clear insight into ES. A reduced form of this balance equation may be obtained by making the assumption that the envelope of the building is in thermal equilibrium with the outdoor environment [21]:

$$\phi_T + \phi_V + \phi_{Sol} + \phi_{Cond} + \phi_{St} = 0 \tag{2}$$

where ϕ_T is the thermal power transferred through the building envelope; ϕ_V accounts for the heat exchanged with the environment through ventilation; ϕ_{Sol} is the heat gain determined by

solar radiation; ϕ_{Cond} is the thermal power of the air conditioning system; and, finally, ϕ_{St} is the internal heat generation.

For the sake of brevity and with reference to the proposed use case application, the following equations refer to the case of cooling ES. It should be pointed out that, by taking care of the sign convention employed in the thermal equations, an analogous framework can easily be derived for the case study of heating ES.

First, the heat fluxes due to the conditioning systems, ϕ_{Cond} and ϕ_V, respectively, are lumped into a single contribution, ϕ_{CLC}, that is:

$$\phi_{CLC} = \phi_{Cond} + \phi_V \tag{3}$$

Next, two fundamental characteristic parameters of buildings, namely the Temperature Sensitivity, k_{TOT}, and the COP are introduced. k_{TOT} represents the coefficient of the slope of an ES that depicts the thermal power transferred through the envelope on the y-axis and the outdoor temperature on the x-axis, and it is defined as:

$$k_{TOT} = \frac{\phi_T}{T_{ext} - T_{in}} \tag{4}$$

where T_{in} and T_{ext} formally represent the indoor and outdoor air temperature, respectively. The COP is defined by the following Equation:

$$COP = -\frac{\phi_{CLC}}{P_{CLC}} \tag{5}$$

where P_{CLC} is the electrical consumption of the cooling system. In accordance with the commonly adopted sign convention, where a heat flux from a zone to the environment is negative, a negative sign appears in Equation 5 to provide positive COP values. Thereafter, the cooling system's consumption quota may be expressed in relation to the variables from the thermal balance Equation 2 and to the outdoor temperature T_{ext}. Indeed, considering Equations 2, 3, 4 and 5, P_{CLC} can be expressed as:

$$P_{CLC} = \frac{\phi_{St} + \phi_{Sol} + k_{TOT} \cdot (T_{ext} - T_{in})}{COP} \tag{6}$$

It is worth noting that the electrical load of the cooling system can only be positive. Similarly, the thermal contribution of the cooling system to the thermal balance can only be negative. Hence, the previous equation is only valid if:

$$k_{TOT} \cdot (T_{ext} - T_{in}) + \phi_{St} + \phi_{Sol} > 0 \tag{7}$$

otherwise, P_{CLC} is equal to 0.

The highest outdoor temperature that determines a thermal load equal to 0 is referred to as the Balance Point temperature, T_{BP}. This can be easily calculated by turning the above inequality into its corresponding equation, as follows:

$$T_{BP} = T_{in} - \frac{\phi_{St} + \phi_{Sol}}{k_{TOT}} \tag{8}$$

In many buildings, the indoor temperature is a controlled variable, which is referred to as Set Point Temperature, T_{SP}. In these cases, the following may be easily derived:

$$k_{TOT} \cdot (T_{BP} - T_{SP}) + \phi_{St} + \phi_{Sol} = 0 \tag{9}$$

Finally, a general piecewise function of the cooling system consumption is derived as:

$$P_{CLC}(T) \begin{cases} 0 & \text{if } T_{ext} <= T_{BP} \\ \frac{k_{TOT} \cdot (T_{ext} - T_{SP}) + \phi_{St} + \phi_{Sol}}{COP} & \text{if } T_{ext} > T_{BP} \end{cases} \tag{10}$$

In the end, by considering Equations 8 and 10 it is found that:

$$P_{CLC}(T) \begin{cases} 0 & \text{if } T_{ext} <= T_{BP} \\ \frac{k_{TOT} \cdot (T_{ext} - T_{BP})}{COP} & \text{if } T_{ext} > T_{BP} \end{cases} \tag{11}$$

It should be noted that, from a mathematically rigorous point of view, T_{BP} is a variable, that can vary slightly, depending on the internal heat generation, ϕ_{St}, and the heat gain determined by solar radiation, ϕ_{Sol}. Hence, the BP value can be considered as a constant parameter of the building in most of the cases. By adopting this simplifying assumption, it may be found that:

$$P_{CLC} = f(T_{ext}) \tag{12}$$

that is, the cooling load only depends on the outdoor air temperature, T_{ext}. The fairness to adopt the hypothesis above should be tested by validating the independence of the cooling load from ϕ_{Sol} and ϕ_{St}. This may be done, for instance, by considering an autocorrelation analysis, as reported at the beginning of Section 5.1. The latter formulation represents a fundamental contribution to the literature approaches to ES. Indeed, most of the previous research focused on deriving ES by considering the thermal fluxes. In this regard, ES are commonly designed as regression models that may be represented by a scatter plot that features thermal power on the y-axis and the outdoor temperature on the x-axis. In these cases, the slope of the conditioned region is k_{TOT}, which is often measured in $kW_{th}/°C$. Conversely, in those cases where the envelope surface is considered, k_{TOT} is measured in $kW_{th}/(°C * m^2)$. Nevertheless, measurements that account for thermal power may not be available in many real-world buildings. Meantime, electricity smart meters are being deployed at an impressive pace. Indeed, 34 % of the metering points in Europe had already been equipped with smart meters by 2018, with a total of over 100 million monitoring devices. This value is expected to exceed 90 % by 2030 [22]. Hence, electricity consumption data regarding almost the whole European building stock will be available in just a few years. For these reasons, Equation 11 indicates opportunities for a wider application of ES to the existing building stock. In this regard, the presented theoretical framework holds true for most buildings, as electricity is the energy vector of the vast majority of the conditioning systems.

2.1. KPI Definition

In the perspective of enhancing the present non-intrusive energy audit approach, appropriate KPI should be employed to describe an ES that exploits electrical load consumption measurements. An important novelty of this paper is the introduction of a new KPI, which is here referred to as Temperature Unstandardized Beta Weight β^*_{Temp}. This indicator is defined as the slope of the normalized rise in the electrical load with respect to the internal heat generation, ϕ_{St}, per Celsius degree, considering temperatures above the BP. Hence, its unit of measure is $°C^{-1}$. So, β^*_{Temp} takes into account not only the thermal behavior of the building envelope, which is described by k_{TOT}, but also the cooling system efficiency and the heat generated within the building. This KPI is defined as:

$$\beta^*_{Temp} = \frac{k_{TOT}}{COP \cdot \phi_{St}} \tag{13}$$

It should be pointed out that k_{TOT} is a constant value, as it is a characteristic parameter of the considered building envelope. The COP instead deserves more attention, since, in some cases, its value may vary slightly depending on the outdoor temperature, T_{ext}. For the aim of this analysis and for the sake of simplicity, the COP is assumed constant. Moreover, a variety of cooling devices, featuring different COP, may exist within the same cooling system, and intervene simultaneously or according to the adopted cooling system control strategy. In the case of multiple conditioning regions (i.e. cooling regimes), the COP is considered constant for each region detected by the proposed methodology. The equation used to estimate the COP value, for any conditioning regime, can be derived considering Equations 13 and 9 as:

$$COP = \frac{1}{\beta^*_{Temp} \cdot (T_{SP} - T_{BP})} \tag{14}$$

Then, by introducing β^*_{Temp}, the piecewise Equation reported in 11 becomes:

$$\frac{P_{CLC}(T)}{\phi_{St}} \begin{cases} 0 & \text{if } T_{ext} <= T_{BP} \\ \beta^*_{Temp} \cdot (T_{ext} - T_{BP}) & \text{if } T_{ext} > T_{BP} \end{cases} \tag{15}$$

In such a formulation of ES, the novel KPI represents the slope coefficient of the scatter plot of the electrical load, normalized by the internal heat generation, on the y-axis, and the outdoor temperature on the x-axis.

In some cases the activation of the cooling system may determine a load step. A constant quota $\Delta P_{act,CLC}$ is added to account for this additional consumption step, at the interval of temperatures corresponding to the conditioned region. temperature range determining the activation of the cooling system, that is if T_{ext} overcomes T_{BP}. Moreover, as this methodology is intended to enhance the detection of multiple cooling regimes, the previous Equation may be generalized by considering different temperature ranges, each with their own β^*_{Temp}. Given the existence of r cooling regimes, the lower boundary temperature that divides the non-conditioned region from the first cooling regime is here referred to as BP, while the successive boundary temperatures, which distinguish two different cooling regimes are referred to as Elbow Points (EPs). The Equation in 15 can therefore be generalized to any number of cooling regimes, by considering different temperature ranges and β^*_{Temp} values, as follows:

$$\frac{P_{CLC}(T)}{\phi_{St}} \begin{cases} 0 & \text{if } T_{ext} <= T_{BP} \\ \beta^*_{Temp,1} \cdot (T_{ext} - T_{BP}) + \Delta P_{act,CLC} & \text{if } \begin{cases} T_{ext} > T_{BP} \\ T_{ext} < T_{EP,2} \end{cases} \\ \quad ... & \\ \beta^*_{Temp,r} \cdot (T_{ext} - T_{BP}) + \Delta P_{act,CLC} & \text{if } T_{ext} > T_{EP,r} \end{cases} \tag{16}$$

where T_{EP} are the EP temperatures.

In the case where the cooling load is the only contribution to the consumption that depends on the outdoor temperature, Equation 16 may be shifted upwards by considering the independent load contributions P_{ind}, introduced in Equation 1 to find the total aggregated load of the building, as follows:

$$\frac{P_{TOT}(T)}{\phi_{St}} \begin{cases} \frac{P_{ind}(T)}{\phi_{St}} & \text{if } T_{ext} <= T_{BP} \\ \frac{P_{ind}(T)}{\phi_{St}} + \Delta P_{act,CLC} + \\ + \beta^*_{Temp,1} \cdot (T_{ext} - T_{BP}) & \text{if } \begin{cases} T_{ext} > T_{BP} \\ T_{ext} < T_{EP,2} \end{cases} \\ \quad ... & \\ \frac{P_{ind}(T)}{\phi_{St}} + \Delta P_{act,CLC} + \\ + \beta^*_{Temp,r} \cdot (T_{ext} - T_{BP}) & \text{if } T_{ext} > T_{EP,r} \end{cases} \tag{17}$$

This formulation results to be the best to employ for non-intrusive energy audit purposes. Indeed, the aggregated load is the most trivial measurement in buildings. Moreover, the outdoor temperature is often available as well or it can be accurately obtained from weather data providers. However, there is still another obstacle to the application of the novel approach to ES, namely the estimation of the internal heat generation. This issue needs to be addressed specifically for each case study. In our application, it was tackled by means of the domain expertise reported in Section 4.

3. Methodology

This Section introduces the adopted methodological approach, from raw data processing to the achievement of the ES results and their interpretation. The overall methodology was conceived as a general ML-based approach for the energy audit of buildings, with a focus on deriving the energy signature of non-residential buildings. Hence, Section 3.1 presents the methodological approach in its general form, as a support for the application of the energy audit to any building sector (i.e. residential, commercial, or industrial). Section 3.2 details

the design and development of the multiple Feed-Forward NN-based regression models for ES, which represents one of the main contributions of this manuscript. Hereafter, the employed notation regarding the ML models is compliant with that reported in [23].

3.1. Methodological Framework

The proposed methodology exploits aggregated meter-level electrical load measurements and the outdoor temperature to retrieve information about the thermal behavior of a non-residential building. As described in Figure 2, it can be used to perform raw data pre-processing, analysis, exploitation, and interpretation. The methodological steps: *i)* the *Pre-Processing*, including the filtering, re-sampling, and normalization procedures; *ii)* the *Machine Learning Workflow*, which is aimed at providing the best architecture and hyperparameters settings for the NN regression models, and *iii)* the *Post-Processing*, that consists of achievement, visualization, and analysis of the experimental results, are described in the following paragraphs.

3.1.1. Pre-Processing

It is necessary to ensure that the raw input data of the power consumption measurements extracted from the dataset are reliable, usable, and exploitable in the subsequent methodological steps, as they are shown in Figure 2. The Pre-Processing pathway is mainly composed of the following blocks: *i)* the *Data-Filtering*, *ii)* the *Re-Sampling*, and *iii)* the *Normalization*.

Data-Filtering. The raw dataset is filtered in order to remove any possible unreliable data and/or outliers. Such data may be caused by measurement errors or by anomalies in the electrical consumption pattern. This step enhances the optimal training of the NN and improves the overall accuracy of the energy audit. The task of detecting abnormal load values is carried out by means of a simple gradient-based statistical approach, which deletes single load values from time series whenever they feature an excessive gradient with respect to the previously recorded data. Specifically, the distribution of the gradients within the single load time series is considered and those values featuring a higher gradient than three times the standard deviation of the distribution from the load profile itself are filtered out.

Re-Sampling. The filtered dataset is then re-sampled to calculate the mean electrical load of each day, as this can represent an optimal set to determine the ES [12]. It is worth noting that considering fine-grained power consumption data may lead to a sharp decrease in the accuracy of the energy audit. Indeed, the thermal capacity of buildings causes their response to outdoor temperature changes to be non-instantaneous.

Normalization. This allows a comparison to be made among different buildings. This step is crucial to achieve two main goals of the proposed methodology, namely the comparative analysis and the provision of retrofit scenarios. Indeed, the awareness of building energy efficiency should be supported by comparing the thermal behavior of the building with that

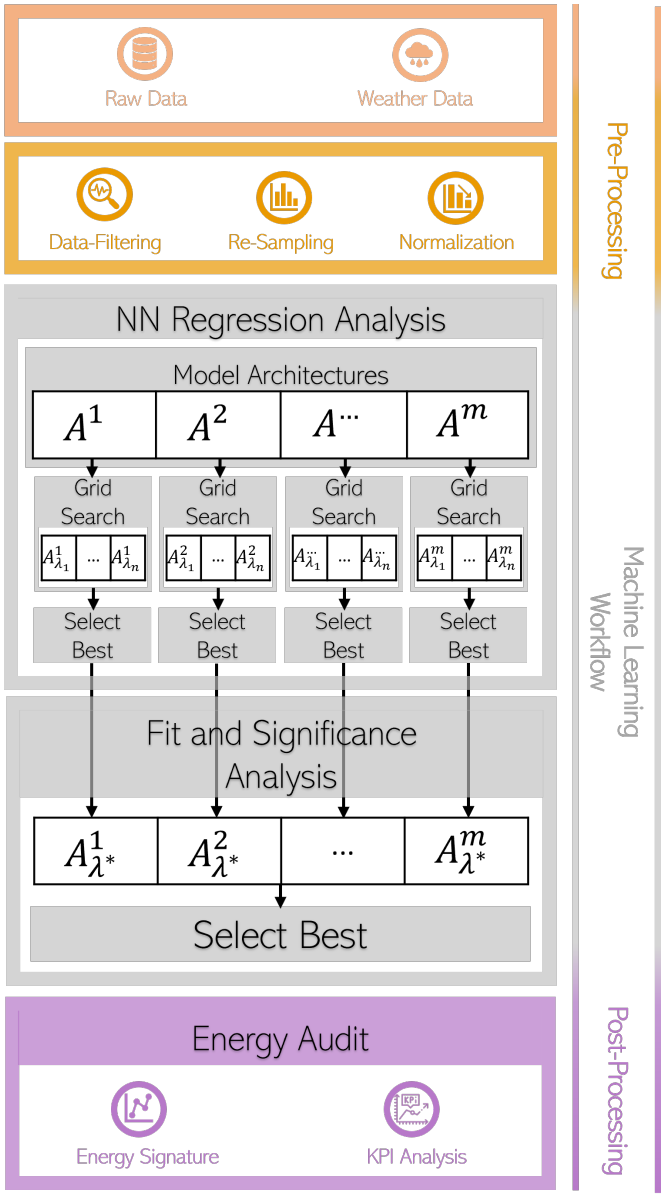

Figure 2: The methodological approach of the NN-based regression models for the non-intrusive energy audit of buildings, from the raw data to the results

of other buildings with a similar usage category. Furthermore, this step is necessary to scale the input data in order to produce an ES in the form expressed by Equation 17. For this reason, the re-sampled data shall be normalized by dividing them by the internal heat generation ϕ_{St}, as in the following equation:

$$\hat{P}_{TOT,norm} = \frac{\hat{P}_{TOT}}{\phi_{St}} \tag{18}$$

where $\hat{P_{TOT}}$ is the re-sampled daily load.

3.1.2. Machine Learning Workflow

In this work, NN are proposed to obtain the most accurate regression model for ES extraction. In this step, several NN architectures are trained to represent different degrees of freedom

of the ES regression model, with respect to the number of cooling regimes or particular discontinuities of the ES. Moreover, each NN architecture is trained with multiple combinations of its hyperparameters to ensure optimal training of each NN architecture. Several architectures are then compared and the best is selected, on the basis of accuracy and interpretability. This step is composed of the following two blocks: *i) NN Regression Analysis*, and *ii) Fit and Significance Analysis*.

NN Regression Analysis. The relationship between the outdoor weather variables (e.g., outdoor temperature) and the electrical load of a building is affected by the response of the HVAC system to the required thermal load. Different regression models were designed, considering the eventual existence of multiple cooling regimes over different ranges of outdoor weather variables, in order to assess the impact of these variables on the electrical load of a building. The *Model Architectures* block involves implementing several regression models, A^i, where i refers to the number of cooling regimes identified by the model. Five main regression models were considered: *i)* a basic regression model A^1, featuring one unconditioned region and one cooling regime; *ii)* three more complex models, that is, A^2, A^3, and A^4, which include one unconditioned region and two, three, and four cooling regimes, respectively; and *iii)* A^{1D}, an evolution of the A^1 regression model to handle step discontinuities. The fundamental designs of these regression models are detailed in Section 3.2.

Each regression model is trained and tested separately. Hence, the outdoor air temperature is considered as input for the various NN, while the aggregated load is provided as output and compared with real values. The dataset is split into the training and test sets, which correspond to two-thirds and one-third of the dataset, respectively. Furthermore, in order to guarantee the optimal training of each single model, a *Grid Search* procedure, in which different sets of values of the NN hyperparameters were tested, was carried out for each regression model. Hence, all the NN regression models, A^i, with $i = 1...m$ were trained with all the hyperparameters set λ_j with $j = 1...n$.

The Grid Search results obtained for each A^i model were then analyzed by the *Select Best* module, which selects the best set of hyperparameters, λ^*, and, consequently, the trained NN regression model, $A^i_{\lambda^*}$. This operation ensures that under or overfitting phenomena are avoided. It is worth noting that several hyperparameters were found to be irrelevant, because of the simplicity of the considered regression problem. Hence, the only hyperparameters considered in the Grid Search procedure were the *learning rate*, which was tested over ranges of values from 0.5 to 0.0001, and the *batch*, which was set equal to the training set length (i.e. batch-learning), to multiple time steps (i.e. mini-batch), or to one (i.e. online learning). The remaining hyperparameters, such as the selection of the Adam optimizer and a 500 epochs training procedure, were fixed. Finally, the Root Mean Squared Error (RMSE) was applied as a loss function.

Fit and significance analysis. The resulting NN architectures $A^i_{\lambda^*}$ with $\{i = 1, ..., m\}$ were compared according to the following criterias. Firstly, a fitting criterion was considered, that is, the accuracy of the regression model. This aspect was quantified by means of the overall coefficient of determination of the model with respect to the input data R^2. More complex models can in fact achieve a better performance, in terms of accuracy. However, in such a case, the interpretability of the model decreases and the significance of the detected cooling regimes may not justify the increased complexity of the regression. Furthermore, a higher fitting level may be caused by an overfitted regression model. In such a circumstance, the overfitting phenomenon derives from the distribution of the data, which may be relevant in statistics, but which is insignificant from an energy audit point of view. Hence, a significance score was defined as:

$$\alpha = \frac{1}{N} \sum_{r=1}^{N} R_r^2 \tag{19}$$

where R_r^2 stands for the coefficient of determination of a cooling regime region r, and N is the number of cooling regime regions. The best model, $A^*_{\lambda^*}$, was selected as a trade-off to ensure both high R^2 and α values.

3.1.3. Post-Processing

Finally, as shown in Figure 2, the resulting best model, $A^*_{\lambda^*}$, was exploited to provide an accurate Energy Audit of the investigated buildings. Firstly, the ES of the buildings was obtained to highlight the behaviors of the HVAC system and of the building. Some fundamental energy-related KPI were then extrapolated by the best NN regression model. Thereby, the Energy Audit consists of two blocks: *i)* the *Energy Signature*, and *ii)* the *KPI Analysis*.

Energy Signature. The reference ES was obtained by the NN regression model, $A^*_{\lambda^*}$, by employing both the training and the validation sets to obtain a higher accuracy. The thus obtained ES can feature different shapes, depending on the model resulting from the previous steps. Indeed, ES can feature one or more sloped lines, regarding different cooling regimes, and may feature a step, which represents a possible discontinuity in the consumption values measured in two different regions. These different ES categories are represented by the above-mentioned A^1, A^2, A^3, A^4, and A^{1D} models that are presented in Section 3.2.

KPI Analysis. Once the outcomes of the analysis had been obtained, the typical building parameters were calculated. This methodology allows the most important parameters that describe the thermal behavior of buildings and the efficiency of their cooling system to be estimated. These parameters can be directly extrapolated by using the trained NN weights and biases values. Section 2.1 details the KPIs that can be estimated and points out the physical issues underlying ES. The buildings KPI that can be extracted by means of the described methodology are: *i)* the β^*_{Temp}, *ii)* the COP, *iii)* the BP, *iv)* and the EPs.

3.2. NN Regression Design

In the last few decades, the research community has exploited iterative calculation methods for the determination of ES, by looking for the best fit considering a number of many possible linear regression models. An efficient, quick, customizable, and easy-to-use ML tool is proposed in this manuscript to accomplish the regression task presented in the previous Sections. This is a NN that was developed to identify the unconditioned region, multiple cooling regimes, and the eventual presence of a load step concerning the activation of the cooling system. A Feed-Forward NN Was selected, among the NN in the literature, to fulfill this task, considering the requirements of the regression model and for the sake of interpretability of the designed ML tool.

Feed-Forward NN are constituted by perceptrons. These are simple elements that are fed with an input vector, x, which is weighted by a trainable vector, w, and summed up with a bias b. An activation function, χ, makes use of the previous contributions and provides the perceptron output, a. This operation may be represented as a function, as in the following Equation:

$$a = f(x) = \chi(\langle w, x \rangle + b) \tag{20}$$

or in its matrix form:

$$\begin{bmatrix} w_1 & ... & w_V \end{bmatrix} \times \begin{bmatrix} x_1 \\ ... \\ x_V \end{bmatrix} + b \xrightarrow{\text{activation}} a \tag{21}$$

where V represents the dimension of the input vector.

Multiple perceptrons may be fed by the same inputs to form a layer. The layer will output a vector with a dimension equal to the number of perceptrons in the layer itself according to the following Equation:

$$\begin{bmatrix} w_{1,1} & ... & w_{1,V} \\ ... & ... & ... \\ w_{P,1} & ... & w_{P,V} \end{bmatrix} \times \begin{bmatrix} x_1 \\ ... \\ x_V \end{bmatrix} + \begin{bmatrix} b_1 \\ ... \\ b_P \end{bmatrix} \xrightarrow{\text{activation}} \begin{bmatrix} a_1 \\ ... \\ a_P \end{bmatrix} \tag{22}$$

where P represents the number of perceptrons in the layer.

Multiple layers may be designed to form a Multi-Layer Perceptron (MLP) network. In this case, the output of a layer is fed as input for the following one, and so on. The last layer of the network is referred to as the output layer, and its output is y. This network is fed with one or more external variables, which are, for the present application, the outdoor weather variables, and it provides an output, that is, the total electrical load. As this analysis focuses on the investigation of ES, the only variable considered in the input layer is the outdoor temperature. This implies assuming that the other weather variables are negligible. The validity of this assumption may be tested, for instance, by means of a correlation analysis, as it will be reported in Section 5.

Hence, this case study has investigated the univariate ES. Considering those buildings where the cooling load, P_{CLC}, is the only load contribution that depends on the outdoor temperature, it follows that a total load variation, determined by a temperature change, is a cooling load variation. Under this condition, the aggregated electrical load measurements of buildings may be employed for energy audit purposes. It is worth noting that employing the aggregated electrical load, instead of the cooling load, could provoke the presence of noise in the regression models, as mentioned in Section 2, with a slight impact on their accuracy.

According to the abovementioned features regarding the desired Feed-Forward NN regression model, an MLP was designed to exploit the outdoor temperature and aggregated load measurements. Firstly, a single node *input layer* was considered to account for a one-dimensional input vector x. This vector was derived by scaling the outdoor mean daily temperatures to ensure effective training of the models. Indeed, the input data were normalized accordingly to both the minimum and maximum values, and were scaled to the range [0,1]. This is a standard step for NN training.

The offset of the regression line, produced by the existence of temperature-independent load contributions, P_{ind} (see Equation 17), was then estimated by providing the network with a single perceptron layer featuring a linear activation. This element in fact estimates the offset of the regression line by adjusting its bias value during the training. Since, according to Equation 17, this offset exists for both unconditioned and conditioned regions, this perceptron represents the network *output layer*, as this allows the estimated offset value to be summed up with any contribution calculated by the previous layers of the network.

Furthermore, the model is capable of detecting multiple cooling regimes, characterized by different temperature ranges and line slopes. To this aim, a layer, consisting of a number of perceptrons with a Rectifier Linear Unit (ReLU) function, was designed. These elements provide an output as follows:

$$a \begin{cases} 0 & \text{if } w \cdot x + b <= 0 \\ w \cdot x + b & \text{if } w \cdot x + b > 0 \end{cases} \tag{23}$$

These perceptrons provide an output equal to zero for any input below a certain threshold. Connecting these layer elements to the input layer, they are fed with the outdoor temperature value. Hence, each ReLU activated perceptron outputs zero for any temperature below a certain EP. The lowest threshold of those detected by the ReLU activated perceptron represents the BP temperature. Indeed, any single perceptron of the layer outputs zero below that temperature, that is, such a temperature does not determine any thermal load. For these reasons, these perceptrons were included in the *hidden layer* of the network. Lastly, it may also be desirable to design a NN that is able to handle discontinuities. Indeed, the regression should be able to detect the eventual existence of an activation step for the cooling system, $\Delta P_{act,CLC}$. To this aim, a perceptron, featuring a sigmoid activation, was provided to the network. This function in fact outputs values close to zero for inputs below zero and close to one for inputs above zero, as it may be derived from Equation 24:

$$a = \frac{1}{1 + e^{w \cdot x + b}} \tag{24}$$

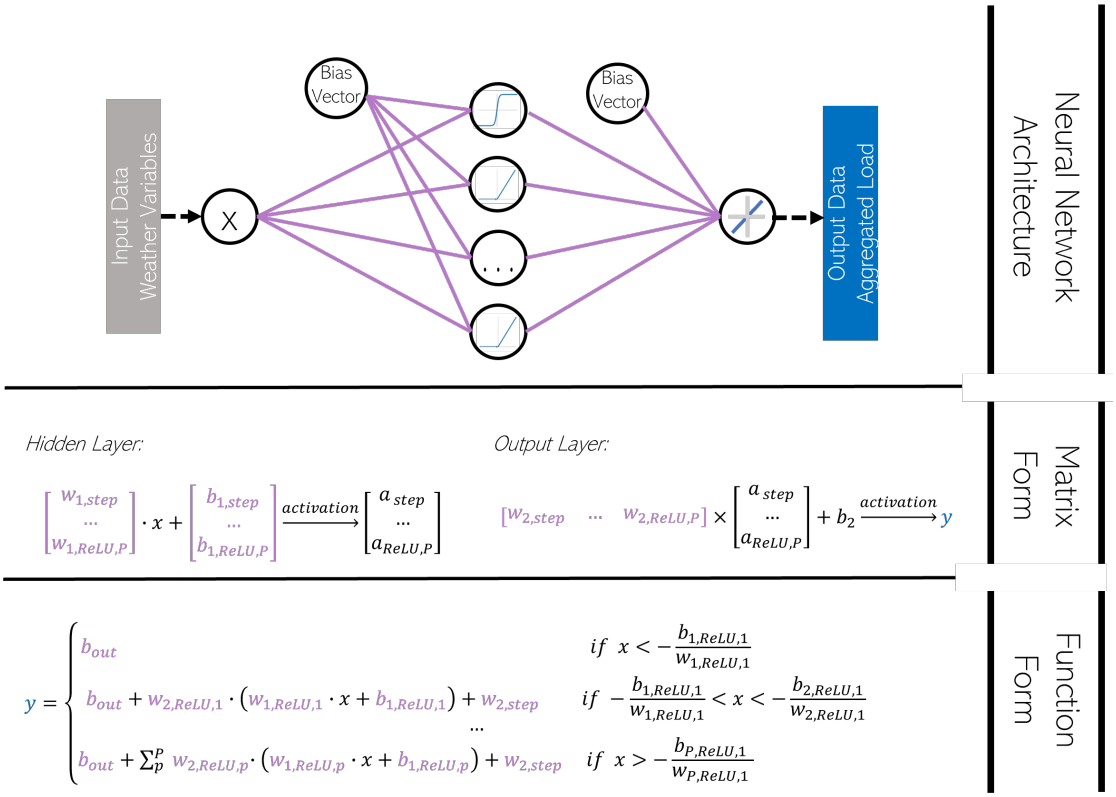

Figure 3: Analogy of the human brain-inspired representation, the matrix form and the function form of an NN-based regression model

Perceptron	Input from	Activation	Output to	Trainable
dense_1_Rectifier Linear Unit (ReLU)_1	Input Layer	ReLU	dense_2	Trainable
dense_1_ReLU_i	Input Layer	ReLU	dense_2	Trainable/freezed
dense_1_step	Input Layer	Sigmoid	dense_2	Trainable/freezed
dense_2	Layer 1	Linear	Output	Trainable

Table 1: Characteristics of the perceptron employed in the NN

A proper training would determine the weight and bias values of these perceptrons to shape a steep rise, as similar as possible to a step function. Namely, if $b \ll 0$ and $|w| \gg |b|$, the sigmoid function approximates to the following Equation:

$$a \begin{cases} 0 & \text{if } w \cdot x + b <= 0 \\ 1 & \text{if } w \cdot x + b > 0 \end{cases} \tag{25}$$

It should be noted that the approximation of the step function with a sigmoid is made necessary by the fact that it is highly recommended to not design an ad-hoc step activation function for NN. Indeed, step functions should not be included in NN, as they do not admit gradient calculation, which is a fundamental mechanism of NN training. This perceptron is included in the hidden layer along with the ReLU perceptrons. The inclusion of a sigmoid-activated perceptron represents an additional novel advance with respect to the literature regression models presented in the literature for ES purposes.

Finally, the MLP is ready when the designed layers are connected. Besides estimating an offset, by means of its bias, the output layer also weights the contribution of the previous elements of the network by means of its weights vector. Considering this and the outputs derived from the elements of the hidden layer, as reported in Equations 23 and 25 with regard to the ReLU-activated neurons and the sigmoid-activated one, respectively, the piecewise function of the network, described

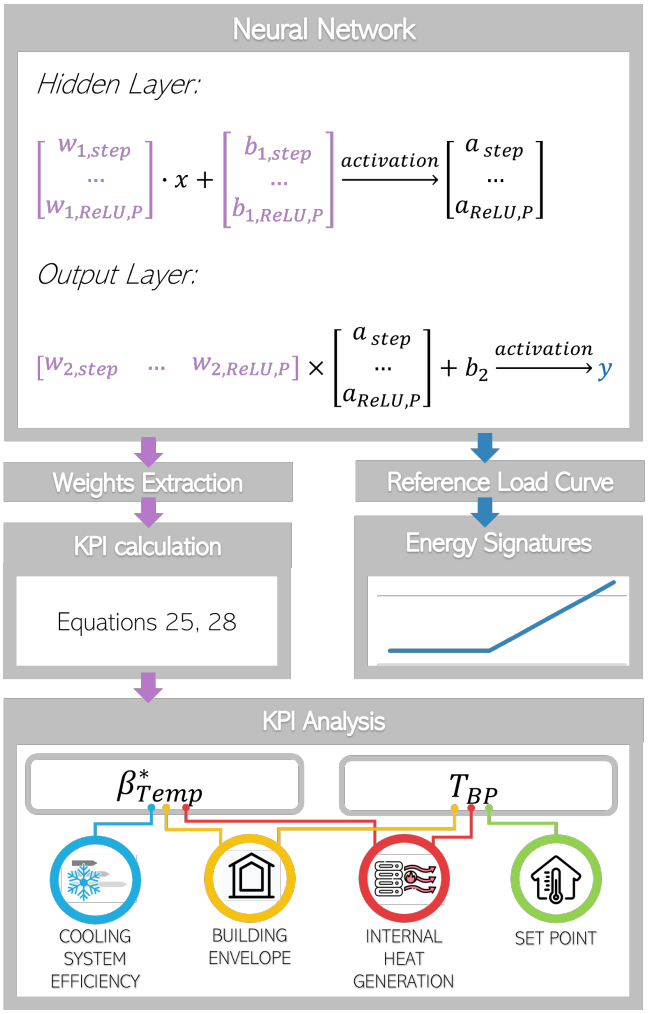

Figure 4: Outlook of the proposed Machine Learning-bases Energy Audit approach

by the following Equation, is obtained:

$$\mathrm{y(x)}\begin{cases} b_{out} & \text{if } x <= x_{BP} \\ b_{out} + \\ + w_{2,ReLU,1} \cdot (w_{1_{R}eLU,1} \cdot x + b_{1,ReLU,1}) + \\ + w_{2,step} & \text{if } \begin{cases} x > x_{BP} \\ x < x_{EP,1} \end{cases} \\ \quad \cdots \\ b_{out} + \\ + \sum_{p}^{P} w_{2,ReLU,p} \cdot (w_{1_{R}eLU,p} \cdot x + b_{1,ReLU,p}) + & \text{if } x > x_{EP,n} \\ + w_{2,step} \end{cases} \tag{26}$$

where $w_{2,ReLU,p}$ is the weight of the connection from perceptron p of the hidden layer to the output layer perceptron, and $w_{1,ReLU,p}$ is the connection from the input layer to the perceptron p of the hidden layer. It is easy to see that Equation 26 is equivalent to with Equation 17, that is the piecewise temperature-depending one. To conclude, Figure 3 highlights the correspondence between the Feed-Forward NN architecture and its matrix and function forms presented in the above description.

The different regression models presented in Section 3.1.2 can be tested by simply activating or deactivating different perceptrons from the hidden layer. The deactivation of branches of the network is achieved by initializing the perceptron bias and weights to output zero and setting them as non-trainable parameters. The possibilities for customizing the models of NN regression are summarized in Table 1. For instance, the basic regression model featuring one unconditioned region and one conditioning regime is achieved by training a single ReLU perceptron. Multiple ReLU perceptrons can be activated to detect multiple cooling regimes, while the sigmoid-activated perceptron can be employed to handle discontinuities. The configured NN architectures were trained separately for each building, subsequent to the Grid Search procedure and the Fit and Significance Analysis described in Section 2.

The energy audit can easily be performed by inspecting the network biases and weights. The outlook of this ML-based approach, is shown in Figure 4. The KPI analysis was carried out considering the weights and biases of the NN regression model. The EPs, BPs, β^*_{Temp} values, and the position and value of the activation step and load offset can be determined directly from the NN trained parameters. An EP is easily found by turning the inequalities that define the conditions in Equation 26 into their corresponding equations. Whenever multiple cooling regimes are considered, the EP should be sorted, and the lowest one represents the BP. Each cooling regime is characterized, not only by its temperature range, but also by the coefficient of the line in that region, that is β^*_{Temp}. These values can be obtained as:

$$\beta^*_{Temp,r} = \Sigma(w_{hidden,p} \cdot w_{out,p}) \qquad \forall i \mid T_{BP,p} < T_{max,r} \tag{27}$$

where $\beta_{Temp,r}$ is the coefficient of the line in the cooling region r; $w_{hidden,p}$ is the weight of the incoming connection of perceptron p; $w_{out,p}$ is the weight of the connection between perceptron p and the perceptron in the output layer, $T_{BP,p}$ is the BP of perceptron p, and $T_{max,r}$ is the maximum temperature of region r. It is worth noting that $\beta^*_{Temp,r}$ should be rescaled, accordingly to the scaling method employed for normalizing input and outputs of the network.

The temperature relative to the cooling regime activation step can be retrieved by considering the flex point of the sigmoid function, which can be found by setting its second derivative equal to zero, as follows:

$$\frac{d^2 \chi_{step}}{d(w \cdot x + b)^2} = -\frac{(e^{w \cdot x + b} - 1)\, e^{w \cdot x + b}}{(e^{w \cdot x + b} + 1)^3} = 0 \tag{28}$$

then:

$$e^{w \cdot x_{act,CLC} + b} = 1 \tag{29}$$

and, finally, the following is obtained:

$$x_{act,CLC} = -\frac{b_{step}}{w_{step}} \tag{30}$$

The value of $\Delta P_{act,CLC}$ can be derived from the weight of the connection between the sigmoid-activated perceptron and the output layer neuron, $w_{2,step}$. Finally, the value of the offset is derived from the bias of the output layer perceptron b_{out}.

4. Case Study

The methodology was applied to a real dataset, consisting of one year of aggregated load measurements with an hourly resolution, obtained from about eighty Data Centers (DCs) of an important Telecommunication (TLC) service provider in Italy. DC are industrial buildings that host the information technology equipment devoted to the management of TLC networks and to other services. The energy demand of TLC networks and their management buildings have been raising dramatically over the last years, featuring an annual increase of about 10 % over the last decade [24].

TLC buildings are characterized by an occasional and irrelevant presence of occupants, which has an insignificant impact on the final electrical demand. In fact, most of the consumption is due to the TLC equipment and by the huge electric demand for the cooling systems that are necessary to avoid overheating of the equipment itself. The contributions that account for these sub loads of the total building power consumption, P_{TOT}, are referred to as P_{TLC} and P_{CLC} respectively. The minor contribution to the load from the auxiliaries and from the lighting system are instead referred to as P_{Aux}, while P_{DISS} takes into account the energy conversion losses and the load due to the Uninterruptible Power Supply (UPS) units [25]. The latter contribution and the direct consumption of the TLC equipment, $P_{TLC}2$, are generally assumed to be constant in time, t. Hence, the energy balance of TLC buildings may be written as:

$$P_{TOT}(t) = P_{TLC} + P_{DISS} + P_{CLC}(t) + P_{Aux}(t) \tag{31}$$

The energy balance may be expressed more usefully as a function of the outdoor temperature, T_{ext}, for the purpose of the present research effort, as it was introduced in Section 2. In this case, the fundamental Equation that describes the behavior of the building is:

$$P_{TOT}(T) = P_{TLC} + P_{DISS} + P_{CLC}(T) + P_{Aux} \tag{32}$$

where P_{CLC} is the only contribution to the total consumption that depends on the outdoor temperature. The constant and temperature-independent load quotas, P_{TLC} and P_{DISS}, can be referred to as the base load:

$$P_{base} = P_{TLC} + P_{DISS} \tag{33}$$

The load contribution of the auxiliary systems, P_{Aux}, which is variable but independent of the outdoor temperature, may then be added to the aforementioned base load to form the independent consumption quota:

$$P_{ind} = P_{base} + P_{Aux} \tag{34}$$

The crucial formulation of ES introduced in Section 2 and expressed in Equation 17 can be introduced for this particular case study as:

$$\frac{P_{TOT}(T)}{\phi_{St}} = \begin{cases} \dfrac{P_{base} + P_{Aux}}{\phi_{St}} & \text{if } T_{ext} <= T_{BP} \\ \quad \dots & \\ \dfrac{P_{base} + P_{Aux}}{\phi_{St}} + & \\ + \beta^*_{Temp} \cdot (T_{ext} - T_{BP}) & \text{if } T_{ext} > T_{EP,r} \end{cases} \tag{35}$$

By considering the eventual existence of discontinuities due to an activation load step of the cooling system, $\Delta P_{act,CLC}$, the piecewise form expressed in Equation 35 becomes:

$$\frac{P_{TOT}(T)}{\phi_{St}} = \begin{cases} \dfrac{P_{base} + P_{Aux}}{\phi_{St}} & \text{if } T_{ext} <= T_{BP} \\ \quad \dots & \\ \dfrac{P_{base} + P_{Aux}}{\phi_{St}} + & \\ + \beta^*_{Temp} \cdot (T_{ext} - T_{BP}) + & \text{if } T_{ext} > T_{EP,r} \\ + \dfrac{\Delta P_{act,CLC}}{\phi_{St}} & \end{cases} \tag{36}$$

It is worth pointing out that the activation of the cooling system determines a step of the electrical load in real-case measurements, independently of the thermal load the system has to deal with. Hence, this quota is taken into account as a contribution to the total consumption in the cooling region, but it is considered constant.

It should be noted that all the simplifying assumptions introduced in Section 2 hold true for this particular case study. The T_{in} value can be considered constant and can be regarded as T_{SP}, as a result of the strict constraints regarding the control of the indoor air temperature of DC. It is important to point out that the P_{Aux} value can depict a certain statistical correlation with the outdoor temperature. Since the P_{Aux} load profile depicts a typical daily pattern [26], the Re-Sampling step described in Section 3.1.1 has removed this statistical correlation. After deriving the mean daily values, Equation 36 can be expressed as:

$$\frac{\hat{P}_{TOT}(T)}{\phi_{St}} = \begin{cases} \dfrac{\hat{P}_{base} + \hat{P}_{Aux}}{\hat{\phi}_{St}} & \text{if } T_{ext} <= T_{BP} \\ \quad \dots & \\ \dfrac{\hat{P}_{base} + \hat{P}_{Aux}}{\hat{\phi}_{St}} + & \\ + \beta^*_{Temp} \cdot (\hat{T}_{ext} - T_{BP}) + & \text{if } T_{ext} > T_{EP,r} \\ + \dfrac{\Delta \hat{P}_{act,CLC}}{\phi_{St}} & \end{cases} \tag{37}$$

where $\hat{P}$ refers to mean daily load, $\hat{T}$ to mean daily temperature and $\hat{\phi}_{St}$ to mean daily internal heat generation.

Next, on the basis of the TLC domain expertise, an additional simplification is introduced. Indeed, the internal heat generation in DC may be considered equal to the constant electrical load of TLC devices, that is:

$$\phi_{St} = P_{TLC}. \tag{38}$$

This assumption eases the calculation of the internal heat generation value, as P_{TLC} can be trivially estimated [26] considering P_{base}, which comprehends the former load quota and the contribution of power losses P_{DISS} [25]. Finally, both sides of piecewise Equation 37 can be normalized by P_{TLC} and expressed, as:

$$\hat{P}_{TOT,norm}(T) = \begin{cases} \hat{P}_{base,norm} + & \\ + \hat{P}_{Aux,norm} & \text{if } T_{ext} <= T_{BP} \\ \quad ... & \\ \hat{P}_{base,norm} + & \\ + \hat{P}_{Aux,norm} + & \\ + \beta^{*}_{Temp} \cdot (\hat{T}_{ext} - T_{BP}) + & \text{if } T_{ext} > T_{EP,r} \\ + \Delta\hat{P}_{act,CLC,norm} & \end{cases} \tag{39}$$

that is, the estimation of the final expression of the normalized total power consumption, $P_{TOT,norm}(T)$.

5. Experimental Results

The proposed methodology has been tested on a real-world dataset pertaining to the case study presented in Section 4. The purpose of this Section is to present and comment on the experimental results of the different steps of the methodological framework, that is, *i)* the Pre-Processing, *ii)* the Machine Learning Workflow, and *iii)* the Post-Processing.

5.1. Pre-Processing

The input data were first refined, in the Data-Filtering step, by replacing any outliers, that is, any out-of-range values whose existence could depend on measurement errors. By doing so, around 5% of the load profiles' values were substituted by means of linear interpolation. The mean daily values were then obtained by resampling both the load and the outdoor temperature time series. Finally, these values were normalized with respect to the internal heat generation, ϕ_{St}. This was done, in the investigated case study, by dividing the load profiles by the DCs' load associated with the TLC equipment, P_{TLC}. Figure 5 shows the annual normalized load profiles of the buildings present in the dataset. Most of the buildings' load profiles are characterized by a low consumption, which is close to the value of the base load for the colder months. A relevant increase in consumption can instead be observed for all the buildings between the end of spring and the beginning of summer.

	P_{TOT}	T	RH	v_{wind}	G
P_{TOT}	1	0.884	-0.409	-0.230	0.376
T	0.884	1	-0.530	-0.286	0.484
RH	-0.409	-0.530	1	0.151	-0.522
v_{wind}	-0.230	-0.286	0.151	1	-0.157
G	0.376	0.484	-0.522	-0.157	1

Table 2: Correlation Analysis of the aggregated electrical load, P_{TOT}, and the weather variables from a DC

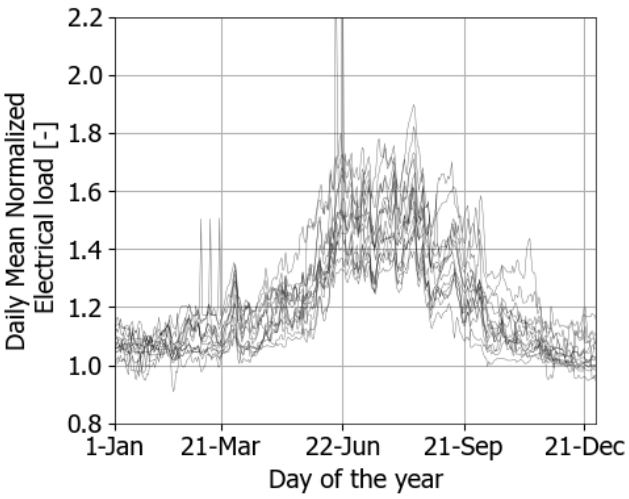

Figure 5: Normalized Daily Mean Electrical load profiles from a subset of buildings included in the analysed real-world dataset

In some cases, the electrical load can exceed twice the value it has during the cold months, due to the additional contribution of the cooling system to consumption. Figure 6 instead shows the load duration curves from a subset of 14 buildings located in Bologna, Italy. The typical behavior of the buildings is characterized by quasi-constant low consumption values for several days, while the load rapidly increases for a relatively low number of hot days. This behavior is particularly interesting when compared with the duration curve of the outdoor temperature of the city recorded for the year 2020. The mean daily temperature duration curve is quite linear, ranging from values below zero to up to 30 °C. At the same time, the load curves depict different slopes on the graph. This suggests that the impact of the outdoor temperature on the load of the building is not linear, but it only becomes relevant just in case a certain temperature threshold is overcome. Moreover, a correlation analysis of the weather variables was carried out, and it pointed out that temperature was the most relevant factor in determining the electrical load in a DC, as reported in Table 2. In this Table, the Pearson correlation factor R between the weather variables and the electrical load are calculated. This analysis clearly highlights the minor relevance of the other weather variables, namely radiation (G), relative humidity (RH), and wind speed (v_{wind}). Indeed, the correlation factor reported between the electrical load P_{TOT} and the solar radiation G depicts low values. For instance, the Pearson coefficient between G and P_{TOT} is 0.376, regarding the building whose resulting correlation factors are reported in Table 2. The mean value retrieved for the whole dataset was 0.411, with a few buildings featuring a correlation factor over 0.5. At the same time, the mean R retrieved between temperature and P_{TOT} was 0.873, with over 90% of the buildings featuring a value over 0.8. Hence, the electrical load is proved to have a slight dependence on the radiation, G, for this case study. Moreover, it should be recalled that the internal heat generation, ϕ_{St}, is constant. These two conditions ensure the fair adoption of the simplifying hypothesis introduced in Section 2, that is,

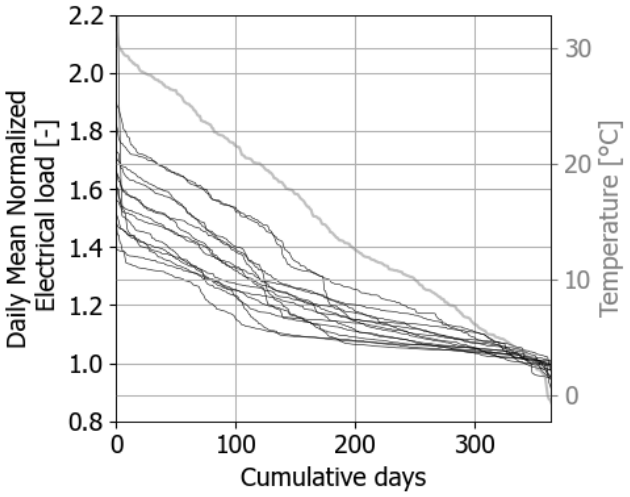

Figure 6: Load duration curves and temperature duration curve from the subset of buildings located in Bologna

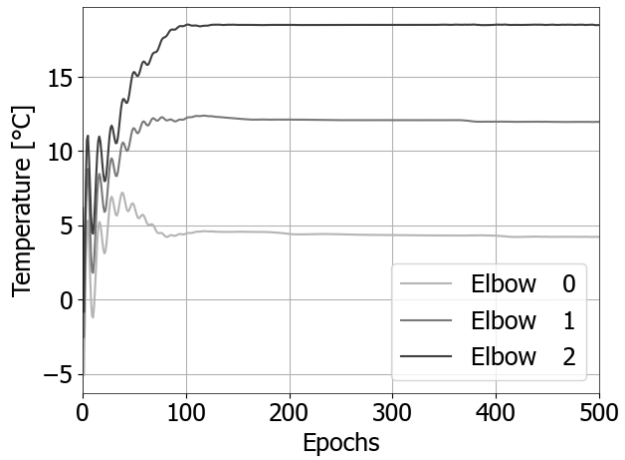

Figure 7: Evolution of the position of the elbow points over the NN training epochs

the BP temperature, T_{BP}, can be considered as a constant parameter of the investigated buildings.

5.2. Machine Learning Workflow

The pre-processed data were fed into the Model Architectures block that redirected it to the NN regression models. Specifically, five models were tested for each building, respectively: *i)* a single cooling regime model, A^1; *ii)* three multiple cooling regimes models with a number of EP from 2 to 4, namely A^2, A^3, and A^4; and, finally, *iii)* a regression model provided with a sigmoid-activated branch to handle discontinuities, considering the existence of a single cooling region, called A^{1D}.

The Grid Search procedure is carried out separately for each building and model to ensure the optimal training of each NN. In all cases, batch learning resulted to be the best solution. The optimal setting for the learning rate is not unique, and it depends on the building and on the employed regression model.

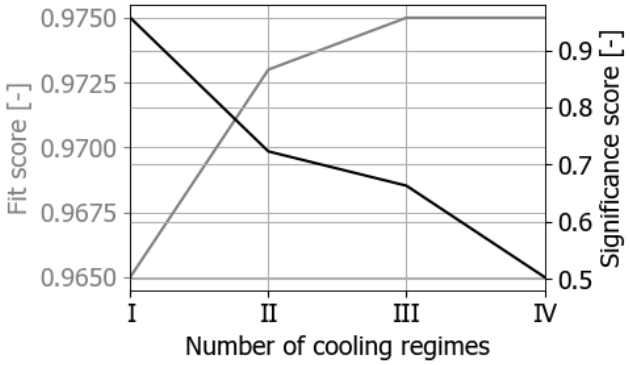

Figure 8: Fit and accuracy scores from different regression models pertaining to Building A

Indeed, higher learning rates, in the order of magnitude of 0.1, resulted in a higher accuracy for simple networks, such as the one employed for the single cooling regime model, A^1. On the opposite, the model requiring the lowest learning rates, in the order of magnitude of 0.001, was the model that was used to handle discontinuities, A^{1D}.

In all cases, at least one reliable model was provided for each building by the Grid Search procedure. This is witnessed by the analogous decreasing evolution of the training and test losses over the epochs, as well as by the achievement of stable KPI values as they were computed by inspecting the network. The extraction of the KPIs is dealt with in more details in Section 5.3. Nevertheless, in order to indicate the stable training that was achieved by the networks, a result from the calculus of the EP for model A^3 is anticipated in Figure 7.

Three EP values were computed for the DC referred to as Building A. As it can be observed by looking at the EP values computed during the first training epochs, all the perceptrons involved in the network are affected by relevant fluctuations in their parameters. Nevertheless, as the training procedure continues, the fluctuations appear to decrease rapidly and finally reach a stable plateau a few epochs after 100. No relevant change in the KPI estimation can be observed for the following 400 epochs, which would appear to indicate that the network had achieved the optimal parameters training to fit the problem.

Hence, an optimal trained network, $A^m_{\lambda^*}$, was obtained for each building and regression model. Next, the coefficient of determination, R^2 (i.e. fit score), and the significance score, α, were computed for each network. It should be pointed out that the fit and significance scores were designed as complementary metrics. The simultaneous analysis of these metrics ensures a reliable selection of the correct regression model. Indeed, on the one hand, the fit score, R^2, is expected to increase as the networks are provided with more perceptrons, that is, they can describe more cooling regimes. On the other hand, the significance score, α, can be expected to decrease as multiple regions are defined.

These expectations are confirmed by the experimental results from the best single and multiple cooling regimes models, as reported in Table 3 for a subset of buildings. The inversely

Model	A^1		A^2		A^3		A^4		A^{1D}	
	R^2	α	R^2	α	R^2	α	R^2	α	R^2	α
Building A	0.965	0.957	0.973	0.723	**0.975**	**0.663**	0.975	0.503	0.967	0.705
Building B	0.952	0.940	**0.971**	**0.877**	0.971	0.615	0.971	0.464	0.956	0.860
Building C	0.912	0.872	**0.924**	**0.697**	0.924	0.468	0.924	0.351	0.919	0.622
Building D	0.950	0.940	**0.957**	**0.770**	0.957	0.469	0.957	0.384	0.950	0.765
Building E	**0.936**	**0.888**	0.936	0.383	0.936	0.346	0.936	0.239	0.936	0.393
Building F	**0.898**	**0.889**	0.904	0.548	0.904	0.358	0.904	0.267	0.901	0.412
Building G	0.940	0.929	**0.960**	**0.762**	0.961	0.625	0.961	0.296	0.941	0.770

Table 3: Fit and significance scores from a subset of buildings taken from the real-world dataset, considering different regression models. The best models are highlighted for each building

proportional relationship between the fit score, R^2, and the significance score, α, with respect to the number of perceptrons is confirmed. Indeed, it can be observed that R^2 increases for any single investigated building, as multiple cooling regimes are added to the regression model. Similarly, α always decreases. Furthermore, R^2 rises, albeit only up to a certain value, after which any further increases in the model complexity do not determine any further improvements. At the same time, it may be observed that, as an increase in complexity does not determine any further boosting of the accuracy of the model, it determines a drop in the significance score.

This can easily be noted by looking at the scores of Building E in configurations A^1 and A^2, Building C in configurations A^2 and A^3, and Building G in configurations A^3 *and* A^4. This aspect is highlighted for Building A in Figure 8. In order to select the best configuration, the following criteria are applied to the Fit and Significance Analysis block: *i)* achieve the highest possible fit score, R^2, *ii)* guarantee a relevant significance score, α, over 0.6, and *iii)* obtain the highest possible significance score.

With regard to the A^{1D} models, it should be pointed out that this approach only resulted to be the best for one building. The model did not provide any meaningful outcomes for the other investigated sites because of the irrelevant magnitude of its output. For this reason, the discussion first focuses on the remaining regression models. The results of the particular case study, whose best regression model resulted to be A^{1D}, are presented and commented in Section 5.3.

For the sake of clarity, Figure 9 reports the resulting ES for the single cooling regime, A^1, and multiple cooling regime regression models, A^2, A^3, and A^4, for Buildings A and B. The background colors of the cooling regions represent the coefficients of determination that were obtained, for each of them, by the regression model. In both cases, the single cooling regime regression model, A^1, depicts high significance score values, but the ES obtained as an output of the NN does not fit the real values optimally. The accuracy of the output improves as the number of cooling regions increases, that is, the coefficient of determination of the model increases. At the same time, it should be observed that the single region coefficient of determination decreases, thereby causing a decline in the significance score.

Considering the criteria mentioned above to select the best model, the three cooling regimes model, A^3, was selected for Building A, while the two cooling regimes model, A^2, was selected for Building B. Indeed, no significant improvement regarding the fit score was observed when a fourth cooling regime was considered for Building A with respect to the three cooling regimes model, A^3. In the A^4 model, the narrow region detected (i.e. red zone) for the second regime shows an extremely low significance value. Hence, this configuration was rejected. Regarding Building B, the three cooling regimes model, A^3, does not enhance the relevant increase in the accuracy of the model as the first and second regions have very similar slopes. Therefore, these two regions are likely to be correctly interpreted as a single one. Lastly, it is interesting to note that the four-cooling regimes model, A^4, from Building B actually depicts three regions. This is due to the collapse of two EP to the same value. This phenomenon is observed for several buildings as multiple cooling regimes' models are tested.

Generally, the single cooling regime model, A^1, was found to be suitable for the majority of the investigated sites, and showed the best performances for almost 60 buildings. Around 20 buildings were described better by the two-cooling regimes model, A^2, while in just a few cases the three-cooling regions model, A^3, resulted to be the best. Finally, in just one case the model featuring a sigmoid-activated building was selected.

All the designed NN regression models achieved good performances for the whole dataset, as attested by the mean coefficient of determination, R^2, values reported from the ES models of the buildings. Indeed, the mean R^2 value obtained for these models was 0.849, while 50% of them achieved values of over 0.90, and 25% over 0.93. This confirms the consistency of the hypothesis reported in Section 3, whereby it is assumed that the electrical load of DC is basically only depends on the outdoor air temperature. The performance of the proposed methodology slightly exceeds that of the most widely adopted algorithm for ES in the literature, that is, when an iterative procedure is used to find the best fit of a two-segment linear regression [9], which achieved a mean R^2 equal to 0.822.

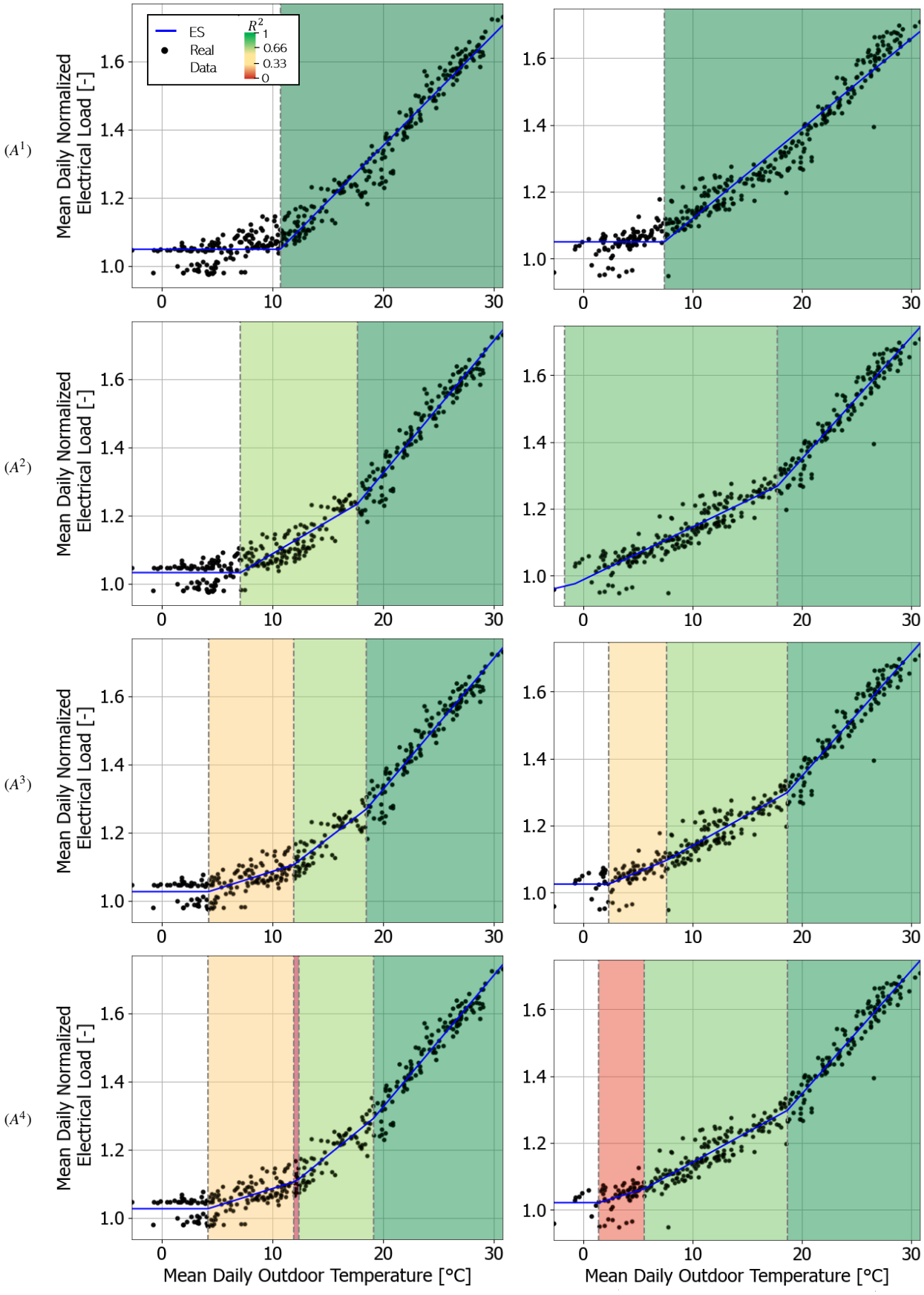

Figure 9: Visual comparative analysis of the four regression models A^1, A^2, A^3, and A^4, applied to Buildings A and B.

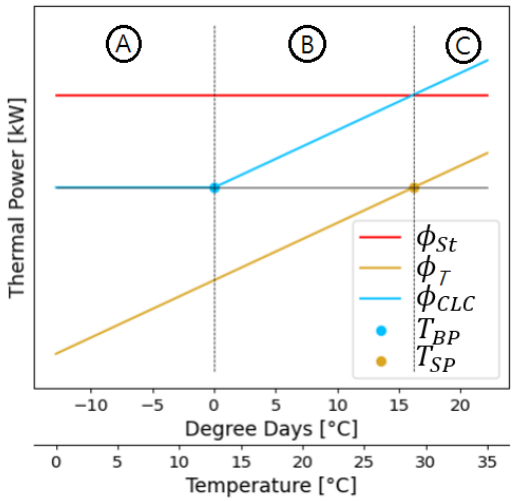

Figure 10: Estimated contributions of thermal balance of a DC, retrieved through the application of the proposed methodology to Building E

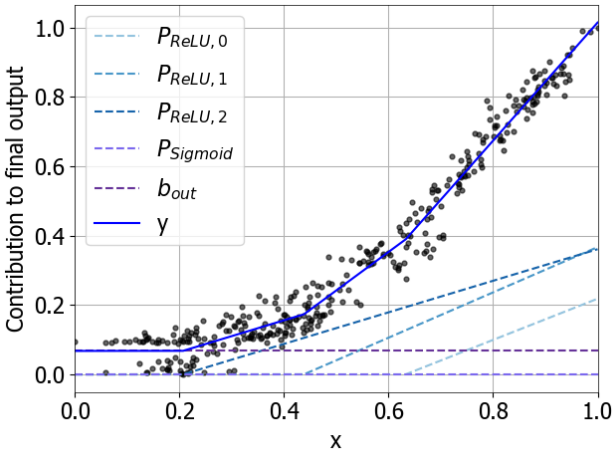

Figure 11: The contribution of each hidden layer perceptron to the calculation of the final output of the NN, regarding the application of the 3 cooling regimes regression model on Building A

5.3. Post-Processing

Finally, as one model had been selected for each building, the Energy Audit step was performed. The objective of this step is twofold: i) to calculate the Energy Signature and, ii) to conduct the KPI Analysis. In fact, the NN regression models, presented in Section 3, were designed to both provide the relationships among the mean daily load and the outdoor temperature (i.e. ES), and to allow the characteristic parameters of the building to be extracted.

A subset of the ES of eight buildings, obtained from the Energy Signature block, are shown in Figure 12, along with the real temperature and load values recorded over the year of the analysis. The following models were selected for this subset of buildings, by means of the Fit and Significance Analysis: *i)* three buildings with a single cooling regime regression model (i.e. Buildings *E*, *F*, and *H*); *ii)* four buildings with a double cooling regimes regression model (i.e. Buildings *B*, *C*, *D* and *G*); and, finally, one regression featuring a three-cooling regimes regression model (i.e. Building *A*).

It is worth noting that the slope of the ES increases moving from the lower temperature interval regime the higher temperatures ones when multiple EP exist in a regression model. Indeed, each of the multiple ReLU-activated perceptrons contributes to the final output as its activation threshold is overcome. The contribution of each NN branch to the final output is depicted in Figure 11 to highlight this behavior, for the case study of *Building A*. The NN regression model selected for this building was A^3. It should be noted that the axis corresponds to the values of the input and output of the NN, which are the input and output values of the regression, scaled to 0 to 1 range. Below the lower BP, which occurs, in this case, at around 0.2, the only contribution to the final output is provided by the bias of the output layer, namely b_{out}. After successive EP are overcome at around 0.43 and 0.63, additional contributions of the ReLU-activated perceptrons are summed and this leads to the progressive, although discretized, increment of the line slope. In this case, the contribution of the sigmoid-activated branch is zero for the whole input domain, as this branch is frozen in this model.

The methodology also allows the energy KPIs characterizing each building to be extracted via the KPI Analysis block. Generally, the investigated buildings reported a T_{BP} mean value equal to 13.8 °C. Less than 20 % of the buildings reported a T_{BP} higher than 17 °C. Finally, a few DCs are characterized by extremely low balance points, even below 10 °C. To this extent, it should be pointed out that DCs can feature extremely dense internal heat generation. This means that cooling may be needed even though the indoor air temperature is much higher than the outdoor one. The indoor air temperature is assumed to be equal to the set point, which, for the investigated DCs, was set equal to 29 °C by the energy managers to ensure the prolonged life of the equipment [27]. By recalling Equation 9, it can be noted that the BP is depending on T_{SP}, the internal heat generation, and on the thermal sensitivity of the building, as depicted in the following equation:

$$T_{BP} = T_{SP} - \frac{P_{TLC}}{k_{TOT}} \tag{40}$$

Since T_{SP} is given and is equal for all the DC, the values of T_{BP} may be interpreted by only considering P_{TLC} and k_{TOT}. The thermal behavior of DC can easily be understood by looking at Figure 10, where the experimental results retrieved for Building *B* are reported. This plot clearly shows three regions that are described hereafter. Region *A* is the unconditioned region, which is characterized by a thermal flux through the envelope toward the external environment that is able to balance the internal heat generation. Region *B* instead is where the cooling system intervenes, as the outdoor temperature overcomes T_{BP}, in order to keep the indoor air temperature below the set point, T_{SP}. Nevertheless, heat transmission through the envelope still plays a role in removing heat from the building. The wider Region *B* is, the higher the potential savings can be achieved

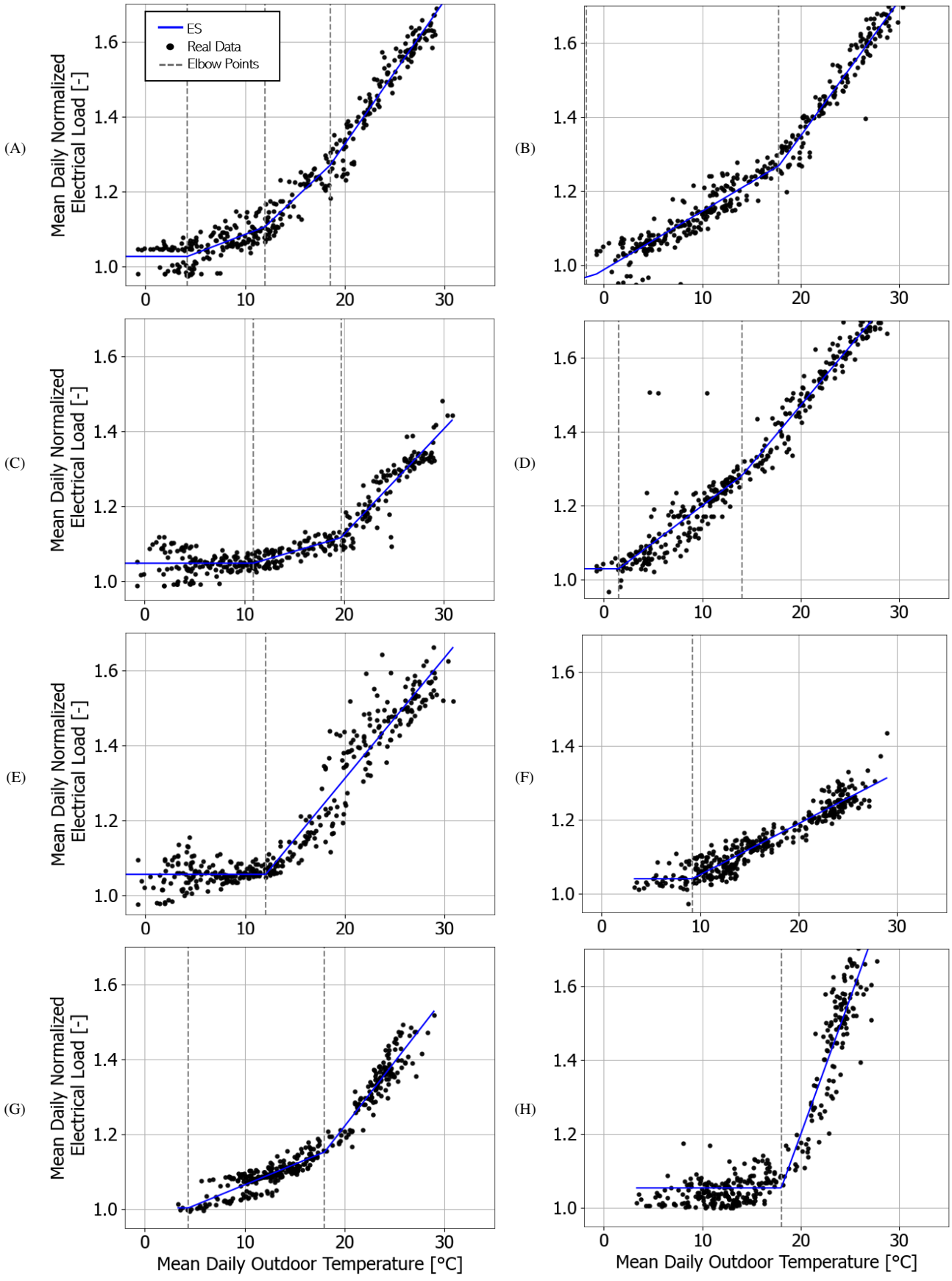

Figure 12: The best NN regression models retrieved for a subset of buildings from the real-world dataset

by means of a free cooling system [28]. Finally, Region C in Figure 10 corresponds to the building condition in which the thermal flux through the envelope, ϕ_T, reverses, since the outdoor temperature is higher than the indoor one, which is kept constant and equal to T_{SP}. In this case, the cooling system has to tackle both the above-mentioned contributions.

Considering the retrofit scenario in which the building envelope is modified to increase thermal flux through the walls (i.e. increasing k_{TOT}), the yellow line, ϕ_T, would feature a steeper tilt. Since T_{SP} would not be modified, this would cause a rise in T_{BP}. In this case, less energy would be wasted if the outdoor air temperature were below the T_{SP}. On the other hand, the electrical load would rapidly rise for hot days. However, the building envelope and its internal heat generation are not the only factors that can affect the electrical consumption caused by the conditioning of DC. Indeed, the thermal fluxes reported in Figure 10 do not take into account the efficiency of the cooling system itself.

As stated in Section 2.1, the proposed KPI β^*_{Temp} is intended to describe this particular building behavior. Considering the lower temperature cooling regime, the values retrieved by the methodology attest that a typical DC features a β^*_{Temp} equal to $0.027\,59\,^{\circ}\mathrm{C}^{-1}$. These values are more easily understood if they are expressed as a percentage of base load. So, the electrical demand for cooling increases by $2.76\,\%\,^{\circ}\mathrm{C}^{-1}$ for any temperature above T_{BP}. For all those cases in which multiple cooling regimes are considered, the β^*_{Temp} values recorded for higher temperature regions are higher than those recorded for lower temperatures. For instance, Building A features β^*_{Temp} values equal to $0.0101\,^{\circ}\mathrm{C}^{-1}$, $0.0232\,^{\circ}\mathrm{C}^{-1}$, and $0.0383\,^{\circ}\mathrm{C}^{-1}$, respectively, for the first, second, and third cooling regime regions. This means the load increases in the first cooling region by 1.01 %, with respect to the base load, as the outdoor temperature rises by one °C. Furthermore, the COP of the cooling system of any region can be estimated by considering Equation 14. The COP values are 3.9, 1.7, and 1.1, with regard to the three cooling regions, respectively. These values appear reasonable as, if more than one chiller is available, the energy management system would initially activate the most efficient one, while the second one would only be called upon to intervene whenever the former was not capable of dealing with the whole thermal load, and so on.

Similarly, Building C features two β^*_{Temp} values equal to $0.0077\,^{\circ}\mathrm{C}^{-1}$ and $0.0280\,^{\circ}\mathrm{C}^{-1}$. These values correspond to COP values of around 7.1 and 2.0. The former may seem improbable but considering the adoption of free cooling systems [29] in many DC and taking into account the range of temperatures where this value holds valid, namely from around 11 °C up to about 20 °C [30], it represents clear evidence of the adoption of such a system within the investigated building.

Considering the extracted KPI, the energy audit procedure may be used to enhance comparative analyses with the aim of detecting possible inefficient buildings and the relative causes of their inefficiencies. With regard to the buildings reported in Figure 12, Buildings C and F appear to be the most virtuous ones, as only slight load increases are recorded as the outdoor

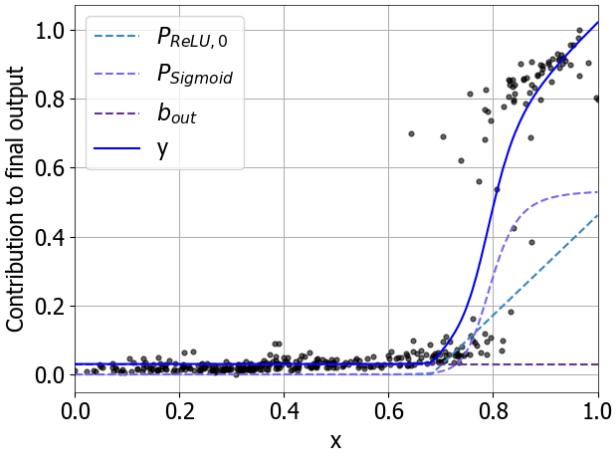

Figure 13: The contribution of each hidden layer perceptron to the calculation of the final output of the NN, regarding the application of the NN including a sigmoid-activated perceptron of a building taken from the dataset

temperature rises. Building H instead is characterized by an extremely high load demand of the cooling system for high outdoor temperatures, as its consumption values may exceed the base load by 70 % during hot days. In fact, the KPI Analysis shows that the estimated COP for this building is 1.2. Similar inefficiencies may be described for Buildings A and E. In the former case, this issue is relevant for the second and third cooling regimes regions, while the building appears to perform more efficiently for low temperatures. In other cases, the main factor that determines the low efficiency of the building is the low BP, which which results in a frequent activation of the cooling system, even in mild weather months. This is the case of Buildings B and D. Building G is also affected by a frequent activation of the cooling system. In this case, a higher overall efficiency is guaranteed by the good performance of the system itself, as a COP equal to 3.7 was estimated by the procedure.

As mentioned above, the regression model featuring a sigmoid-activated perceptron to describe the cooling system activation step only resulted to be significant for one building in the dataset. Indeed, setting this configuration for the rest of the buildings does not enhance the regression accuracy. In fact, the flex of the sigmoid is often pushed outside the input domain and the weight from the sigmoid-activated perceptron to the output layer, $w_{2,step}$, depicted extremely low values. This configuration resulted as stable and significant for the case study whose regression model is shown in Figure 13, along with the single contributions of the perceptrons of the employed network. In this case, the activation step of the cooling system is clearly described by the contribution of the sigmoid-activated perceptron. Furthermore, one ReLU-activated perceptron becomes active as the outdoor temperature increases. Considering the theoretical background expressed in Section 3.2, it is expected that the BP temperature and the step regarding the activation of the cooling system could overlap. This expectation was verified by considering the evolution of the estimated position of the elbow and flex from the regression model reported in Figure 14. After about 200 epochs, the positions of these points converge to a

value of about 21.5 °C. Thereafter, their position acquires stability, and the models converge to their optimal training. This building is characterized by medium efficiency. Indeed, the performance of the cooling system is affected by the high impact of $\Delta P_{act,CLC}$, which consists of about the same load consumption as the base load. On the other hand, low consumption values are recorded for many months due to the high value of T_{BP}, which determines an infrequent activation of the cooling system.

6. Conclusion and future work

A simple and effective ML tool, based on Feed-Forward NN, is proposed to calculate ES and the energy KPIs by only considering the aggregated electrical load of building premises and the outdoor temperature. This methodology can be used to support the widespread application of energy audits and enhances the awareness of the energy behavior and potential inefficiencies of buildings without sensorization. Several NN regression models have been designed to describe the different possible thermal behaviors of a building. The NN regression models were set up to take into account the eventual existence of multiple cooling regimes and to handle possible discontinuities due to the activation of the cooling system. Appropriate fit and significance scores were introduced to support the selection of the best NN regression model. Moreover, the methodological framework enhances the extraction of fundamental information from the bias and connection weights of NN. Indeed, it was used to calculate the following KPIs: *i)* the BP temperature; *ii)* the EP temperature, whenever a multiple cooling region model was selected; *iii)* the novel KPI β^*_{Temp} that accounts for multiple aspects of the thermal behavior of buildings; and *iv)* the cooling system COP, calculated for all the identified cooling regions.

The methodology has been tested on a real case dataset from an energy-intensive industrial sector, namely TLC. It was applied considering around 80 DC, and was able to produce accurate results, achieving a mean R^2 of 0.849, and to accurately detect the significant cooling regimes. Besides, the mean BP temperature retrieved through the methodology was 13.8 °C. Above this threshold, the electrical demand of the buildings rose considerably, as the mean estimated β^*_{Temp} was 2.76%. In those cases in which multiple cooling regimes existed, the COP values showed higher values regarding the lower temperatures conditioning regions. First, this analysis has confirmed the fundamental role of the cooling load in the total electrical demand of the investigated buildings. In particular, as we tested the methodology by considering industrial buildings featuring relevant internal heat generation, it was observed that the cooling system intervenes even with mild outdoor air temperatures. Consequently, the opportunity to adopt free cooling systems may represent a valuable energy efficiency measure, as it can substantially reduce the number of hours chillers are used. Even so, chillers will be needed in cases where free cooling is not capable of handling the whole thermal load, as well as when the outdoor temperature increases over a certain threshold, thus hindering the possibility of employing the former system. In this regard, this analysis may support the identification of possible retrofit actions regarding the existing cooling systems, as

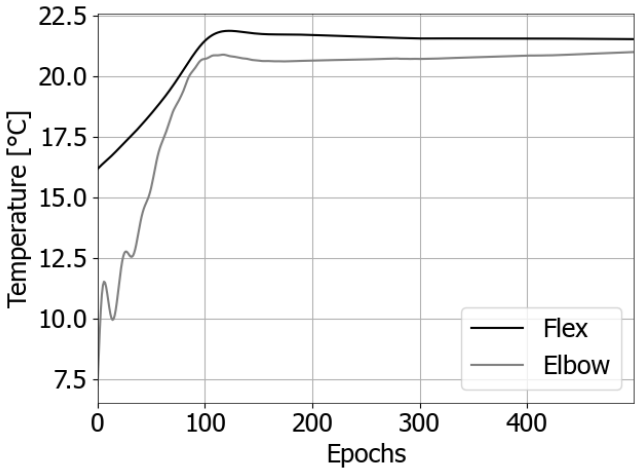

Figure 14: Evolution of the flex and elbow point position over NN training epochs

it allows the calculation of two efficiency indexes, namely COP and β^*_{Temp}, regarding multiple cooling regimes of any building whose total daily electrical demand is measured. In addition, the proposed methodology could also be adopted considering different types of energy carriers, and different types of ES, for instance, considering heating ES, and applied to different case studies.

Future works will take into account the validation of the KPI estimated value, derived by means of NN. Moreover, an automated selection of the possible retrofit actions could provide a fundamental extension of the outcomes of this investigation. Finally, the regression models employed could be further improved by including additional input weather variables. Such an approach, which we refer to as multivariate ES, should consider in particular those variables that can affect the thermal behavior of buildings, for instance, solar radiation. Since the latter variable can be of particular importance for other case studies, for instance from the residential and commercial sector, its analysis and inclusion in the models are of special interest with the aim of widespread of the proposed approach, both regarding cooling and heating ES.

Credit Authorship Contribution Statement

S. Eiraudo: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Visualization **D. S. Schiera:** Conceptualization, Software, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing, Supervision **L. Mascali:** Software, Investigation, Data Curation **L. Barbierato:** Conceptualization, Writing - Review & Editing, Supervision **R. Giannantonio:** Resources, Supervision, Project administration **E. Patti:** Writing - Review & Editing **L. Bottaccioli:** Writing - Review & Editing, Supervision **A. Lanzini:** Conceptualization, Formal Analysis, Writing - Review & Editing, Supervision, Project administration

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that has been used is confidential.

Acknowledgments

Simone Eiraudo acknowledges support from TIM S.p.A. through the Ph.D. scholarship.

References

[1] Energy efficiency directive, (Accessed: May. 24 2021). URL https://ec.europa.eu/energy/topics/energy-efficiency/targets-directive-and-rules/energy-efficiency-directive_en

[2] IEA (2018), The Future of Cooling, IEA, Paris https://www.iea.org/reports/the-future-of-cooling.

[3] I. Gaetani, P.-J. Hoes, J. L. Hensen, Estimating the influence of occupant behavior on building heating and cooling energy in one simulation run, Applied Energy 223 (2018) 159–171. doi:10.1016/j.apenergy.2018.03.108.

[4] P. Nageler, A. Koch, F. Mauthner, I. Leusbrock, T. Mach, C. Hochenauer, R. Heimrath, Comparison of dynamic urban building energy models (ubem): Sigmoid energy signature and physical modelling approach, Energy and buildings 179 (2018) 333–343. doi:10.1016/j.enbuild.2018.09.034.

[5] A. P. Vargas, L. Hamui, Thermal energy performance simulation of a residential building retrofitted with passive design strategies: a case study in mexico, Sustainability 13 (14) (2021) 8064. doi:10.3390/su13148064.

[6] J. K. Hwang, G. Y. Yun, S. Lee, H. Seo, M. Santamouris, Using deep learning approaches with variable selection process to predict the energy performance of a heating and cooling system, Renewable Energy 149 (2020) 1227–1245. doi:10.1016/j.renene.2019.10.113.

[7] A. A. A. Gassar, S. H. Cha, Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review, Energy and Buildings 224 (2020) 110238. doi:10.1016/j.enbuild.2020.110238.

[8] S. R. Asaee, V. I. Ugursal, I. Beausoleil-Morrison, Development and analysis of strategies to facilitate the conversion of canadian houses into net zero energy buildings, Energy Policy 126 (2019) 118–130. doi:10.1016/j.enpol.2018.10.055.

[9] F. Jacobsen, Energy signature and energy monitoring in building energy management systems, in: Proceeding of CLIMA 2000 world congress, Vol. 3, Copenhagen, Denmark, August 1985, pp. 25–31.

[10] P. Westermann, C. Deb, A. Schlueter, R. Evins, Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data, Applied Energy 264 (2020) 114715. doi:10.1016/j.apenergy.2020.114715.

[11] L. Tronchin, M. Manfren, B. Nastasi, Energy analytics for supporting built environment decarbonisation, Energy Procedia 157 (2019) 1486–1493. doi:10.1016/j.egypro.2018.11.313.

[12] A. Acquaviva, D. Apiletti, A. Attanasio, E. Baralis, L. Bottaccioli, F. B. Castagnetti, T. Cerquitelli, S. Chiusano, E. Macii, D. Martellacci, et al., Energy signature analysis: Knowledge at your fingertips, in: 2015 IEEE International Congress on Big Data, IEEE, New York City, NY, USA, 2015, pp. 543–550. doi:https://doi.org/10.1109/BigDataCongress.2015.85.

[13] R. Hitchin, I. Knight, Daily energy consumption signatures and control charts for air-conditioned buildings, Energy and Buildings 112 (2016) 101–109. doi:10.1016/j.enbuild.2015.11.059.

[14] O. Pasichnyi, F. Levihn, H. Shahrokni, J. Wallin, O. Kordas, Data-driven strategic planning of building energy retrofitting: The case of stockholm, Journal of Cleaner Production 233 (2019) 546–560. doi:10.1016/j.enbuild.2015.11.059.

[15] A. Anjomshoaa, M. Salmanzadeh, Estimation of the changeover times and degree-days balance point temperatures of a city using energy signatures, Sustainable Cities and Society 35 (2017) 538–543. doi:10.1016/j.scs.2017.08.028.

[16] A. Belleri, M. Avantaggiato, T. Psomas, P. Heiselberg, Evaluation tool of climate potential for ventilative cooling, International Journal of Ventilation 17 (3) (2018) 196–208. doi:10.1080/14733315.2017.1388627.

[17] S. Park, J. Shim, D. Song, Issues in calculation of balance-point temperatures for heating degree-days for the development of building-energy policy, Renewable and Sustainable Energy Reviews 135 (July 2020) (2021) 110211. doi:10.1016/j.rser.2020.110211.

[18] H. Sha, P. Xu, C. Hu, Z. Li, Y. Chen, Z. Chen, A simplified hvac energy prediction method based on degree-day, Sustainable Cities and Society 51 (2019) 101698. doi:10.1016/j.scs.2019.101698.

[19] G. Krese, Ž. Lampret, V. Butala, M. Prek, Determination of a building's balance point temperature as an energy characteristic, Energy 165 (2018) 1034–1049. doi:10.1016/j.energy.2018.10.025.

[20] S. Eiraudo, L. Barbierato, R. Giannantonio, E. Patti, L. Bottaccioli, A. Lanzini, A neural network-based methodology for non-intrusive energy audit of telecom sites, in: 2022 International Conference on Smart Energy Systems and Technologies (SEST), IEEE, Eindhoven, Netherlands, September 2022, pp. 1–6. doi:10.1109/SEST53650.2022.9898459.

[21] A. Petraglia, A. Spagnuolo, C. Vetromile, A. D'Onofrio, C. Lubritto, Heat flows and energetic behavior of a telecommunication radio base station, Energy 89 (2015) 75–83. doi:10.1016/j.energy.2015.07.044.

[22] F. Tounquet, C. Alaton, Benchmarking Smart Metering Deployment in EU-28, no. December, 2020, (Accessed: May. 30 2021). URL https://op.europa.eu/en/publication-detail/-/publication/b397ef73-698f-11ea-b735-01aa75ed71a1/language-en

[23] E. Bartz, T. Bartz-Beielstein, M. Zaefferer, O. Mersmann, Hyperparameter Tuning for Machine and Deep Learning with R: A Practical Guide, Springer Nature, 2023. doi:10.1007/978-981-19-5170-1.

[24] S. Lambert, W. Van Heddeghem, W. Vereecken, B. Lannoo, D. Colle, M. Pickavet, Worldwide electricity consumption of communication networks, Optics Express 20 (26) (2012) B513–B524. doi:10.1364/OE.20.00B513.

[25] M. Sorrentino, M. Bruno, A. Trifirò, G. Rizzo, An innovative energy efficiency metric for data analytics and diagnostics in telecommunication applications, Applied Energy 242 (2019) 1539–1548. doi:10.1016/j.apenergy.2019.03.173.

[26] S. Eiraudo, L. Barbierato, R. Giannantonio, E. Patti, A. Lanzini, L. Bottaccioli, Non-intrusive load disaggregation of industrial cooling demand with lstm neural network, in: 2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), IEEE, Prague, Czech Republic, July 2022, pp. 1–6. doi:10.1109/EEEIC/ICPSEurope54979.2022.9854581.

[27] D. Demetriou, Effectively applying the expanded ashrae guidelines in your data center, IBM Systems Report (Accessed: June. 23 2023) (2015). URL https://www.ibm.com/downloads/cas/1Q94RPGE

[28] H. M. Daraghmeh, C. C. Wang, A review of current status of free cooling in datacenters, Applied Thermal Engineering 114 (2017) 1224–1239. doi:10.1016/j.applthermaleng.2016.10.093.

[29] H. Zhang, S. Shao, H. Xu, H. Zou, C. Tian, Free cooling of data centers: A review, Renewable and Sustainable Energy Reviews 35 (2014) 171–182. doi:10.1016/j.rser.2014.04.017.

[30] A. F. Santos, P. D. Gaspar, H. J. de Souza, Ecoenergetic comparison of hvac systems in data centers, Climate 9 (3) (2021) 42. doi:10.3390/cli9030042.