NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY POLITEHNICA BUCHAREST
Faculty of Power Engineering
Department of Hydraulics, Hydraulic Machinery and Environmental Engineering
Master's programme:
Hydro-Informatics and Water Engineering – MS13

Class of
July 2026

Analysis of the Influence of Solar Shading on the Energy Consumption of Buildings Using Artificial Neural Networks

Master's Student
Eng. Aurelian-Andrei Panait

Scientific Coordinator
Assoc. Prof. Dr. Eng. Liana Ioana VUŢĂ

Abstract

The prediction of building energy consumption has traditionally relied on detailed thermal simulations, which are time-consuming and require a significant modelling effort. The present work evaluates the ability of artificial neural networks to predict energy performance while explicitly accounting for the solar shading effects generated by neighbouring structures. A baseline dataset is generated for a reference building without shading, and additional datasets introduce multiple shading configurations. Two predictive models are trained and evaluated using the RMSE, CV(RMSE) and R^2 indicators, and their results are compared against dynamic thermal simulations. Finally, a renovation scenario is analysed in order to quantify whether the AI-based models maintain their accuracy when the building characteristics are modified. The results highlight the influence of shading on prediction accuracy and the ability of machine learning models to integrate the physical context. The detailed analysis quantifies how the impact of shading depends on the geometry of the obstruction and on its interaction with the insulation level of the building envelope. Finally, an economic analysis translates these results into terms of return on investment in insulation, for the conditions specific to Romania.

Keywords: solar shading, energy consumption, artificial neural networks, dynamic thermal simulation, Pléiades, energy renovation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN Artificial Neural Network

BEM Building Energy Modeling

BESTEST Building Energy Simulation Test (standardised comparative testing procedure for thermal simulation programs)

COMFIE The dynamic thermal calculation engine of the Pléiades suite

CV(RMSE) Coefficient of Variation of the RMSE

DTS Dynamic Thermal Simulation

IEA International Energy Agency

LSTM Long Short-Term Memory (network)

MLP Multilayer Perceptron

MSE Mean Squared Error

NPV Net Present Value

PINN Physics-Informed Neural Network

PMV/PPD Predicted Mean Vote / Predicted Percentage of Dissatisfied (thermal comfort indicators)

RMSE Root Mean Squared Error

RT2012 / RE2020 French thermal regulations for buildings

SELU Scaled Exponential Linear Unit

SHAP SHapley Additive exPlanations (feature contribution interpretation method)

SVM Support Vector Machine

UBEM Urban Building Energy Modeling

WWR Window-to-Wall Ratio

CHAPTER 1
INTRODUCTION

The building sector accounts for a significant share of global energy consumption. Improving its energy efficiency is essential for reducing the environmental impact and for meeting climate targets. Dynamic thermal simulation tools remain the main reference for assessing energy performance, but they require detailed physical models and long computation times. By contrast, machine learning approaches can provide much faster predictions once trained, provided that the dataset used captures the relevant physical behaviour of the system. Solar shading is one of the parameters that significantly influence the energy demand of buildings, especially in dense urban environments. It modifies solar gains, daylight availability and, implicitly, the heating and cooling loads. Despite its importance, shading is often insufficiently represented in predictive models. The present work focuses on analysing how the explicit inclusion of shading parameters influences the performance of neural network models trained to predict the energy consumption of buildings.

CHAPTER 2
STATE OF THE ART

The prediction of building energy performance lies at the intersection of several rapidly developing research fields: detailed thermal modelling, its extension to the urban scale, machine learning methods and, specifically for the present work, the treatment of solar shading effects. This chapter synthesises the current state of knowledge in these fields, with the aim of positioning the contribution of the present work in relation to the existing literature and of highlighting the gap that it addresses.

2.1 Building energy modelling

Building Energy Modeling (BEM) is the classical approach for estimating energy consumption, based on solving the physical equations that govern heat transfer through the envelope, solar gains, ventilation and internal thermal exchanges. Established tools such as EnergyPlus, TRNSYS or Pléiades/COMFIE implement detailed dynamic thermal models, capable of simulating the hourly behaviour of a building over an entire year.

These physics-based methods provide high accuracy and full interpretability, since every input parameter has a direct physical meaning. However, they exhibit two limitations widely acknowledged in the literature. First, setting up a detailed model is a laborious, time-consuming and costly process, especially for existing buildings, where information about the components is often incomplete. Second, the long computation times make these tools less suitable for the iterative exploration of a large number of configurations, as required in early design stages or in optimisation studies [1].

An additional difficulty, extensively documented, is the gap between the predicted and the actually measured energy consumption (the *performance gap*). This gap can be significant, sometimes exceeding double or triple the predicted values, due to the uncertainties associated with workmanship quality, occupant behaviour and the variability of weather conditions [2].

2.2 Urban-scale energy modelling

Extending energy modelling from the individual building to urban ensembles has led to the development of Urban Building Energy Modeling (UBEM). UBEM approaches allow the estimation of energy demand at the neighbourhood or city scale, being essential for urban planning, the sizing of energy networks and the assessment of the renovation potential of the building stock [3, 4].

The literature distinguishes two main categories of UBEM approaches. Physics-based *bottom-up* approaches aggregate individual thermal models of buildings in order to obtain the energy demand at the urban scale, preserving the physical foundation but at a high computational cost. Data-driven approaches instead use statistical and measured consumption data to estimate energy demand, with a low computational cost but with more limited interpretability [3].

A recognised challenge at the urban scale is that, unlike the modelling of a single building, collecting accurate data about all the buildings in a district is practically impossible, which leads to increased uncertainties. In this context, choosing the appropriate level of model complexity becomes a trade-off between accuracy and the feasibility of parametrisation [4]. An important aspect in this respect is that the level of detail with which the shading from the surrounding environment is represented constitutes a modelling choice with a significant impact on accuracy. Dedicated studies have shown that the way the shading environment is treated, together with the level of geometric detail and the thermal zoning, directly affects the performance of UBEM models at the district scale [5].

2.3 Machine learning methods in energy prediction

In order to overcome the computational limitations of detailed simulation, machine learning (ML) approaches have experienced exponential growth over the last decade in the field of building energy consumption prediction [6]. Once trained, these models provide almost instantaneous predictions, which makes them particularly attractive for iterative applications.

2.3.1 Types of models used

The literature documents a wide range of algorithms applied to this problem. Review studies identify multiple linear regression, support vector machines (SVM), decision trees and Random Forests, gradient boosting methods, as well as artificial neural networks as the most widely used methods [7, 2]. For problems with a temporal component (short-term prediction of hourly consumption), recurrent neural networks, in particular LSTM architectures, are frequently used [6].

Multilayer Perceptron (MLP) neural networks remain a robust and widely adopted choice for regression problems in which the input variables are static descriptors of the building (envelope properties, geometry, systems), as is the case in the present work. Comparative studies show that, although more complex methods may offer marginal improvements, the choice of the optimal architecture depends strongly on the nature and size of the dataset [6].

2.3.2 Surrogate models for thermal simulation

A research direction particularly relevant to the present work is the use of machine learning models as *surrogate models* (or *metamodels*) of thermal simulation. In this approach, an ML model is trained on a dataset generated through simulation, learning to reproduce the results of the simulator at a fraction of the computational cost. This strategy combines the physical grounding of the training data with the prediction speed of ML models.

Recent studies confirm the effectiveness of this approach. For example, by training models on parametrically generated datasets for rectangular buildings, it has been shown that a Random Forest model can achieve a root mean squared error of approximately 2.7 kWh/m^2, delivering predictions in less than a thousandth of a second [8]. Other studies report computation time reductions of over 99% compared to direct simulation, while maintaining comparable accuracy

[1]. Similar strategies have been applied at the neighbourhood scale: a surrogate model based on polynomial regression, trained on 1000 configurations generated through Latin Hypercube sampling and accounting for microclimate and shading effects, enabled the rapid assessment of energy consumption in the early stage of urban design [9]. These results directly support the methodology adopted in the present work, in which the MLP models are trained on data generated with the COMFIE engine.

2.3.3 Integrating physical constraints

An emerging methodological direction is represented by Physics-Informed Neural Networks (PINNs), which incorporate physical laws directly into the training process, in the form of penalty terms in the cost function [10]. These hybrid approaches aim to combine the advantages of data-driven models with the robustness and extrapolation capability of physical models [11]. Although the application of PINNs in building energy modelling is still at an early stage, it represents a promising direction for future research.

2.4 The treatment of solar shading in the literature

Solar shading generated by neighbouring structures is recognised as an important factor in the energy balance of buildings, especially in dense urban environments. Simulation-based studies have quantified this effect, showing that ignoring shading can lead to significant errors in consumption estimates.

2.4.1 The quantified impact of shading

The results in the literature converge on the idea that inter-building effects have a substantial impact. It has been shown that mutual occlusion and reflections between buildings can influence heating and cooling loads by up to 65%, and lighting loads by approximately 17% [12]. A case study carried out in Hong Kong suggests that neglecting these effects can lead to a 13% gap in total energy consumption [12].

The concept of the *Inter-Building Effect* (IBE), introduced to describe the mutual influence between buildings in spatial proximity, has been systematically analysed in the literature. Dedicated studies have shown that ignoring mutual shading can generate significant errors in energy performance prediction, of up to 42% in summer and 22% in winter [13]. Separating the contribution of shading from that of inter-façade reflections has revealed that the shading effect has a larger impact on energy consumption than that of reflections [14].

Shading affects not only thermal solar gains, but also daylight availability. Simulation-based studies in dense urban environments have demonstrated that neighbouring obstructions significantly influence the energy savings associated with daylighting, especially for the perimeter zones of buildings [15].

The direction of the impact depends strongly on orientation and climate. Shading on the southern façade significantly increases the heating and lighting demand in cold climates, whereas shading on the eastern and western façades reduces the cooling demand in hot regions [16]. Urban shading simulation studies show that omitting shading can lead to an average underestimation of the heating demand by approximately 13% during winter and to an overestimation of the cooling demand by about 15% in summer [12]. These values are consistent with the results obtained in the present work, which highlight the dominant role of the southern orientation during the heating season (Section 6.3).

2.4.2 Shading in data-driven models

An essential aspect, recently highlighted in the literature, is that most data-driven UBEM studies have focused predominantly on building attributes, weather data and historical consumption, without taking into account relational urban factors such as mutual shading or neighbourhood density [17]. While physics-based UBEM modelling has always incorporated information about the urban context, data-driven approaches have often treated shading in a simplified manner or neglected it altogether.

A large-scale study carried out on more than 250,000 buildings in Seoul investigated precisely this aspect, demonstrating that the inclusion of simulation-generated shading indices improves the performance of data-driven prediction models [17]. This study also raises the question of the optimal structure of the shading descriptors — per exterior wall orientation or per time interval — for the most efficient modelling.

2.5 Positioning of the present work

Several observations emerge from the literature review presented above, which underpin the contribution of the present work.

First, although ML-based surrogate models for thermal simulation are well documented, few studies isolate the effect of solar shading from the other envelope parameters in a controlled manner. The strategy adopted in this work, of reusing identical building configurations across all shading scenarios (Section 5.1.2), allows a rigorous isolation of this effect, which constitutes a distinctive methodological element.

Second, the recent literature highlights that shading is frequently underrepresented in data-driven models [17]. The present work directly addresses this gap, systematically comparing two MLP models — one without and one with explicit shading descriptors — and quantifying the resulting improvement in predictive accuracy.

Third, unlike many studies that focus on a single aggregated indicator, the present work analyses in detail the variation of the shading impact as a function of the obstruction geometry (azimuth and distance) and of its interaction with the envelope properties (Sections 6.3 and 6.4), providing an in-depth physical perspective on the mechanisms involved.

Therefore, the present work falls within the research direction of surrogate models informed by geometric context, contributing a controlled and detailed analysis of the role of solar shading in the prediction of the heating demand of buildings.

CHAPTER 3
THEORETICAL FOUNDATIONS OF NEURAL NETWORKS

The predictive model used in this work is a multilayer perceptron artificial neural network. In order to substantiate the methodological choices described in the following chapters, this chapter presents the essential theoretical concepts: the structure of the artificial neuron, its organisation into layers, the activation function used and the principle of network training.

3.1 The artificial neuron

The elementary computational unit of a neural network is the artificial neuron, a mathematical model inspired by the functioning of the biological neuron. A neuron receives a set of input values $x_1, x_2, \ldots, x_n$, each weighted by a coefficient w_i called a *weight*. The neuron computes the weighted sum of the inputs, to which it adds a constant term b called the *bias*:

$$z = \sum_{i=1}^{n} w_i x_i + b \tag{3.1}$$

The result z is then passed through an *activation function* f, which introduces the non-linearity required for the network to be able to model complex relationships:

$$y = f(z) = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{3.2}$$

Figure 3.1 illustrates this model schematically. The weights w_i and the bias b constitute the parameters that are adjusted during the training process, so that the network produces the desired outputs.

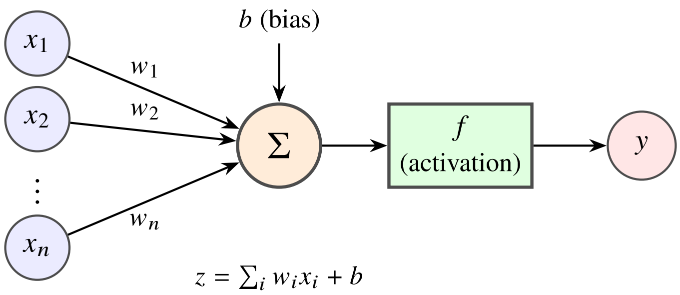


Fig. 3.1: The artificial neuron model: the inputs x_i are weighted by the coefficients w_i, summed together with the bias term b, and the result z is passed through an activation function f to produce the output y.

3.2 The architecture of the multilayer perceptron

A Multilayer Perceptron (MLP) is a neural network in which the neurons are organised into successive layers: an *input layer*, one or more *hidden layers* and an *output layer*. Each neuron in a layer is connected to all the neurons in the next layer, a configuration called *fully connected*. Information flows in a single direction, from input to output, which is why the MLP belongs to the category of *feedforward* networks.

The representational capacity of an MLP network depends on its depth (the number of hidden layers) and on its width (the number of neurons per layer). A deeper network can, in principle, model more complex relationships, but at the risk of increasing the number of parameters and of overfitting.

A relevant result of the present work, illustrated in Figure 3.2, is that the two compared models require architectures of very different complexity. The model trained without shading descriptors requires a deep architecture, with nine hidden layers of 80 neurons each, in order to implicitly infer the shading effects from the other variables. In contrast, the model that explicitly receives the shading information can achieve superior accuracy with a much simpler architecture, having a single hidden layer with 20 neurons. This drastic reduction in the required complexity constitutes indirect evidence of the informational value of the shading descriptors.

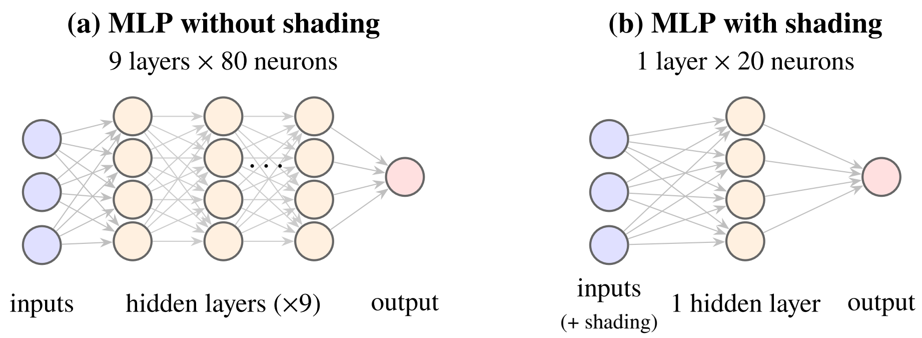


Fig. 3.2: The architecture of the two MLP models used. The model without shading descriptors (a) requires a deep architecture (9 hidden layers), whereas explicitly providing the shading information allows a much simpler architecture (b), with a single hidden layer. The number of input neurons is reduced here for visual clarity.

3.3 The activation function

The activation function determines how a neuron responds to the weighted sum of its inputs. Without a non-linear activation function, a network with any number of layers would be equivalent to a simple linear transformation, incapable of modelling complex relationships.

In the present work, the SELU (*Scaled Exponential Linear Unit*) activation function is used, defined by:

$$\mathrm{SELU}(x) = \lambda \begin{cases} x, & \text{if } x > 0 \\ \alpha\,(e^x - 1), & \text{if } x \leq 0 \end{cases} \tag{3.3}$$

where $\lambda \approx 1.0507$ and $\alpha \approx 1.6733$ are fixed constants. For positive values, the function is linear, whereas for negative values it saturates exponentially towards an asymptotic value, as can be seen in Figure 3.3.

The distinctive feature of the SELU function lies in its *self-normalising* property: under certain conditions regarding weight initialisation (the `lecun_normal` initialiser, used in this work), the activations within the network automatically tend to maintain a mean close to zero and unit variance across the layers. This property contributes to the stability of the training process, especially for deep networks.

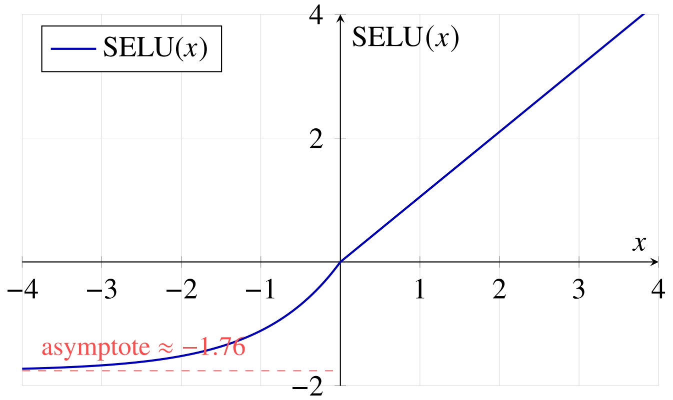


Fig. 3.3: The SELU (*Scaled Exponential Linear Unit*) activation function. It is linear for positive inputs, while for negative inputs it saturates exponentially towards an asymptotic value, a property that contributes to the self-normalisation of the activations within the network.

3.4 Training the network

Training a neural network consists of iteratively adjusting the parameters (weights and biases) so that the network's predictions come as close as possible to the actual values. This process relies on minimising a *loss function*, which quantifies the error between the predictions and the target values.

For regression problems, such as the prediction of heating demand, the usual loss function is the Mean Squared Error (MSE):

$$\mathrm{MSE} = \frac{1}{N}\sum_{j=1}^{N}\left(y_j - \hat{y}_j\right)^2 \tag{3.4}$$

where y_j are the actual values, $\hat{y}_j$ the network's predictions, and N the number of examples.

3.4.1 Backpropagation of the error

The minimisation of the loss function is carried out through the *backpropagation* algorithm. It operates in two stages repeated at every iteration. In the *forward pass* stage, the input data traverse the network and the prediction and the associated error are computed. In the *backward pass* stage, the error is propagated from the output layer towards the input layers, computing the gradient of the loss function with respect to each parameter by applying the chain rule of differentiation. The parameters are then updated in the direction opposite to the gradient, in order to reduce the error.

The magnitude of the update step is controlled by the *learning rate*, an important hyperparameter: a value that is too large can destabilise training, while a value that is too small slows it down excessively. The actual parameter update is handled by an *optimiser*; in this work, the

Adam optimiser is used, a stochastic gradient method with adaptive learning rates, recognised for its efficient convergence.

3.4.2 Overfitting and early stopping

A central challenge in training neural networks is *overfitting*: the situation in which the model memorises the particularities of the training data, including noise, instead of learning generalisable relationships. An overfitted model performs excellently on the training data but poorly on new data.

To prevent this phenomenon, the present work uses the *early stopping* strategy: training is monitored on a separate validation set, and the process is stopped when the error on this set no longer decreases, even if the error on the training set would continue to decrease. This technique ensures that the model with the best generalisation capability is retained.

In addition, the quality of the input data plays an essential role. The normalisation of the input variables (described in Section 5.1.2) ensures a balanced contribution of all features to the optimisation process, preventing the process from being dominated by variables with large numerical scales.

CHAPTER 4
THE PLÉIADES SOFTWARE AND THE COMFIE CALCULATION ENGINE

The entire database used in this work was generated through dynamic thermal simulation in the *Pléiades* software. Since the validity of the results obtained through machine learning depends directly on the physical fidelity of the training data, this chapter presents the theoretical foundations and the operating principle of the simulation software, as well as the justification for choosing it as the source of reference data.

4.1 General overview

Pléiades is a building energy simulation software suite, intended for eco-design and energy performance optimisation, widely used in France by architects and technical design offices. The software is developed and distributed by the French company IZUBA énergies, in collaboration with the Centre for Energy, Environment and Processes (*Centre Énergie, Environnement, Procédés* – CEEP) of *Mines Paris – PSL*.

The history of the software begins in the late 1980s – early 1990s, when the COMFIE dynamic thermal calculation engine was developed at the Energy Centre of the École des Mines de Paris, based on the model reduction techniques elaborated in that period [18]. Initially equipped with a command-line interface and used only by researchers, COMFIE was complemented in 1994 with a graphical interface, thus giving rise to the package known as Pléiades. Over time, the suite has been enriched with additional modules, and since 2012 it has integrated the regulatory calculation according to the French thermal regulation (RT2012, later RE2020).

The Pléiades suite integrates several complementary modules, of which the ones relevant to the present work are:

- **The geometric modelling module**, which allows the graphical description of the building level by level and its three-dimensional visualisation, including the calculation of shading and daylighting;
- **The COMFIE calculation engine**, which performs the multi-zone dynamic thermal simulation of the building;
- **The AMAPOLA module**, dedicated to statistical analysis, parametric studies, sensitivity and uncertainty analysis, as well as optimisation.

The choice of the Pléiades software for the present work is justified by three main considerations. First, the COMFIE engine is a scientifically validated tool, with a rigorous theoretical basis, developed in an academic setting and in use for more than three decades. Second, the model reduction technique on which it is based (described in Section 4.2) provides short computation times, which made the generation of the 32 000 simulations required by this study feasible. Third, the AMAPOLA module natively offers parametric variation functionalities, essential for the automatic construction of the database.

4.2 The COMFIE calculation engine

COMFIE is the dynamic thermal simulation (DTS) engine of the Pléiades suite. It computes the time evolution of the indoor temperatures and of the building's energy demand, based on a physical model of heat transfer through the envelope and between thermal zones. At each time step, the algorithm determines the heating and cooling demand, the humidity and the temperatures in each zone of the building, performing a thermal balance that integrates the exchanges between zones and takes into account the thermal inertia at the level of each wall. Figure 4.1 schematically illustrates this zoning principle.

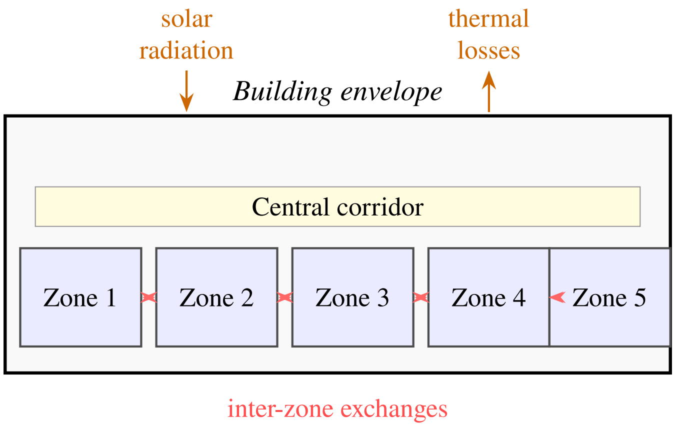


Fig. 4.1: Schematic representation of the multi-zone thermal zoning: each room constitutes a thermal zone for which COMFIE solves the energy balance, taking into account the exchanges between zones, the solar gains and the losses through the envelope.

4.2.1 The numerical method: nodal model reduced through modal analysis

From a mathematical point of view, COMFIE is based on a multi-zone **finite difference** model, to which a **reduction technique through modal analysis** is applied, theoretically grounded by Peuportier and Blanc-Sommereux [18]. The building is discretised into a mesh of nodes, each associated with a control volume characterised by its own thermal capacity and by a temperature assumed to be homogeneous. The energy conservation equations are written at each node of the mesh, expressing the balance between the variation of the stored energy and the sum of the heat fluxes through advection, lumped convection and long-wave radiation.

The resulting system of differential equations is, in its complete form, of high dimension, which would lead to long computation times. To overcome this limitation, COMFIE resorts to a *modal analysis*: the system is projected onto the basis of its dominant eigenmodes, retaining only those modes that contribute significantly to the thermal dynamics of the building. This order reduction allows a substantial decrease in the computational cost while preserving the accuracy of the predictions. Non-linear phenomena or phenomena with time-varying parameters (such as solar gains or ventilation schedules) are reintroduced in the actual simulation stage.

In matrix form, the reduced model of each zone can be expressed synthetically through the state system:

$$C\,\frac{\mathrm{d}T}{\mathrm{d}t} = A\,T + E\,U, \qquad Y = J\,T + G\,U \tag{4.1}$$

where T is the vector of nodal temperatures, U the vector of excitations (climate, internal gains, ventilation), Y the output quantities (zone temperatures, energy demand), and C, A, E, J, G are the matrices describing the thermal properties of the model. The reduced modal models of each thermal zone are then coupled in order to build the global system of the multi-zone building.

The time step usually employed is half an hour, with the possibility of reducing it to a few minutes for a finer temporal resolution.

4.2.2 Output quantities

For each thermal zone, COMFIE provides the annual and hourly heating and cooling demand, the hourly and average temperatures, as well as thermal comfort indicators (of the PMV/PPD type). Within the present work, the quantity of interest is the annual heating demand, expressed in kWh and subsequently normalised to the building's floor area (Section 5.1.2).

4.3 The treatment of solar shading

Since the central subject of this work is the influence of solar shading, the way in which the software treats this phenomenon is of particular relevance.

Within the geometric modelling module, for each exterior wall of the building a *solar mask* is automatically generated, based on the geometry entered by the user. This mask describes the obstructions to solar radiation produced by elements belonging to the building itself (for example, overhangs or façade setbacks) or by obstacles in the vicinity (neighbouring buildings, terrain). The software distinguishes two types of masks: *near masks*, generated by obstacles in the immediate vicinity (such as the obstruction building modelled in the present study), and *far masks*, associated with the terrain or the horizon. The geometric principle of this mechanism is illustrated in Figure 4.2.

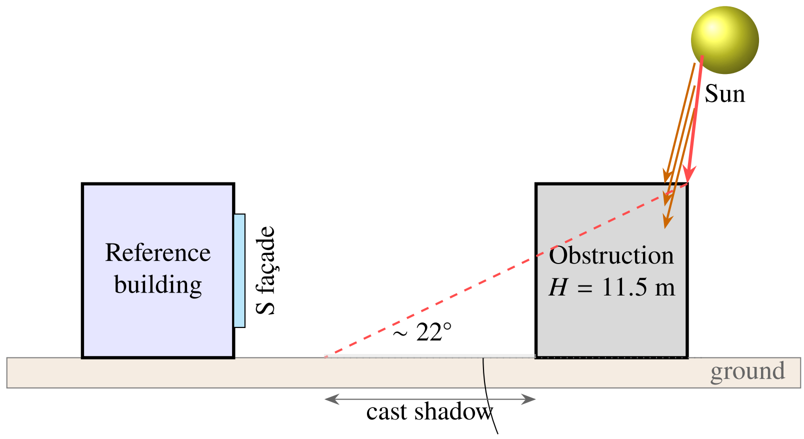


Fig. 4.2: The shading calculation principle: at the latitude of Bucharest, the maximum solar altitude at the winter solstice is approximately 22°. An obstruction located to the south casts a shadow that reduces the direct solar radiation incident on the southern façade of the reference building.

The calculation of the incident radiation, both direct and diffuse, takes these masks into account for every surface of the envelope. Thus, when an obstruction is introduced into the model, the software automatically recalculates the solar flux that effectively reaches the building's façades, reducing the solar gains in accordance with the geometry of the obstruction. This mechanism underlies the methodology described in Section 6.3, in which the shading obstruction is systematically moved around the reference building.

The software's three-dimensional rendering also allows the visualisation of the cast shadows and of the sunshine duration on the façades, providing a visual check of the consistency of the geometry entered.

4.4 The scientific validation of the COMFIE engine

The reliability of a simulation tool is an essential condition for its use as a source of reference data. Over time, the COMFIE engine has undergone several validation procedures, which support its scientific credibility:

- **Experimental validation** through comparison with measurements carried out in instrumented test cells, in particular on the INCAS platform and within the PASSYS test cells [19];
- **Intercode comparisons**, through the standardised BESTEST procedure of the International Energy Agency (IEA), which confronts the results of different thermal simulation programs on common reference cases;
- **Compliance with international standards**, the engine having been evaluated against the ISO 52016-1 and ASHRAE 140 standards;
- **Application in peer-reviewed studies**, both in academic research and in French regulatory practice (RT2012 and RE2020 compliance calculations).

These successive validations, together with the computational efficiency provided by the modal reduction technique, make COMFIE a suitable tool for generating a large volume of physically consistent simulations, as required by the methodology of the present work.

4.5 The workflow used in this work

The database generation process followed the steps below, integrating the three modules described previously and summarised in Figure 4.3:

1. **Geometric modelling** of the reference building (described in Section 5.1.2), including the definition of the thermal zones, façades and windows;
2. **Definition of the parametric variations** through the AMAPOLA module, for the input variables (insulation, ventilation, glazing), sampled according to the strategy described in Section 5.1.2;
3. **Introduction of the shading obstruction**, systematically moved through the 15 azimuth–distance configurations defined in Section 6.3;
4. **Launching the dynamic thermal simulations** with the COMFIE engine, for each of the 32 000 configurations (2 000 envelope configurations × 16 shading states: one baseline and 15 with an obstruction);
5. **Extraction of the annual heating demand** for each simulation and export of the results as the data files subsequently used for training the machine learning models.

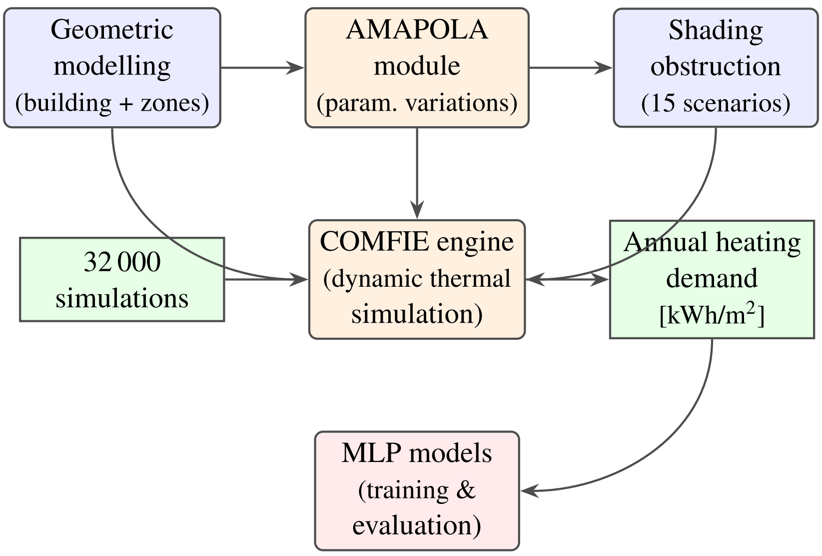


Fig. 4.3: The workflow used for generating the database: geometric modelling and definition of the parametric variations (AMAPOLA), followed by the dynamic thermal simulation (COMFIE) and the export of the heating demand for model training.

This automated procedure, exploiting the ability of the AMAPOLA module to handle massive parametric variations and the computation speed of the COMFIE engine, made it possible to build a large and physically consistent dataset, which constitutes the foundation of the analyses presented in the following chapters.

CHAPTER 5
METHODOLOGY

5.1 The reference building model and the simulation database

5.1.1 The baseline dataset without shading

The geometry and parameters of the reference building

The reference building was modelled using the *Pléiades* thermal simulation software. A four-storey building template with nearly identical levels was used, slightly adapted for the purpose of this study.

Figures 5.1 and 5.2 show, respectively, the three-dimensional exterior view of the building model and the two-dimensional plan of the ground floor.

Each storey consists of two rows of five adjacent rooms, each with an individual floor area of 17.04 m^2. The two rows are separated by a central corridor 1 m wide. The building is oriented strictly towards the south. Windows are placed on all four façades, providing exterior exposure for every room.

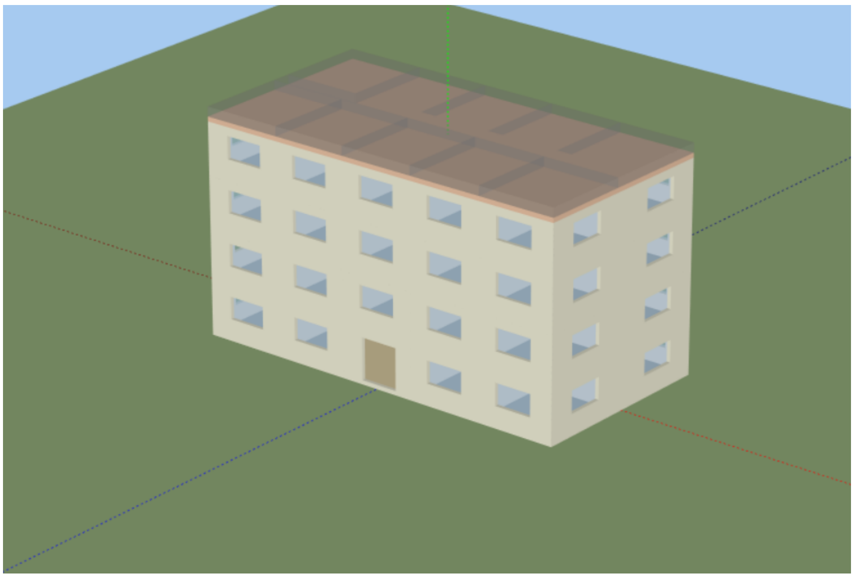


Fig. 5.1: Three-dimensional view of the reference building model

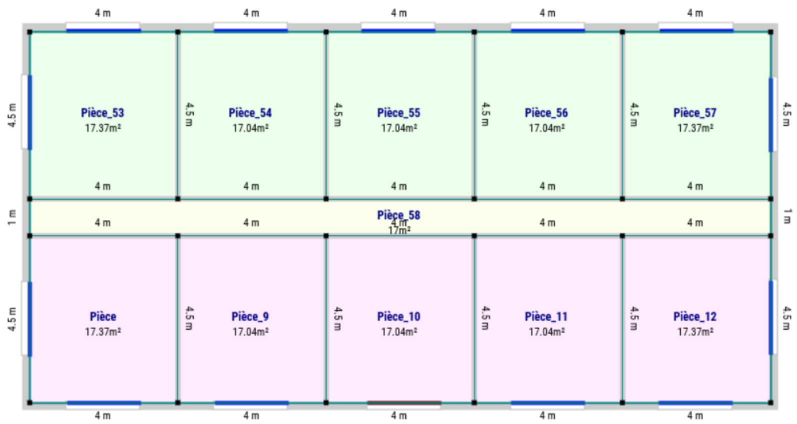


Fig. 5.2: Ground floor plan of the reference building

The dataset generation process

To build the baseline dataset, several physical and operational variables were randomly sampled within predefined ranges. These variables are grouped into three categories: ventilation parameters, insulation properties and glazing characteristics.

For each sampled configuration, a dynamic thermal simulation was carried out over a full year, using the AMAPOLA engine integrated in *Pléiades*.

The dataset was generated as three subsets of equal size:

- **First subset:** the parameter values were sampled using a uniform distribution through the Latin Hypercube Sampling method, ensuring broad coverage of the input space.
- **Second subset:** the insulation was set to zero for all simulations, while the other variables were sampled according to a normal distribution.
- **Third subset:** the insulation was set to zero and the ventilation was fixed at its maximum value.

Following this procedure, a total of 2000 simulations were obtained. This baseline dataset represents the heating demand of the reference building under conditions in which no external shading is present.

5.1.2 The shading dataset

Definition of the shading scenarios

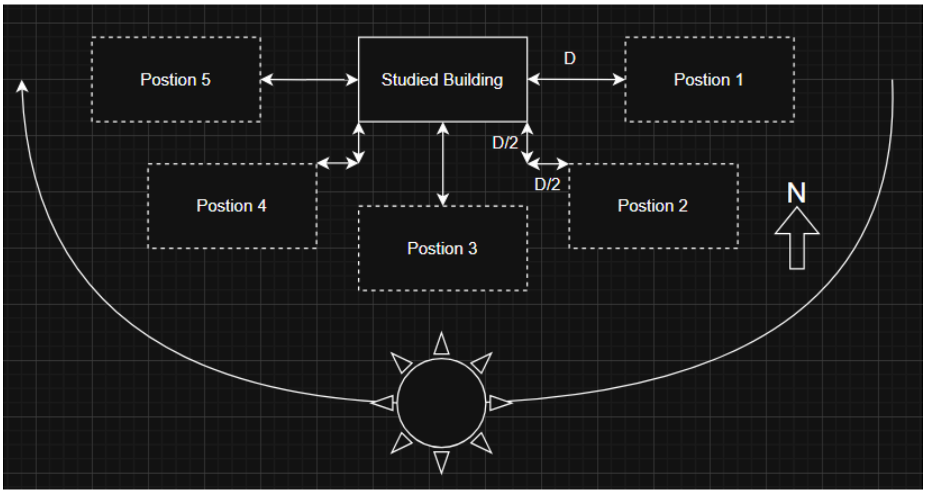
In order to investigate the impact of neighbouring structures on the energy demand, a parametric study involving a single shading obstruction was carried out. This obstruction is modelled as a rectangular parallelepiped, with dimensions identical to those of the reference building (a footprint of 20×10 m) and a fixed height $H = 11.5$ m.

The position of the obstruction is defined relative to the reference building using two variables: the azimuth angle and the separation distance D.

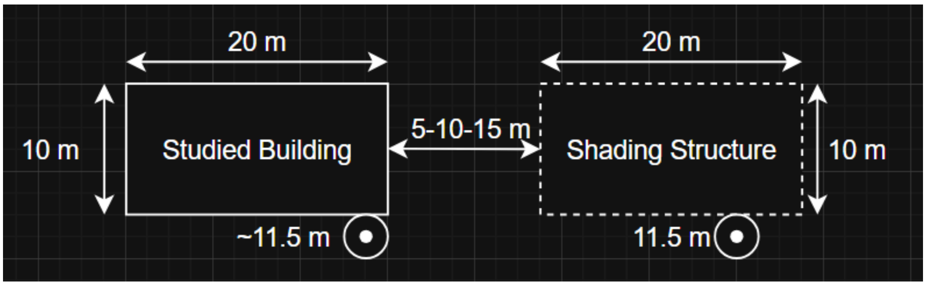
- **Azimuth:** Five positions relative to the south direction are considered: $-90°$ (West), $-45°$ (South-West), $0°$ (South), $+45°$ (South-East) and $+90°$ (East).

- **Distance (D):** Three façade-to-façade distances are tested: 5 m, 10 m and 15 m.

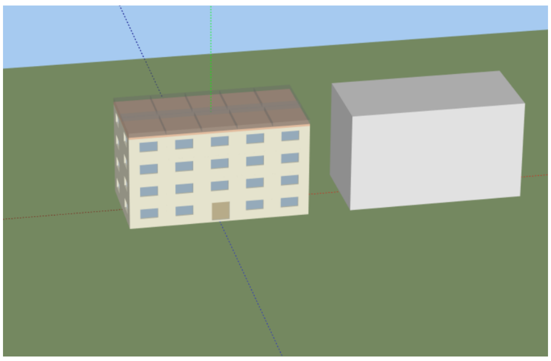
For the cardinal directions, D represents the perpendicular distance. For the diagonal orientations (±45°), a Manhattan-distance-type logic is applied, with an offset of $D/2$ on both axes.



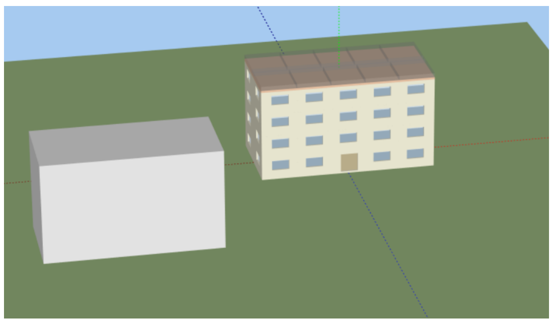
(a) Top view of the 5 angular positions



(b) Definition of the distances (D) and the geometry of the configuration



(c) East, D = 5 m



(d) South-West, D = 15 m

Fig. 5.3: The geometric configuration of the shading scenarios. The shading structure (H = 11.5 m) is identical to the reference building.

This leads to a total of 15 distinct shading scenarios (5 positions × 3 distances).

The dataset extension strategy

A specific sampling strategy was used in order to isolate the effect of shading from the influence of the building envelope parameters. Instead of generating new, random building configurations for the shading scenarios, exactly the same 2000 building configurations generated for the baseline dataset were reused.

For each of the 15 shading scenarios, the 2000 simulations were re-run with the addition of the obstruction. This "cloning" approach ensures that any variation in energy performance is attributed strictly to the shading effect, since the ventilation rates, the insulation thicknesses (parameters E2–E10) and the glazing properties (E11, E19) remain strictly identical between the baseline and the shaded case.

Consequently, the total dataset comprises 32 000 simulations:

- 2 000 baseline simulations (without shading).
- 30 000 shaded simulations (15 scenarios $\times$ 2 000 buildings).

The training/test split established in the baseline phase (1600 for training / 400 for testing) is preserved for all scenarios. The model is trained and evaluated using the same corresponding building identifiers, in order to prevent information leakage between the datasets.

The simulation output

The target variable is the annual heating demand [kWh] required to maintain a reference indoor temperature of 21°C. This value is normalised by the total floor area of the building (754.8 m^2), resulting in a consumption expressed in kWh/m^2. The shading masks are opaque and oriented at 0° (South), significantly influencing the direct solar radiation.

5.1.3 Renovation scenario

Building envelope and window upgrade measures

The renovation scenario is introduced as a robustness test subsequent to the training phase. To keep the analysis consistent with the learning framework, the comparison is carried out exclusively at the level of the model inputs. All geometric descriptors and window-to-wall ratios are kept unchanged, while the insulation properties of the envelope and the glazing characteristics are modified to represent an upgraded building.

Table 5.1: The renovation scenario: the modified model input variables (before / after).

Parameter	Unit	Before	After
E2: Interior insulation thickness (ext. walls)	cm	0.0	2.0
E3: Exterior insulation thickness (ext. walls)	cm	15.0	10.0
E6: Exterior insulation thickness (stairwell)	cm	0.0	0.01
E7: Interior insulation thickness (roof)	cm	0.0	2.0
E8: Exterior insulation thickness (roof)	cm	15.0	15.0
E10: Insulation thickness (ground floor slab)	cm	15.0	2.0
E11: Thermal transmittance of windows (North)	$W/(m^2K)$	2.606	1.095
E19: Solar factor of the glazing (North)	–	0.502	0.533
E31: Air change rate (typical floors)	h^{-1}	1	1
E32: Air change rate (ground floor/corridor)	h^{-1}	1	1
G4: Heating active on the ground floor	binary	1	1

The renovation scenario mainly affects the parameters associated with the building envelope insulation and the glazing, while the geometric, ventilation and shading parameters remain unchanged. Interior insulation layers are introduced for the exterior walls and the roof, while the exterior wall insulation is reduced by approximately 33%. The most pronounced change concerns the ground floor slab insulation (E10), which decreases by almost 87%. The thermal transmittance of the windows (E11) is reduced by approximately 58%, reflecting a substantial improvement in glazing performance. The solar factor of the glazing (E19) increases slightly, by approximately 6%, modifying the solar gains. Overall, the renovation induces significant relative variations in a limited set of model input variables, providing a relevant robustness test for the prediction models.

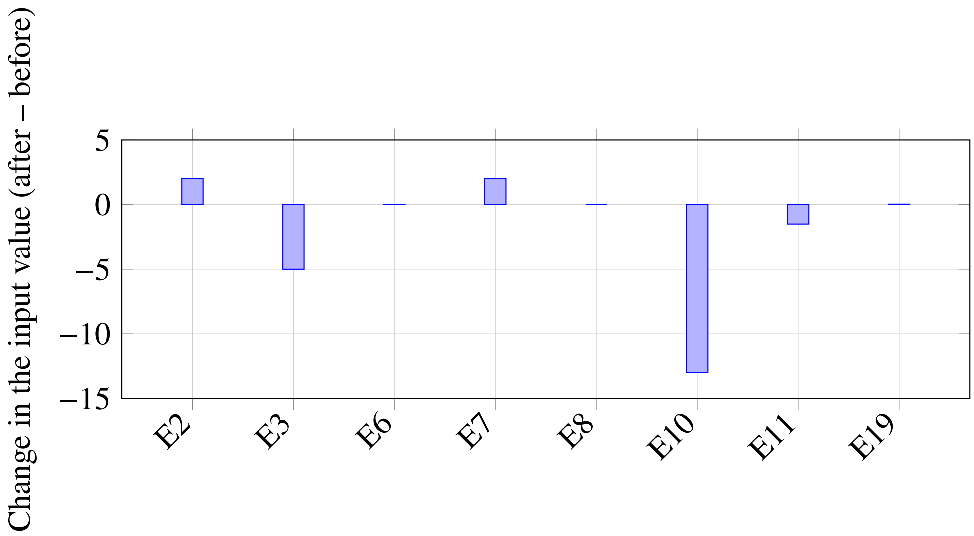


Fig. 5.4: The change in the model input variables under the renovation scenario.

5.2 Data preprocessing

5.2.1 Data transformation

Before training the neural network, the raw dataset was preprocessed using a unified transformation pipeline. The input variables include numerical, categorical and binary descriptors of the building configuration. In order to ensure consistent scaling and a stable learning process, the data were processed through a `ColumnTransformer` composed of four elements:

- **Indicator transformation:** applied to the eight parameters associated with the glazing and the window-to-wall ratios (*Ratio fenêtre Nord, Sud, Est, Ouest* and their counterparts for the ground floor). This transformation generates binary or threshold-based indicators, which help the model capture abrupt changes in the façade configuration.
- **Standardisation:** all the other numerical variables were normalised using a z-score transformation, in order to ensure a balanced contribution during the optimisation process.
- **One-hot encoding:** the categorical variable *Type de mur* (wall type) was encoded without introducing artificial ordinal relationships.
- **Passthrough:** the binary variable *Chauffage RDC* (ground floor heating) was kept unchanged, since no scaling was necessary.

5.2.2 Training, validation and evaluation metrics

Each candidate MLP architecture is trained for up to 100 epochs, using an early stopping strategy based on the validation set.

After the hyperparameter tuning stage, the best-performing model is evaluated on the test set using the root mean squared error (RMSE) and the coefficient of determination (R^2).

These indicators provide a comprehensive assessment of the model's accuracy and of its generalisation capability in the no-shading scenario.

5.3 Network architecture and hyperparameters

The predictive model used is a Multilayer Perceptron (MLP), whose architecture is automatically tuned using `Keras Tuner` through a `RandomSearch` strategy.

The hyperparameter search explores:

- the number of hidden layers: from 1 to 10,
- the number of neurons per layer: from 20 to 500,
- the activation function: SELU (fixed),
- the weight initialiser: `lecun_normal`,
- the learning rate: from 10^{-4} to $3 \cdot 10^{-2}$,
- the optimiser: stochastic gradient descent with a momentum term.

CHAPTER 6
RESULTS

6.1 Summary of the hyperparameter search

MLP without shading inputs
Optimal configuration (random search):

Hidden layers	**9**
Neurons/layer	**80**
Learning rate	**0.001**
Optimiser	**Adam**

MLP with shading inputs
Optimal configuration (random search):

Hidden layers	**1**
Neurons/layer	**20**
Learning rate	**0.01**
Optimiser	**Adam**

Fig. 6.1: The optimal hyperparameter configurations selected through random search for the two input scenarios.

The random hyperparameter search leads to a compact configuration for the model that includes shading information, consisting of a single hidden layer with 20 neurons, trained with the Adam optimiser and a learning rate of 0.01. This result suggests that, once the shading information is explicitly provided at the input level, the relationship between the input variables and the annual heating demand can be captured using a reduced model complexity. The same search procedure is also applied to the no-shading case, allowing a direct comparison of how the optimal model capacity depends on the chosen input space.

6.2 Model performance

This section compares the predictive performance of the two proposed models: (i) a baseline MLP model, trained without shading-related inputs, and (ii) an MLP model that explicitly includes geometric shading parameters. The performance of the models is evaluated using graphical indicators and standard regression metrics, computed on an independent test set.

6.2.1 The baseline model (without shading inputs)

The baseline model is evaluated first, in order to establish a reference level of performance. Figure 6.2 compares the predicted and simulated values of the annual heating demand on the test dataset.

The predictions are, in general, well aligned with the identity line, indicating a high overall agreement between the model outputs and the simulation results. However, a slight increase in dispersion can be observed for certain configurations, especially for those associated with pronounced solar shading. Since the shading information is not explicitly provided as a model input, the model is forced to implicitly infer these effects from the other variables, which can limit its robustness in situations with strong shading.

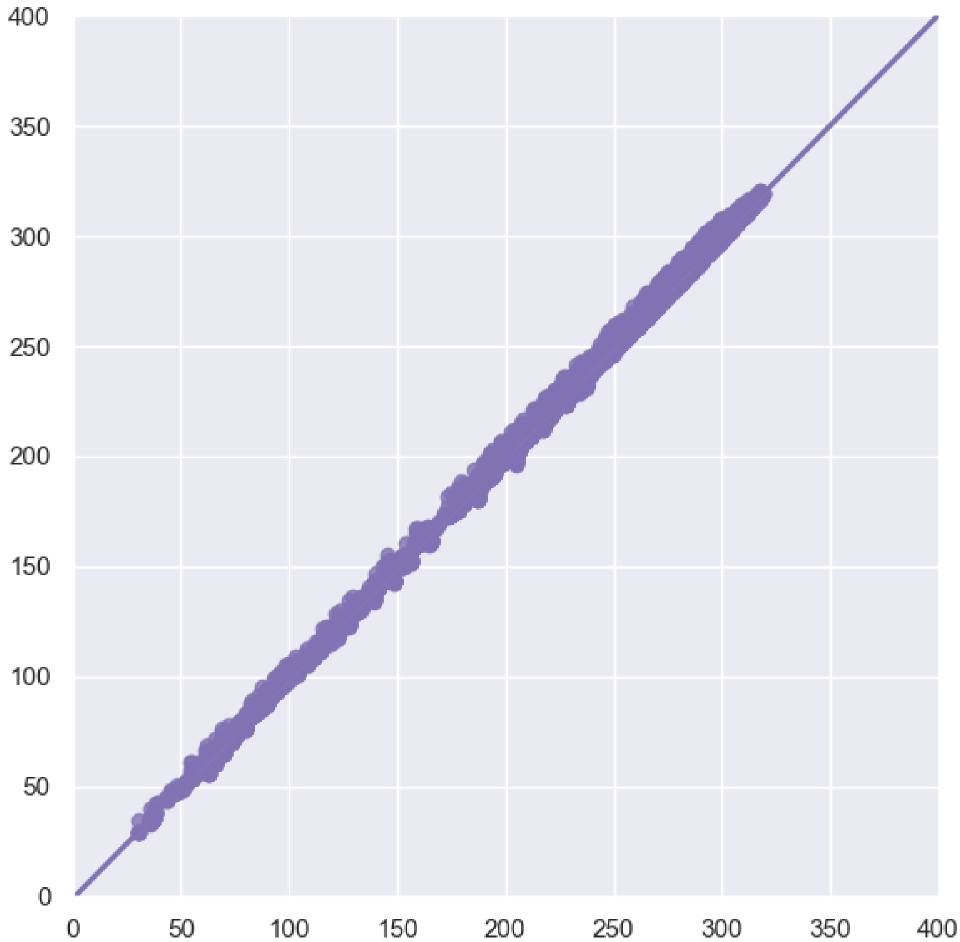


Fig. 6.2: Comparison of the predicted and simulated heating demand values for the baseline model, expressed in kWh/m^2 (horizontal axis: values simulated with COMFIE; vertical axis: values predicted by the model). The solid line represents the identity line.

6.2.2 The model including shading information

The model that includes shading information is evaluated using the same test dataset and the same performance indicators as the baseline model. Figure 6.3 shows the comparison between the predicted and simulated heating demand values.

Compared to the baseline configuration, the predictions are distributed more tightly around the identity line across the entire range of values. This reduced dispersion indicates that the explicit inclusion of shading-related inputs improves the model's ability to capture the variations in solar gains induced by neighbouring obstructions, leading to a more consistent predictive behaviour.

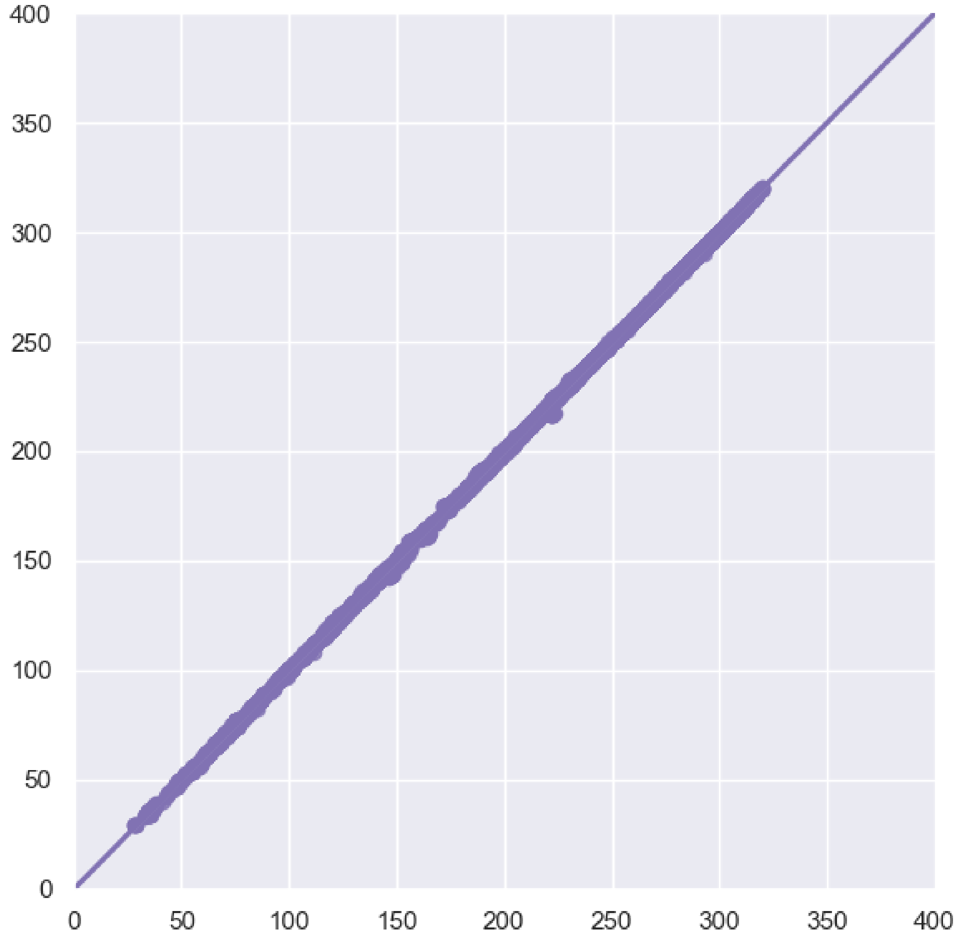


Fig. 6.3: Comparison of the predicted and simulated heating demand values for the model including shading information, expressed in kWh/m^2 (horizontal axis: values simulated with COMFIE; vertical axis: values predicted by the model).

6.2.3 Quantitative performance indicators

The quantitative performance indicators are summarised in Table 6.1. Both models achieve high predictive accuracy on the test set, as evidenced by the very high values of the coefficient of determination.

Table 6.1: The prediction performance of both models on the test dataset.

Model	RMSE	CV(RMSE) [%]	R^2
MLP without shading	2.1047	0.96%	0.9991
MLP with shading	0.5977	0.27%	0.9999

The model that includes shading information systematically outperforms the baseline model, achieving a reduction of approximately 72% in the RMSE value and a CV(RMSE) below 0.3%. These results confirm that integrating the shading information leads not only to an improvement in accuracy, but also to a more stable and more robust predictive behaviour under different operating conditions.

Both models achieve high predictive accuracy on the test set. However, the model that includes shading information consistently exhibits lower error levels and a more stable coefficient of determination, especially for the configurations in which shading significantly influences the heating demand.

6.2.4 The impact of the solar shading scenarios

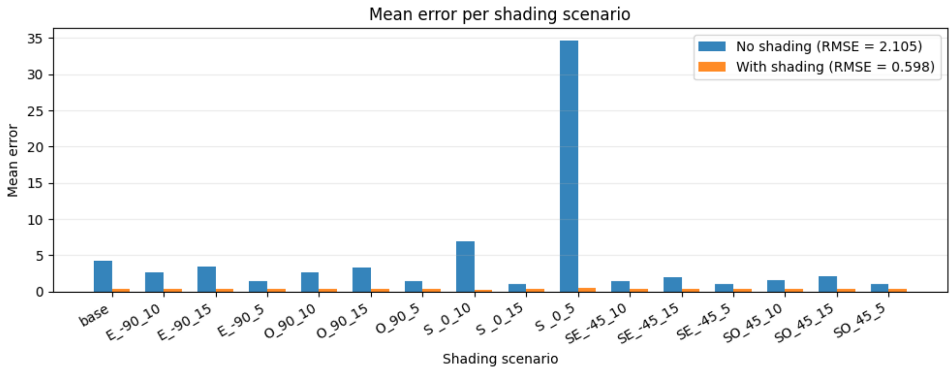


Fig. 6.4: The distribution of the prediction error for both models as a function of the shading scenarios.

6.2.5 Model interpretability

To improve the interpretability of the model, SHAP values were computed for the model including shading information. Figure 6.5 presents a global summary of the feature contributions to the predicted heating demand.

The envelope insulation thicknesses and the glazing-related parameters are identified as the dominant contributors, in agreement with physical expectations. The shading descriptors also show a significant contribution, with an impact that varies with the configuration. The dispersion of the SHAP values highlights the non-linear interaction between shading, envelope properties and heating demand, confirming that the geometric context plays a relevant role in the model's predictions.

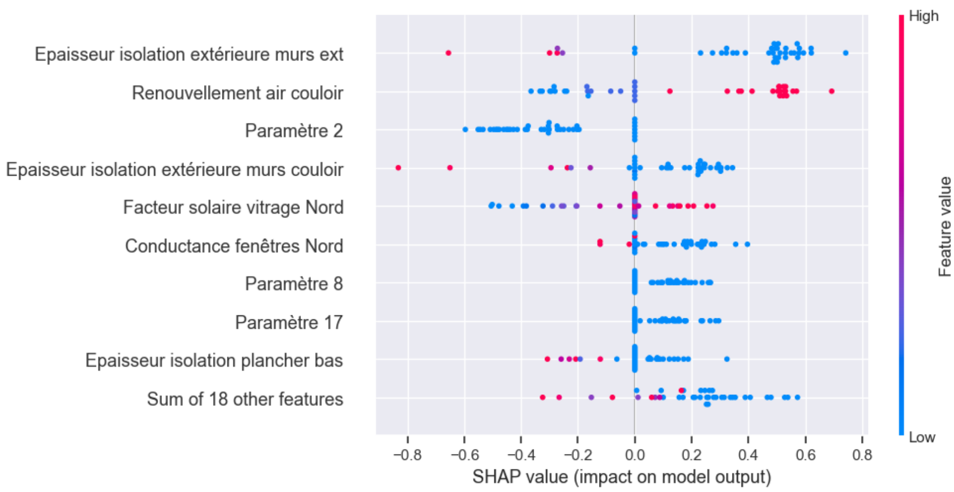


Fig. 6.5: SHAP summary plot for the model including shading information.

6.2.6 Performance under renovation conditions

Table 6.2: The predicted heating demand [kWh/m^2] before and after renovation, for a reference building.

Model	Demand before [kWh/m^2]	Demand after [kWh/m^2]
MLP without shading	108.59	109.77
MLP with shading	106.24	104.52

Finally, both models are evaluated under the renovation scenario, in order to analyse their behaviour when the building envelope and glazing characteristics are modified. Table 6.2 compares the heating demand predicted by each model, before and after the application of the renovation measures, for a reference building.

The behaviour of the two models differs in a significant and meaningful way. The model that includes shading information predicts a reduction of the heating demand following the renovation (from 106.24 to 104.52 kWh/m^2), a physically consistent result, since the improvement of the insulation and of the glazing should reduce the energy demand. In contrast, the baseline model, without shading information, predicts a slight *increase* of the demand (from 108.59 to 109.77 kWh/m^2), a behaviour contrary to physical expectations. This difference suggests that the explicit inclusion of the shading descriptors gives the model a more coherent representation of the physical relationships, allowing it to respond correctly to envelope modifications, even under conditions different from those encountered during training.

6.3 Detailed analysis per shading scenario

The global indicators presented in the previous section summarise the model's accuracy over the whole set of scenarios, but they do not reveal how the physical impact of shading varies with the geometric configuration. To explore this dimension further, we analyse the simulation data directly: for each of the 2000 building configurations, we compare the annual heating demand obtained with and without the obstruction, for all 15 scenarios defined in Section 5.1.2.

We denote by Δ_s the increase in heating demand induced by scenario s:

$$\Delta_s = \frac{1}{A}\left(Q_s - Q_{\text{base}}\right) \quad [\text{kWh/m}^2] \tag{6.1}$$

where Q_s is the annual demand in scenario s, Q_{base} is the demand in the absence of shading, and $A = 754.8$ m^2 is the total floor area of the reference building. Since the 2000 envelope configurations are identical between Q_s and Q_{base}, Δ_s strictly isolates the geometric effect of the obstruction.

6.3.1 The spatial distribution of the impact

Figure 6.6 shows the heatmap of the mean value $\bar{\Delta}_s$ for the 15 azimuth–distance combinations. The pattern is strongly structured: the South position categorically dominates the other orientations, with an impact of **6.73 kWh/m^2 at 5 m** — approximately **27 times larger** than the least severe case (East at 15 m, $\bar{\Delta} = 0.25$ kWh/m^2). The perfect symmetry between the East–West and South-East–South-West orientations is consistent with the symmetric geometry of the building about the North–South axis and constitutes an implicit validation test for the consistency of the dataset.

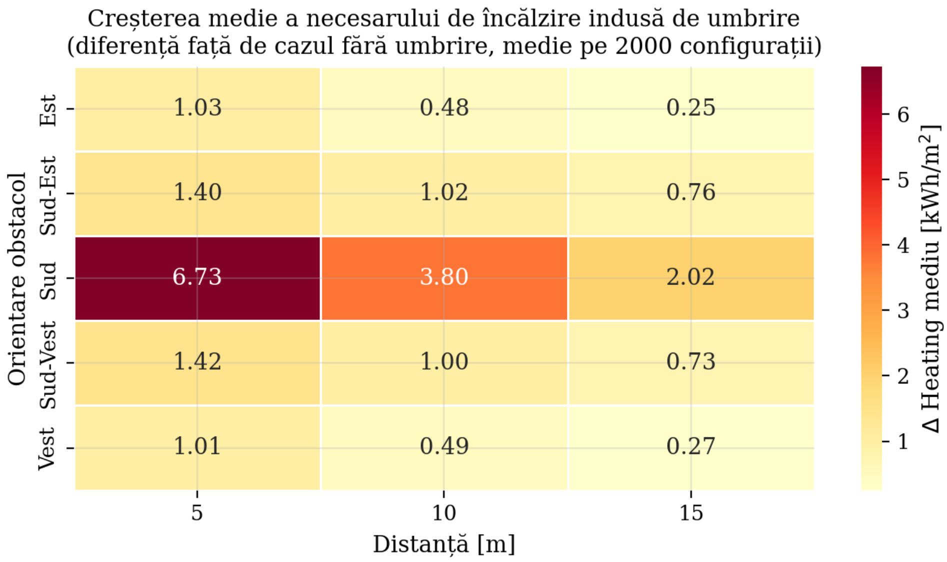


Fig. 6.6: Heatmap of the mean increase in heating demand ($\bar{\Delta}_s$) for the 15 shading scenarios, computed over 2000 envelope configurations. The South position at 5 m dominates, with an impact 27 times larger than the most favourable case (East at 15 m).

6.3.2 *The angular profile and the decay with distance*

Figure 6.7 illustrates the three impact curves parametrised by the separation distance. The shape of the curves suggests a sharp bell-shaped function centred on the South, with a rapid drop towards the lateral orientations: for $D = 5$ m, the impact decreases from 6.73 kWh/m^2 (South) to 1.40 kWh/m^2 (South-East/South-West), i.e. a reduction of almost 80% for an angular variation of only 45°.

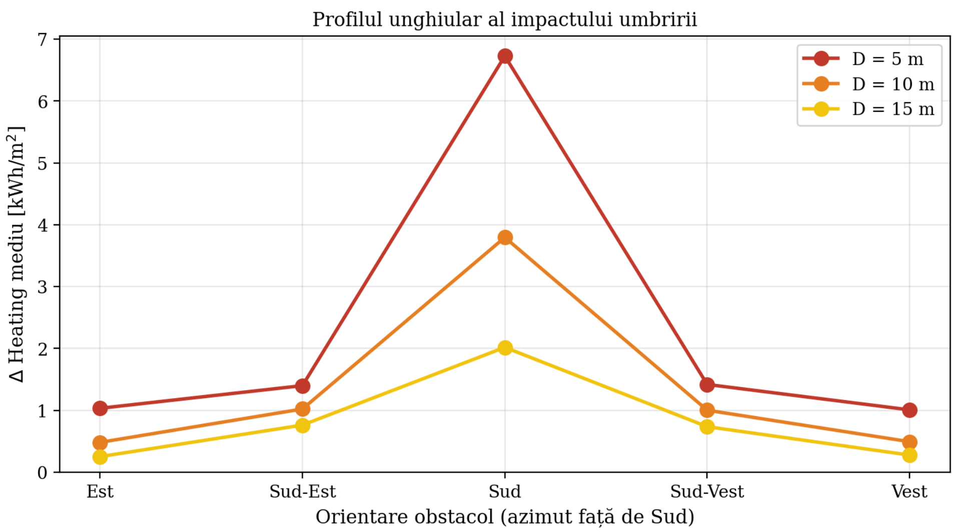


Fig. 6.7: The angular profile of the shading impact for the three tested distances. The bell shape centred on the South reflects the predominance of the southern façade in capturing solar gains at the latitude of Bucharest (~ 44.4° N).

This azimuth–distance asymmetry has a direct physical explanation: at the latitude of Bucharest (44.4° N), the solar path during the heating season (October–March) remains predominantly within the southern sector, with a maximum solar altitude of approximately 22° at the winter solstice. An obstruction located strictly to the south preferentially blocks the useful direct solar flux during the periods of maximum thermal contribution, whereas eastern or western obstructions affect only the hours with obliquely incident radiation, which is far less valuable energetically.

6.3.3 *Inter-configuration variability*

In order to assess not only the mean trend but also the dispersion of the impact over the population of 2000 configurations, Figure 6.8 shows the complete distribution of the Δ_s values grouped by distance. Two observations are relevant.

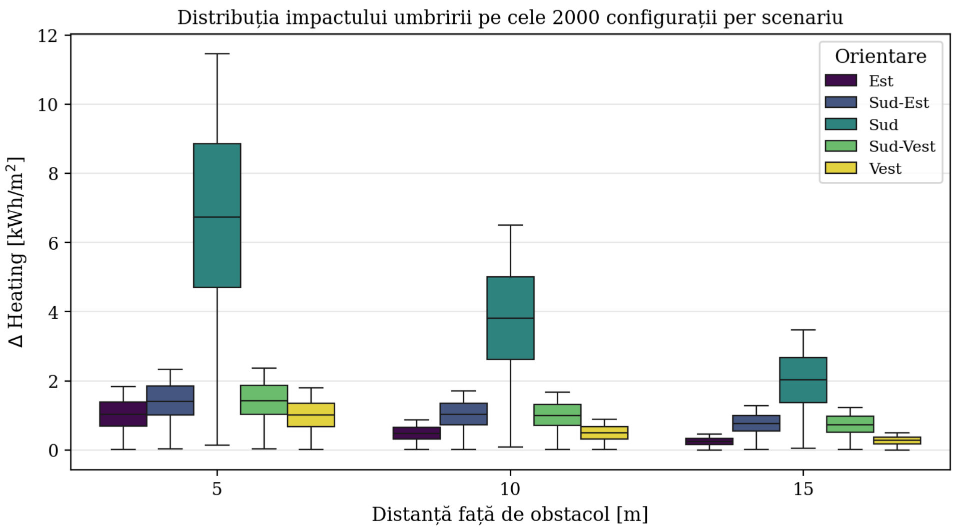


Fig. 6.8: The distribution of the shading impact Δ_s over the 2000 configurations, for each scenario (3 distances × 5 azimuths). The South scenario at 5 m exhibits both the highest median value and the largest inter-configuration dispersion.

First, the South scenario at 5 m has not only the largest mean, but also the largest standard deviation (2.68 kWh/m^2), which indicates a strong sensitivity to the envelope characteristics. In other words, the impact of the same physical obstruction can vary from 0.14 kWh/m^2 to 11.46 kWh/m^2 depending on the configuration of the shaded building — 71.5% of the configurations experience an increase of more than 5 kWh/m^2, and 14.1% exceed 10 kWh/m^2. Second, for the lateral scenarios (East, West), the distribution remains tightly concentrated near zero regardless of the distance, which confirms that the physical model correctly captures the lack of a significant solar contribution on these orientations during the heating season.

6.3.4 *Comparison of the extreme distributions*

Figure 6.9 overlays the histograms of the annual heating demand for the no-shading case and for the scenario with the largest impact (South, 5 m). The shift of the distribution towards the right is visible especially in the region of moderate values (200–280 kWh/m^2), corresponding to buildings with medium insulation, for which the solar gains make an important relative contribution to the energy balance. For very poorly insulated buildings (>290 kWh/m^2), the effect remains measurable but relatively attenuated, since the losses through the envelope dominate the balance.

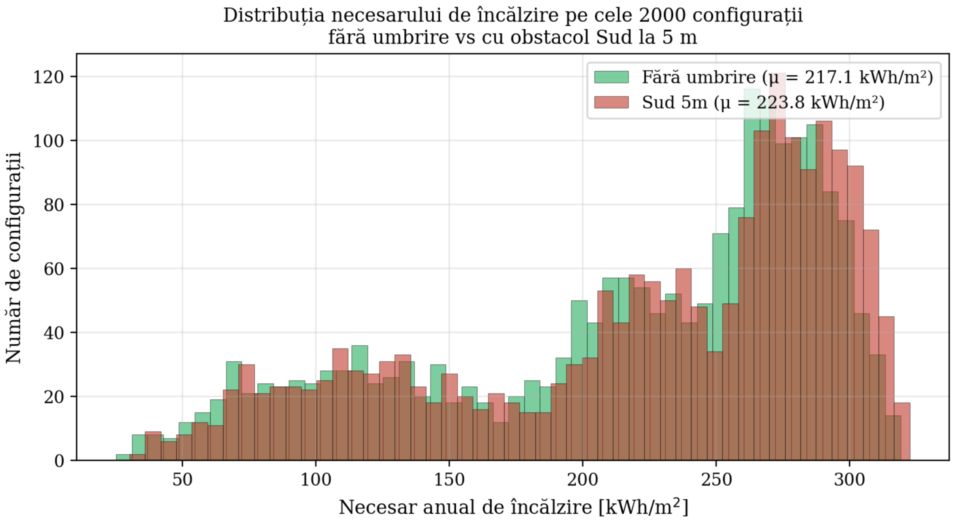


Fig. 6.9: The distribution of the annual heating demand over the 2000 configurations in the absence of shading (green) and under an obstruction placed strictly South at 5 m (red). The overall mean increases by 6.7 kWh/m^2 (+3.1%).

6.4 The interaction between envelope insulation and shading

The scenario-level analysis suggests that the impact of shading is not uniform over the building population, but depends on the envelope characteristics. To isolate and quantify this interaction, we group the 2000 configurations by the exterior wall insulation thickness (E_3) and analyse the distribution of Δ_s in the most unfavourable scenario (South, 5 m).

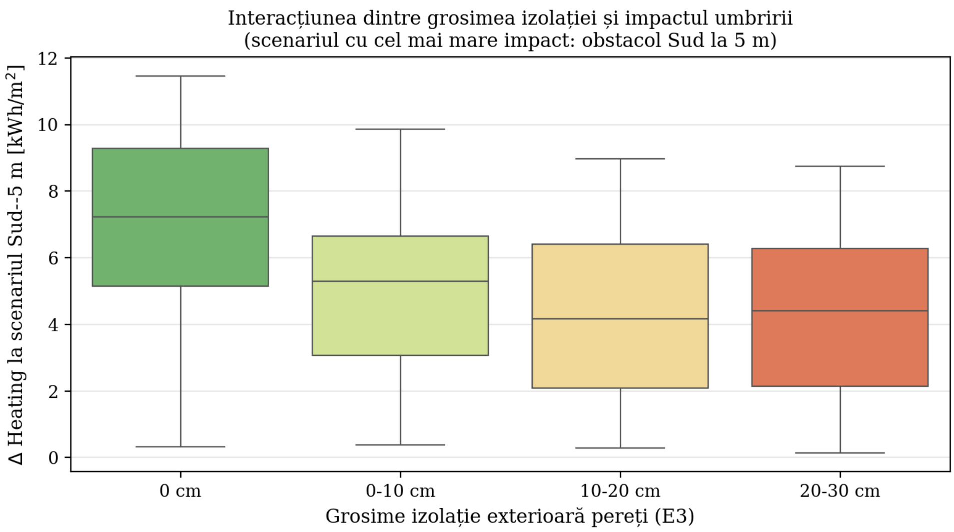


Fig. 6.10: The shading impact as a function of the exterior wall insulation thickness (E_3), for the South scenario at 5 m. As the insulation thickness increases, the absolute impact of shading decreases significantly.

The quantitative result is summarised in Table 6.3. The mean impact of shading decreases from 7.17 kWh/m^2 (for buildings without exterior insulation, $E_3 = 0$) to 4.29 kWh/m^2 (for those with $E_3 > 10$ cm), a **relative reduction of approximately 40%**.

Table 6.3: The mean shading impact (South scenario at 5 m) as a function of the exterior wall insulation thickness.

Insulation thickness E_3	No. of config.	$\bar{\Delta}$ [kWh/m^2]	σ [kWh/m^2]
0 cm	1667	7.17	2.51
0–10 cm	109	4.98	2.37
10–20 cm	114	4.28	2.45
20–30 cm	110	4.29	2.37

This behaviour has a direct physical interpretation. A poorly insulated building loses a significant amount of heat through the envelope, and the solar gains through the glazing represent a significant compensating source in the thermal balance; suppressing these gains through shading has a high absolute impact. As the insulation increases, the share of the losses through the envelope decreases, and the relative contribution of the solar gains becomes, paradoxically, less critical for covering the heating season — some buildings even enter a slightly overheated regime, which shading attenuates.

From a design perspective, this observation has an important practical consequence: **the benefit of envelope renovation is amplified by the presence of shading**. In other words, a building located in a dense urban context, with neighbours to the south, achieves a larger reduction of the heating demand as a result of envelope insulation than the same building placed in an open field. This confirms the need to include the shading effect in the feasibility studies of energy renovations, especially in dense urban contexts.

6.4.1 Implications for ML-based modelling

For a predictive model, this non-linear interaction is particularly relevant: the relationship between the geometric inputs (azimuth, distance) and the output (heating demand) cannot be described as a sum of independent contributions, but requires the ability to capture terms of the type $E_3 \times \mathbb{1}_{\text{shading}}$. This fact explains, retrospectively, why the explicit inclusion of the shading descriptors leads not only to an improvement of the mean accuracy (Table 6.1), but also to a reduction of the prediction dispersion, especially for the sub-population of poorly insulated buildings exposed to southern obstructions — the most frequently encountered case in the renovation of the existing urban building stock.

CHAPTER 7
DISCUSSIONS

7.1 The influence of shading on the learning and prediction process

The results clearly demonstrate that solar shading has a significant influence on the learning process and on the predictive accuracy of neural-network-based models for estimating the energy demand of buildings, a fact evidenced by the substantial reduction of the prediction error obtained by introducing the shading-related inputs (Table 6.1). When the shading information is not explicitly included, the baseline model is forced to implicitly infer the solar masking effects from envelope and operational variables that are not directly related to the geometry of the surroundings, since the input space contains no geometric descriptors of the neighbouring obstructions. This limitation leads to an increase in the dispersion of the predictions, especially for the configurations characterised by strong or asymmetric shading, as can be observed in the comparison between the predicted and simulated values for the baseline model (Figure 6.2).

In contrast, the model that includes shading information benefits from an input space that is better grounded physically. Through the explicit encoding of the geometric shading descriptors, such as the azimuth of the obstruction and the separation distance, the model is able to directly associate the variations in solar gains with the changes in heating demand. This leads to a substantial reduction of the prediction error (Table 6.1) and to a more stable regression behaviour across the different shading configurations, as illustrated by the tighter alignment with the identity line (Figure 6.3) and by the more uniform distribution of the errors across the shading scenarios (Figure 6.4). The significant improvement of the RMSE and CV(RMSE) values confirms that shading is not a secondary effect, but a determining factor of the model's performance in predicting the energy demand in dense or obstructed environments.

7.2 Strengths and limitations of the proposed approach

One of the main strengths of the proposed approach lies in the controlled construction of the simulation dataset. By reusing identical building configurations for all the shading scenarios, the study isolates the impact of solar shading from the influence of the other envelope and operational parameters. This design ensures that the differences observed in the predictive performance can be attributed directly to the shading effects, and not to spurious variability in the input space.

Another important strength is the integration of physically meaningful descriptors into a data-driven modelling framework. The inclusion of the geometric shading parameters allows the model to learn relationships that are coherent with building physics, a fact reflected by the improvement in prediction accuracy and by the reduction in dispersion observed for the model that includes shading information (Figures 6.3 and 6.2). The SHAP analysis further supports this interpretation, highlighting the contribution of the shading-related features alongside the envelope and glazing parameters (Figure 6.5).

Despite these advantages, certain limitations must also be acknowledged. The shading configurations analysed in this study are deliberately simplified, involving a single neighbouring obstruction, with a fixed height and an idealised geometry. Real urban environments typically include multiple surrounding buildings with variable shapes, heights and surface properties, which can lead to more complex shading interactions than those captured in this analysis.

In addition, the models are trained exclusively on simulation-generated data. While this ensures coherence and physical consistency, it may limit the direct transferability to real buildings, where the uncertainties associated with workmanship quality, occupant behaviour and the variability of weather conditions are more pronounced. Finally, although the model that includes shading information demonstrates high predictive performance within the analysed parameter space, extrapolation beyond the training ranges cannot be guaranteed and must be approached with caution.

7.3 Applicability in early design stages and renovation analysis

The proposed modelling framework is particularly relevant for early design stages and for renovation analysis, in which the rapid evaluation of a large number of configurations is required, while detailed information about the building components may still be limited. In these phases, full dynamic thermal simulations remain time-consuming and computationally expensive, which makes them less suitable for the iterative exploration of design alternatives.

Once trained, the neural-network-based models provide almost instantaneous predictions of the annual heating demand, while preserving the sensitivity to the main physical determinants, including the building envelope properties and the solar shading effects. The high agreement between the predicted and simulated values observed for the model including shading information (Figure 6.3) indicates that this approach can reliably approximate the simulation results at a much reduced computational cost.

The renovation scenario further illustrates the applicability of the method in decision-support contexts. Despite the significant changes in the insulation and glazing parameters (Table 5.1), the model that includes shading information predicts a physically consistent evolution of the heating demand, indicating its reduction following the renovation (Table 6.2). This behaviour suggests that the model captures generalisable relationships between the input variables and the energy demand, rather than being overfitted on a narrow range of building configurations.

Consequently, the proposed approach can be used effectively as a screening or preliminary assessment tool for comparing design and renovation options, identifying the relative importance of the shading effects and prioritising the configurations that require more detailed simulation-based analyses. Such a workflow is particularly valuable in dense urban contexts, where solar masking plays an essential role in the energy performance of buildings.

7.4 Sources of error and possible improvements

This study demonstrates that the explicit inclusion of solar shading information in data-driven building energy estimation models leads to a significant improvement in predictive accuracy. Compared to a baseline neural network trained without shading-related inputs, the model that includes shading information achieves a substantial reduction of the prediction error, with a decrease of the RMSE value of approximately 72% and a CV(RMSE) below 0.3%.

The results also show that neglecting shading causes the baseline models to implicitly infer the solar masking effects from envelope and operational variables that are not directly related to the geometry of the surroundings, which leads to an increased dispersion of the predictions for the configurations characterised by pronounced shading. In contrast, the inclusion of the geometric shading descriptors allows the model to directly capture the variations in solar gains induced by the neighbouring obstructions, resulting in a more coherent predictive behaviour across the different shading scenarios.

Finally, the analysis under renovation conditions indicates that the model including shading information predicts a physically consistent evolution of the demand when the building envelope and glazing properties are modified. This result suggests that the proposed approach learns generalisable relationships between the physical inputs and the energy demand, supporting its potential use as a fast and reliable decision-support tool in early design and renovation stages.

The results of this study highlight the central role of solar shading as a structuring variable in the predictive modelling of energy consumption. Instead of acting as a secondary correction, shading directly influences the solar gains and therefore constitutes a determining factor of the heating demand, especially in dense or obstructed urban environments.

From a modelling perspective, the explicit inclusion of the shading-related descriptors allows data-driven approaches to more faithfully reflect the fundamental physical mechanisms that govern the energy performance of buildings. When shading is not represented in the input space, the predictive models are forced to indirectly approximate its effects through the envelope or operational parameters, which can reduce robustness and increase the dispersion of the predictions. By contrast, input spaces that include shading information allow a clearer correlation between geometry, solar exposure and energy demand.

These findings suggest that future energy prediction models, especially those intended for urban-scale applications or for supporting early-stage design, should systematically integrate shading information alongside the conventional envelope and system parameters. Such an integration improves not only the predictive accuracy, but also the interpretability of the model and its physical consistency, strengthening the relevance of hybrid approaches that combine the geometric context with machine learning methods.

Several directions can be explored to extend the scope and relevance of the proposed approach. First, the representation of shading could be refined by considering more complex urban configurations, including multiple neighbouring obstructions with variable heights, shapes and spatial arrangements. Such extensions would allow the model to better capture the diversity and geometric complexity of real urban environments, in line with recent developments in the field of urban building energy modelling (UBEM) [3, 4].

Future work could also investigate the generalisation capability of the proposed models under different climatic conditions and for various building typologies. Training and evaluating the modelling framework on datasets generated for several climates, or validating the predictions through comparison with measured building energy consumption, would provide valuable insights into the robustness and transferability of the model beyond the controlled context of the simulations.

From a methodological perspective, the integration of hybrid modelling strategies represents a promising direction. In particular, combining data-driven approaches with simplified physical constraints or with Physics-Informed Neural Networks (PINNs) could improve the extrapolation capability while maintaining computational efficiency [10, 11]. Finally, extending the analysis framework to include the cooling demand and multi-objective performance indicators would further increase its relevance for decision support in early design and renovation stages.

CHAPTER 8
COMPARATIVE STUDY ON CONCRETE CASES

The statistical analyses presented in the previous chapters characterised the impact of shading on the entire population of 2000 configurations. To make these results more accessible and easier to interpret, this chapter selects four concrete cases from the simulation set and analyses the behaviour of each in detail: the variation of the heating demand as a function of the insulation level, of the orientation of the obstruction and of its distance. All the values presented are extracted directly from the thermal simulations performed with the COMFIE engine.

8.1 Selection of the representative cases

The four buildings were chosen so as to cover a wide spectrum of envelope insulation levels, from completely uninsulated buildings to well-insulated modern ones. Table 8.1 presents the main parameters of each case, where E_2 and E_3 are the interior and exterior wall insulation thicknesses, and E_{11} is the thermal transmittance of the glazing.

Table 8.1: The characteristics of the four analysed cases (actual values from the simulations).

Case	Description	E_2 [cm]	E_3 [cm]	E_{11}	Envelope type
A	Uninsulated, high demand	0.0	0.0	6.14	poor
B	Uninsulated, poor glazing	0.0	0.0	3.27	poor
C	Modern, well insulated	12.3	29.1	2.83	high-performance
D	Medium insulation	0.0	5.6	1.84	medium

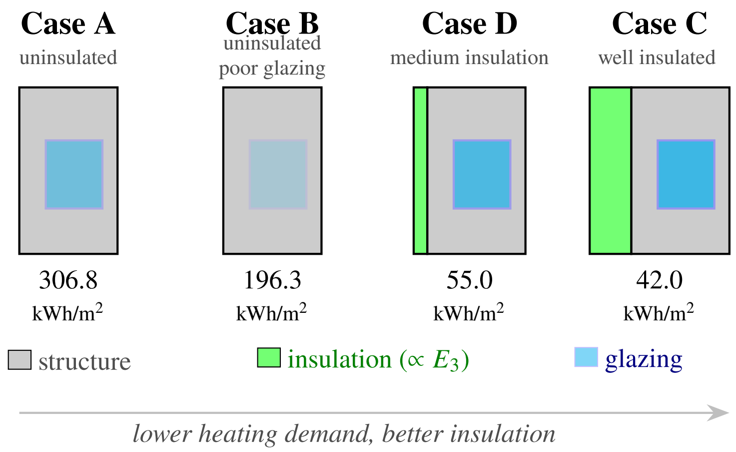


Fig. 8.1: Schematic representation of the four analysed cases, ordered by decreasing heating demand. The thickness of the green layer is proportional to the exterior insulation (E_3): cases A and B are uninsulated (high demand), case D has medium insulation, and case C is well insulated (the lowest demand).

8.2 Comparison of the heating demand

Table 8.2 presents the annual heating demand for each case, with and without shading, for the scenario with the most severe impact (obstruction placed strictly to the south, at a distance of 5 m).

Table 8.2: The annual heating demand and the shading impact for the four cases (South scenario at 5 m).

Case	Without shading [kWh/m^2]	With shading [kWh/m^2]	Impact Δ [kWh/m^2]	Relative impact [%]
A	306.8	313.0	6.20	2.0
B	196.3	207.6	11.33	5.8
C	42.0	48.7	6.73	16.0
D	55.0	61.0	5.97	10.9

Figure 8.2 illustrates these values graphically, highlighting both the baseline demand and the increase induced by shading for each case.

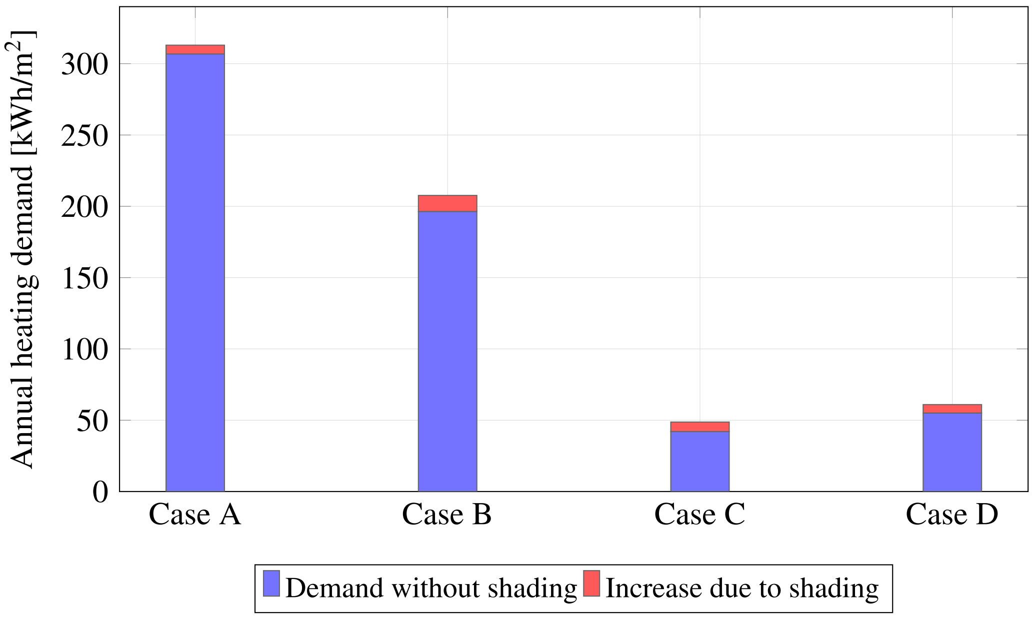


Fig. 8.2: The annual heating demand for the four analysed cases. The red segment represents the increase induced by an obstruction placed to the south, at 5 m.

The comparison highlights an important observation: although in absolute terms the increase due to shading is on the order of 6–11 kWh/m^2 for all cases, its relative share differs dramatically. For the uninsulated building in case A, the increase of 6.20 kWh/m^2 represents only 2.0% of the demand. In contrast, for the well-insulated building in case C, a similar increase in absolute terms (6.73 kWh/m^2) represents 16% of the total demand — eight times more in relative terms.

8.3 The variation of the impact with the orientation of the obstruction

To highlight the dependence of the impact on the orientation of the obstruction, Table 8.3 presents the increase in heating demand for each of the five analysed orientations, at the fixed distance of 5 m.

Table 8.3: The shading impact Δ [kWh/m^2] as a function of the orientation of the obstruction (at 5 m), for the four cases.

Case	South	South-East	South-West	East	West
A	6.20	1.34	1.35	0.93	0.90
B	11.33	2.30	2.33	1.78	1.74
C	6.73	1.22	1.22	0.99	0.97
D	5.97	1.14	1.15	0.87	0.86

Figure 8.3 represents these values graphically. The pattern is consistent across all four cases: the southern orientation categorically dominates, with an impact approximately 6 times larger than the lateral orientations. The nearly perfect symmetry between the East and West orientations, and between South-East and South-West respectively, is also notable, in agreement with the symmetric geometry of the building about the North–South axis.

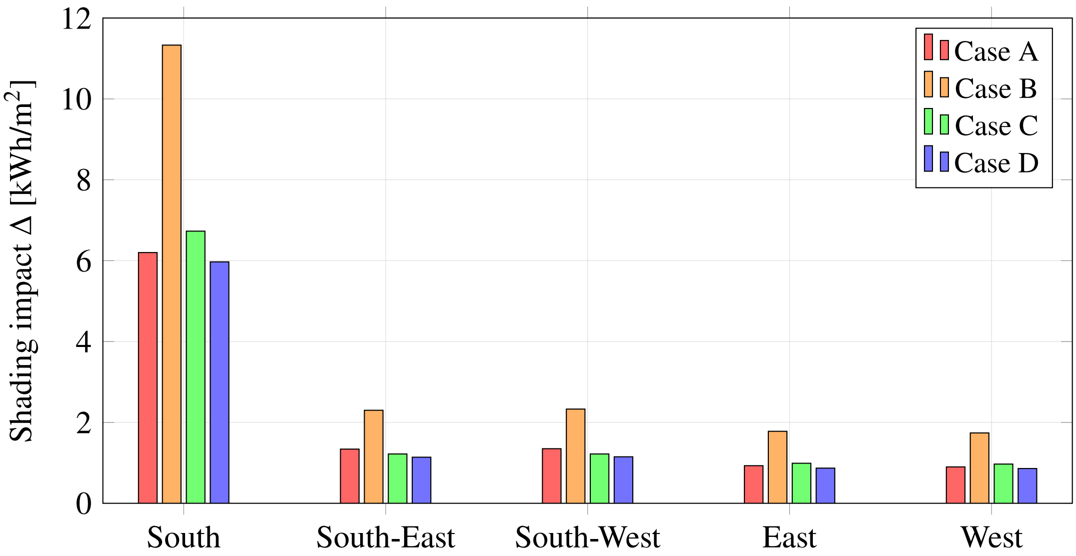


Fig. 8.3: The shading impact as a function of the orientation of the obstruction (at 5 m). For all cases, the southern orientation dominates, while the lateral orientations (East, West) have a minimal impact.

8.4 The variation of the impact with distance

A second essential geometric parameter is the distance between the obstruction and the building. Table 8.4 presents the shading impact in the southern direction for the three analysed distances.

Table 8.4: The shading impact Δ [kWh/m^2] in the southern direction, as a function of the distance of the obstruction.

Case	South 5 m	South 10 m	South 15 m
A	6.20	3.46	1.82
B	11.33	6.43	3.46
C	6.73	3.86	2.10
D	5.97	3.44	1.86

A monotonic decrease of the impact with increasing distance can be observed: for all cases, doubling the distance from 5 to 10 m reduces the impact by approximately 45%, and at 15 m the impact drops to approximately 30% of the initial value. This attenuation reflects the fact that, as the obstruction moves further away, the angle under which it blocks the direct solar radiation decreases, and the southern façade recovers an increasingly large part of the solar gains.

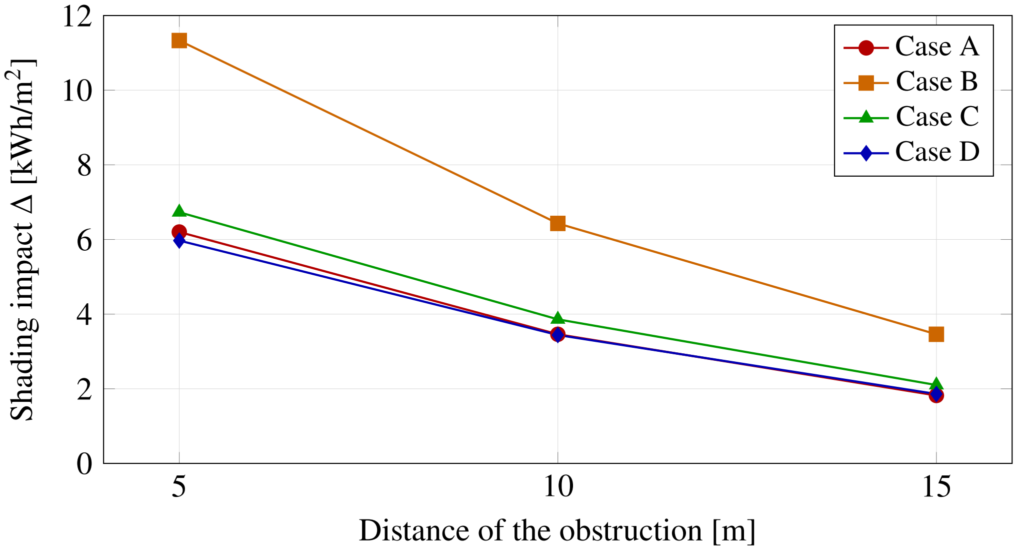


Fig. 8.4: The variation of the shading impact with the distance of the obstruction (southern direction), for the four cases. The impact decreases monotonically with distance: doubling the distance from 5 to 10 m reduces the impact by approximately 45%.

8.5 The general trend: absolute vs. relative impact

The four individual cases suggest a trend that can be confirmed at the level of the entire population of 2000 configurations. Figure 8.5 presents, for the South scenario at 5 m, both the mean absolute impact and the mean relative impact, grouped by exterior insulation level.

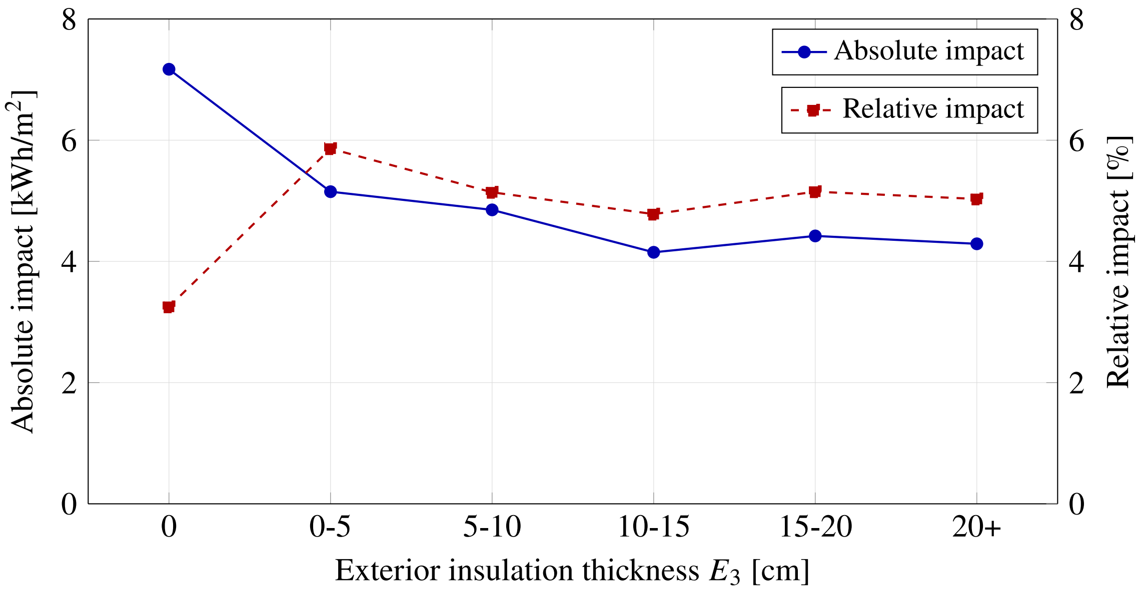


Fig. 8.5: The absolute impact (blue, left axis) and the relative impact (red, right axis) of shading as a function of the insulation level, computed over the 2000 configurations (South scenario at 5 m). As the insulation increases, the absolute impact decreases, while the relative impact increases.

The figure confirms the opposite trends of the two indicators: the *absolute* impact decreases from 7.17 kWh/m^2 (uninsulated buildings) to approximately 4.3 kWh/m^2 (well-insulated buildings), while the *relative* impact increases from 3.25% to over 5%. In other words, although an insulated building loses less energy in absolute terms because of shading, this loss represents a more important fraction of its total demand, which is already reduced.

8.6 Interpretation and implications

The comparative study on concrete cases leads to three main conclusions.

Geometry dominates. Both the orientation and the distance of the obstruction are determining factors of the impact. A southern obstruction at 5 m can have an impact up to 10 times larger than the same obstruction placed to the east or west, or more than 3 times larger than the same southern obstruction at 15 m. This geometric sensitivity justifies the explicit inclusion of the position descriptors (azimuth, distance) in the prediction models.

Insulation changes the nature of the problem. For uninsulated buildings, shading is a secondary factor (2–3% of the demand). For modern, well-insulated buildings, shading becomes a proportionally major factor (10–16% of the demand). Case B illustrates an interesting particular case: an uninsulated building with poor-quality glazing, in which the pronounced dependence on the direct solar gains through the windows leads to the largest absolute impact among all the cases (11.33 kWh/m^2).

The relevance increases over time. As energy efficiency standards become stricter and the building stock is modernised, the relative importance of the shading context in the energy balance will increase. This strengthens the long-term relevance of the approach proposed in the present work, which explicitly integrates the shading effect into the prediction of energy consumption.

CHAPTER 9
ECONOMIC ANALYSIS

The results presented in the previous chapters quantify the influence of solar shading and of the envelope insulation on the heating demand. In order to assess the practical relevance of these results, this chapter translates the energy savings into financial terms, estimating the cost of the investment in envelope insulation, the payback period and the net present value of the investment. The analysis is carried out for the conditions specific to Romania, using energy prices valid for the period 2025–2026 and the actual values extracted from the author's own simulations.

9.1 Assumptions and input data

9.1.1 The geometry of the building

The cost of thermal insulation is related to the façade area, not to the usable floor area. Starting from the dimensions of the reference building (a footprint of 20×10 m and a height $H = 11.5$ m), the façade area is estimated as follows:

$$S_{\text{façade}} = P \cdot H = 2\,(20 + 10) \cdot 11.5 \approx 690 \text{ m}^2 \tag{9.1}$$

where P is the perimeter of the building. Considering an average window-to-wall ratio of approximately 28%, the opaque area that can effectively be insulated is approximately 500 m^2.

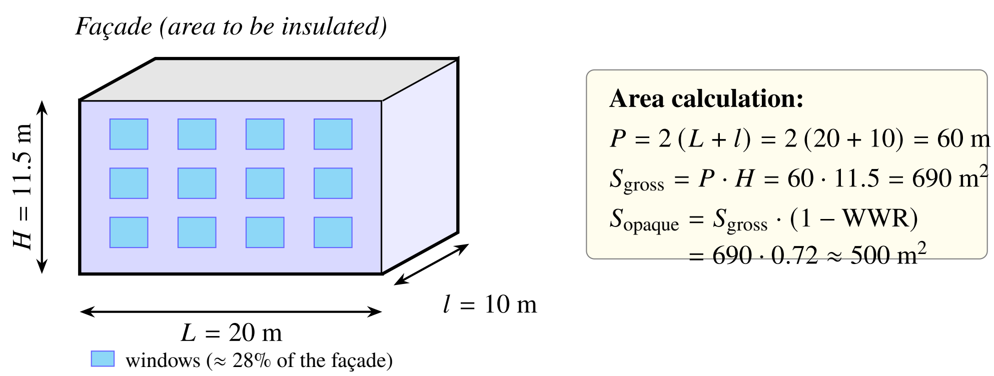


Fig. 9.1: The geometry of the reference building and the calculation of the area to be insulated. The gross façade area results from the perimeter and the height, and the opaque area that can effectively be insulated is obtained by subtracting the glazed area (window-to-wall ratio ≈ 28%).

Splaiul Independenţei 313, Sector 6, 060042 Bucharest, Romania
Tel: +40 21 402 9433; Fax: +40 21 318 1015; http://www.energ.upb.ro/

9.1.2 Energy prices

To cover the most common situations encountered in Romania, the analysis considers two heating sources: natural gas (the most widespread heating fuel in urban areas) and electricity. The prices used, presented in Table 9.1, reflect the regulated and market values for the period 2025–2026.

Table 9.1: The energy prices used in the analysis (Romania, 2025–2026).

Energy source	Price [lei/kWh]	Remarks
Natural gas	0.31	capped price, VAT included
Electricity	1.30	average post-liberalisation price

For gas heating, a boiler efficiency of 0.92 was considered, while for electric resistance heating the efficiency was taken as unity. The conversion into euros was carried out at an indicative exchange rate of 5.0 lei/euro.

9.2 Annual savings at the level of the building population

The heating demand values used in this analysis are extracted directly from the author's own simulations, as averages over the building sub-populations from Chapter 8. The mean demand of the uninsulated buildings ($E_3 = 0$) is 239.0 kWh/m^2, while for the well-insulated buildings ($E_3 > 15$ cm) it drops to 105.2 kWh/m^2 — a reduction of approximately 56%. The absolute reduction of the demand through envelope insulation is thus:

$$\Delta Q = (239.0 - 105.2)\ \text{kWh/m}^2 \times 754.8\ \text{m}^2 \approx 101\,000\ \text{kWh/year} \tag{9.2}$$

The financial value of these savings depends on the heating source, as shown in Table 9.2.

Table 9.2: The annual savings resulting from envelope insulation, for the two heating sources.

Energy source	Savings [lei/year]	Savings [€/year]
Natural gas	~34 000	~6 810
Electricity	~131 300	~26 260

The significant difference between the two values reflects the ratio between the price of electricity and that of natural gas, amplified by the conversion efficiency. This highlights the fact that the profitability of insulation is considerably higher for electrically heated buildings.

9.3 Economic analysis on individual cases

The population average provides an overall picture, but the savings potential varies considerably from one building to another, depending on the initial condition of the envelope. Table 9.3 presents the annual savings for the four cases from Chapter 8, assuming a renovation that brings each building to the level of a well-insulated envelope (a target demand of approximately 105 kWh/m^2).

Table 9.3: The annual savings potential through renovation, for the four cases (renovation to ~105 kWh/m^2).

Case	Initial demand [kWh/m^2]	Reduction [kWh/m^2]	Gas savings [€/year]	Electricity savings [€/year]
A	306.8	201.6	~10 260	~39 560
B	196.3	91.1	~4 630	~17 880
C	42.0	0	—	—
D	55.0	0	—	—

The results highlight an important practical conclusion: **not all buildings benefit equally from renovation**. The uninsulated buildings (cases A and B) present a major savings potential, while the buildings that are already well insulated (cases C and D) do not justify an additional investment in insulation, since their demand is already low. This observation confirms the importance of assessing each building individually before deciding on an intervention — exactly the type of rapid analysis enabled by the prediction model developed in the present work.

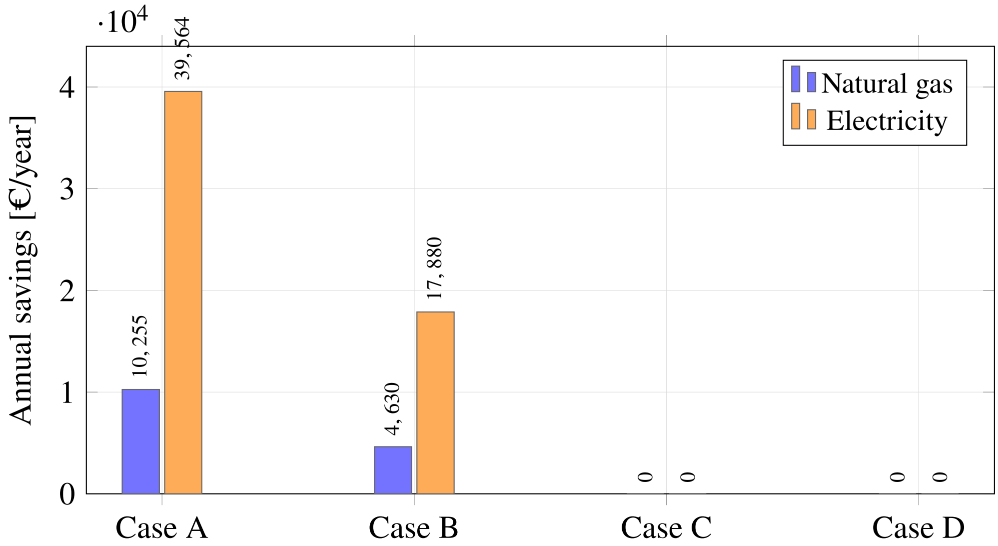


Fig. 9.2: The potential annual savings through renovation for the four cases. The uninsulated buildings (A, B) present a high potential, while the buildings that are already insulated (C, D) do not justify an additional investment.

9.4 The cost of the investment and the payback period

The cost of exterior thermal insulation in Romania, including material and labour, usually varies between 60 and 120 €/m^2 of façade. For case A (the building with the highest savings potential) and the opaque area of 500 m^2, Table 9.4 presents the total investment and the payback period.

Table 9.4: The cost of the investment and the payback period for case A, as a function of the unit cost and of the heating source.

Unit cost [€/m^2]	Total investment [€]	Payback period [years]	
		Gas	Electricity
60	~29 800	2.9	0.8
90	~44 700	4.4	1.1
120	~59 600	5.8	1.5

For case A, owing to the very high initial demand, the investment pays for itself quickly: in 3–6 years for gas heating and in less than 1.5 years for electric heating.

9.4.1 Comparison of the insulation materials

The choice of the insulation material influences both the cost and the performance. Table 9.5 compares two of the most widely used materials in Romania, for case A and gas heating.

Table 9.5: Comparison of insulation materials (case A, gas heating).

Material	Cost [€/m^2]	Payback [years]
Expanded polystyrene (EPS), 16 cm	~75	3.6
Basalt mineral wool, 16 cm	~105	5.1

Expanded polystyrene offers a faster payback owing to its lower cost, while basalt mineral wool, although more expensive, presents additional advantages in terms of fire resistance and acoustic performance, which are not quantified in this purely energy-focused analysis.

9.5 Sensitivity analysis to the energy price

Energy prices are subject to significant fluctuations, especially in the current context of energy market volatility. To assess the robustness of the conclusions, Table 9.6 presents the payback period of the investment (case A, cost 90 €/m^2, gas heating) for different energy price increase scenarios.

Table 9.6: The sensitivity of the payback period to the energy price (case A, gas, 90 €/m^2).

Price scenario	Savings [€/year]	Payback [years]
Current price	~10 260	4.4
Increase +20%	~12 310	3.6
Increase +50%	~15 380	2.9
Increase +100%	~20 510	2.2

The analysis confirms that the profitability of the investment increases with the energy price: a doubling of the gas price would reduce the payback period from 4.4 to 2.2 years. This means that, in a context of rising energy prices, the investment in insulation becomes even more attractive, while also offering protection against the future volatility of costs.

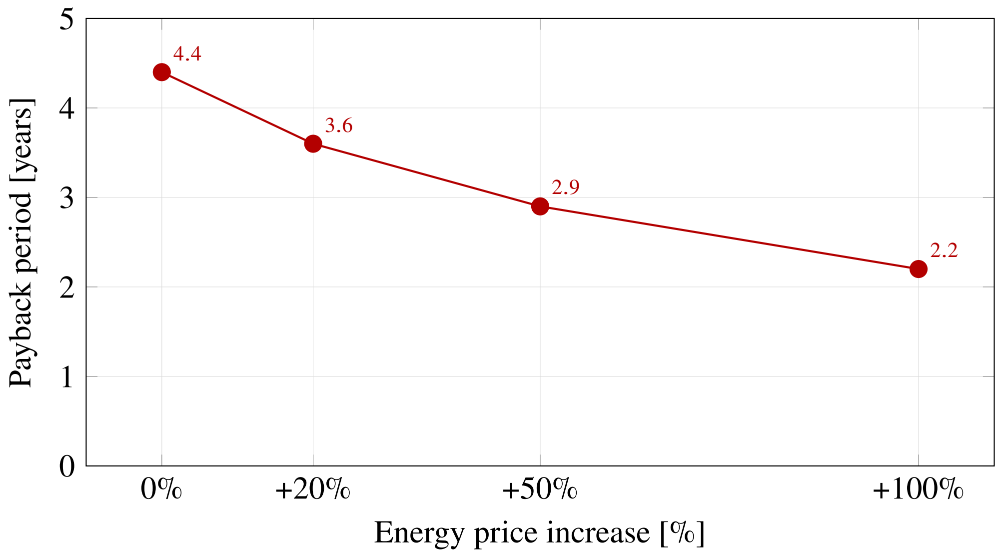


Fig. 9.3: The sensitivity of the payback period to the energy price increase (case A, gas heating, cost 90 €/m^2). A price increase significantly shortens the payback period of the investment.

9.6 The net present value of the investment

The payback period offers a simple measure of profitability, but it does not account for the time value of money or for the total lifetime of the investment. A more complete indicator is the Net Present Value (NPV), which discounts the future savings to their present value:

$$\mathrm{NPV} = -I_0 + \sum_{t=1}^{N} \frac{E_t}{(1+r)^t} \tag{9.3}$$

where I_0 is the initial investment, E_t the savings in year t, r the discount rate, and N the analysis horizon. Considering a horizon of 20 years (the typical lifetime of an insulation system), a discount rate of 5% and the reference investment (case A, 90 €/m^2, $I_0 \approx 44\,700$ €), the results are presented in Table 9.7.

Table 9.7: The Net Present Value (NPV) over 20 years for case A ($r = 5\%$).

Scenario	NPV gas [€]	NPV electricity [€]
No energy price increase	~83 000	~448 000
Price increase 3%/year	~124 000	~606 000

The positive and substantial NPV values confirm that the investment in insulation is profitable over its entire lifetime, generating a net value several times larger than the initial investment. In the case of electric heating, the NPV is approximately ten times larger than the investment, which makes thermal insulation one of the most profitable interventions for an uninsulated, electrically heated building.

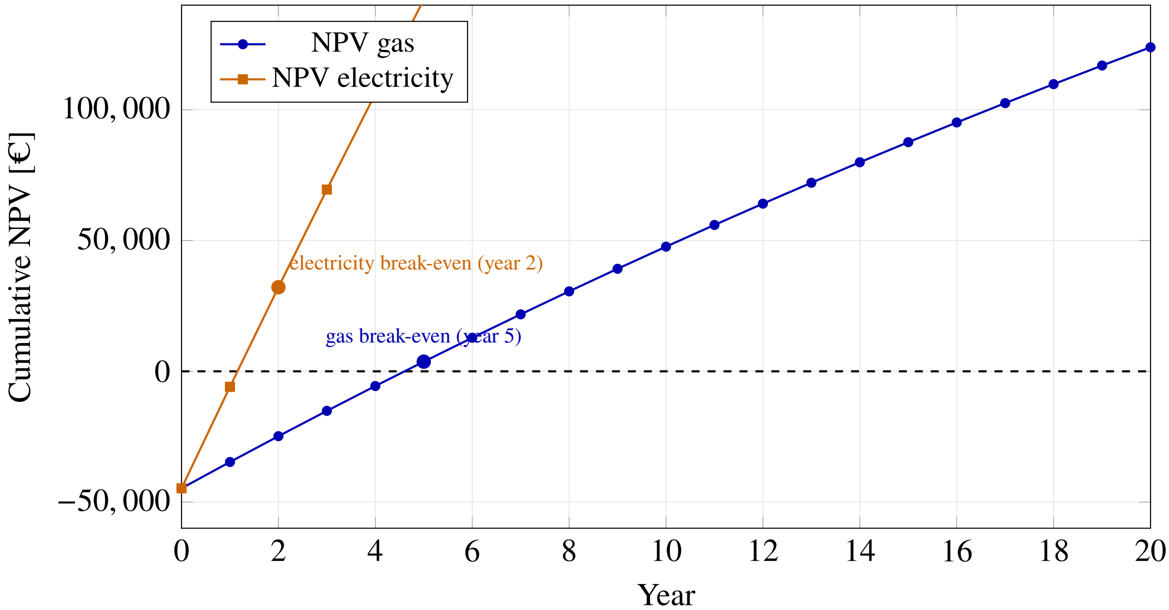


Fig. 9.4: The evolution of the cumulative Net Present Value (NPV) over 20 years for case A (cost 90 €/m^2, energy price increase 3%/year). The investment is recovered (NPV = 0) in year 5 for gas heating and in year 2 for electric heating, after which it generates positive net value.

9.7 The economic value of the shading effect

The central result of the present work — the amplification of the insulation benefit in the presence of shading (Section 6.4) — also has a direct economic dimension. The analysis showed that the impact of shading decreases from 7.17 kWh/m^2 (uninsulated building) to 4.29 kWh/m^2 (insulated building), a difference of 2.88 kWh/m^2.

This difference represents an additional energy benefit of insulation, attributable exclusively to the interaction with shading:

$$\Delta Q_{\text{shading}} = 2.88\ \text{kWh/m}^2 \times 754.8\ \text{m}^2 \approx 2\,170\ \text{kWh/year} \tag{9.4}$$

In financial terms, this additional benefit translates into approximately 146 €/year for gas heating and approximately 565 €/year for electric heating. Although at the level of a single building the value may seem modest, it becomes significant when extrapolated to the scale of the urban building stock, especially in dense neighbourhoods where mutual shading is frequent.

This observation has an important practical consequence for the assessment of energy renovation projects: **ignoring the shading effect leads to an underestimation of the profitability of insulation** for buildings located in a dense urban context, with obstructions in the southern direction. The explicit inclusion of shading in feasibility analyses, as enabled by the methodology developed in this work, ensures a more realistic estimation of the savings and, implicitly, of the payback period.

9.8 Summary

The economic analysis confirms that envelope insulation is a profitable investment, with a payback period ranging from less than 1 year (an uninsulated, electrically heated building) to approximately 6 years (gas heating, maximum cost), for the buildings with high savings potential.

The analysis on individual cases also showed that not all buildings justify an intervention, as the buildings that are already well insulated have a low savings potential.

The advanced financial indicators — the sensitivity analysis and the net present value — confirm the robustness of these conclusions: the investment remains profitable even under conservative scenarios and becomes increasingly attractive in the context of rising energy prices. Moreover, the quantification of the economic value of the shading effect demonstrates that including the geometric context in energy analyses has concrete financial implications for renovation decisions in dense urban environments. These findings strengthen the practical usefulness of the machine-learning-based approach proposed in the present work, as a rapid tool for evaluating renovation options.

CHAPTER 10
CONCLUSIONS

The present work investigated the influence of solar shading on the energy consumption of buildings, using artificial neural networks trained on data generated through dynamic thermal simulation. This chapter summarises the main results and contributions, highlights the limitations of the approach and proposes directions for future research.

10.1 Summary of the results

The central objective of the work was to quantify the extent to which the explicit inclusion of the shading parameters improves the accuracy of machine-learning-based predictive models. The results obtained allow the following main conclusions to be drawn.

Including shading substantially improves prediction. The comparison of the two MLP models — one without and one with explicit shading descriptors — revealed a reduction of approximately 72% in the prediction error (RMSE from 2.10 to 0.60), accompanied by a coefficient of variation CV(RMSE) below 0.3%. This result confirms the basic hypothesis of the work: solar shading is not a negligible secondary effect, but a determining factor of energy performance that must be explicitly represented in prediction models.

Explicitly providing the shading simplifies the model. A remarkable result is that the model receiving the shading information achieves superior accuracy with a much simpler architecture (a single hidden layer with 20 neurons), compared to the model without shading, which requires a deep architecture (nine layers of 80 neurons each). This contrast demonstrates that the shading information provided directly eliminates the need for the network to implicitly infer these effects from the other variables.

The shading geometry plays a dominant role. The detailed analysis over the 15 scenarios showed that the impact of shading depends strongly on the geometry of the obstruction. The southern orientation categorically dominates, with an impact up to 27 times larger than the lateral orientations (East, West), and the impact decreases monotonically with the distance of the obstruction. This geometric sensitivity, explained by the solar path at the latitude of Bucharest, justifies the inclusion of the position descriptors (azimuth and distance) in the input space of the model.

The insulation–shading interaction is significant. The work highlighted a non-linear interaction between the envelope insulation level and the impact of shading. In absolute terms, the impact of shading decreases by approximately 40% as the insulation increases (from 7.17 to 4.29 kWh/m^2). In relative terms, however, the impact increases significantly for well-insulated buildings, coming to represent up to 16% of the total demand. This observation has an important practical consequence: as the building stock is modernised, the relative role of shading in the energy balance becomes increasingly important.

The model maintains its robustness under renovation. The robustness test through the renovation scenario showed that the model with shading preserves its accuracy even when the envelope characteristics are significantly modified, which suggests that the model learns generalisable physical relationships, rather than merely memorising the training configurations.

Concrete economic relevance. The economic analysis demonstrated that the results have direct financial implications. Envelope insulation pays for itself in 1–6 years, depending on the heating source, and the net present value over 20 years is positive and substantial. The quantification of the economic value of the shading effect showed that ignoring it leads to an underestimation of the profitability of renovations in dense urban contexts.

10.2 Contributions

The main contributions of the present work can be summarised as follows:

- **The controlled isolation of the shading effect**, by reusing the same 2000 envelope configurations for all the shading scenarios, which allows the rigorous attribution of the performance variations exclusively to the geometric effect of the obstruction;
- **The systematic quantification of the improvement** brought by the explicit inclusion of the shading descriptors, through the direct comparison of two MLP models on a large dataset (32 000 simulations);
- **The detailed analysis of the geometric dependence** of the shading impact as a function of azimuth and distance, providing an in-depth physical perspective on the mechanisms involved;
- **Highlighting the insulation–shading interaction**, a result with direct relevance for the feasibility studies of energy renovations;
- **The translation of the results into economic terms**, by estimating the payback period and the net present value for the conditions specific to Romania.

10.3 Limitations

The limitations of the approach, discussed at length in the Discussions chapter, are summarised here. The shading configurations are deliberately simplified (a single obstruction, fixed height), the models are trained exclusively on simulation data, and the study addresses exclusively the heating demand, without the cooling component. In addition, the economic analysis is based on prices valid at the time of writing and is indicative in nature.

10.4 Directions for future research

Starting from the results and the limitations identified, several promising directions for future research emerge.

A first direction is the extension of the shading representation towards realistic urban configurations, with multiple obstructions of varied geometries, in order to capture the complexity of real built environments. Closely related, training and validating the models on datasets corresponding to several climate zones and building typologies would allow the assessment of the generalisation capability of the approach.

A second direction concerns the extension of the analysis framework to include the cooling demand and multi-objective performance indicators, which would provide a complete picture of the annual energy balance, especially in the context of global warming and of the increasing cooling demand.

From a methodological perspective, the integration of physical constraints into the learning process, through approaches such as Physics-Informed Neural Networks (PINNs), represents a valuable direction for improving the extrapolation capability of the models beyond the training domain.

Finally, the experimental validation of the predictions through comparison with measured energy consumption of real buildings would constitute an essential step towards the practical application of the developed methodology, strengthening the bridge between simulation and the built reality.

10.5 Final remarks

The present work has demonstrated that the explicit inclusion of solar shading in machine-learning-based prediction models leads to significant improvements in accuracy, simplifies the required model architecture and provides a more faithful representation of the physical mechanisms that govern the energy performance of buildings. Beyond the methodological contribution, the results have direct practical relevance for the design and renovation of buildings in dense urban environments, where mutual shading plays an essential role. As the energy efficiency of buildings becomes an increasingly pressing priority, the integration of the geometric context into rapid prediction tools represents a useful contribution to the broader effort of reducing energy consumption in the building sector.

APPENDIX A
THE COMPLETE CATALOGUE OF THE SHADING SCENARIOS

This appendix presents the complete values of the shading impact for all 15 analysed scenarios, computed as averages over the 2000 envelope configurations. The impact $\bar{\Delta}$ represents the mean increase of the annual heating demand, and Δ_{max} the maximum value recorded over the population of configurations.

Table A.1: The shading impact for the 15 scenarios (averaged over 2000 configurations).

Orientation	Azimuth	Distance [m]	$\bar{\Delta}$ [kWh/m^2]	Δ_{max} [kWh/m^2]
South	0°	5	6.73	11.46
South	0°	10	3.80	6.50
South	0°	15	2.02	3.47
South-East	−45°	5	1.40	2.34
South-East	−45°	10	1.02	1.71
South-East	−45°	15	0.76	1.28
South-West	+45°	5	1.42	2.37
South-West	+45°	10	1.00	1.68
South-West	+45°	15	0.73	1.23
East	−90°	5	1.03	1.84
East	−90°	10	0.48	0.87
East	−90°	15	0.25	0.45
West	+90°	5	1.01	1.80
West	+90°	10	0.49	0.89
West	+90°	15	0.27	0.50

The nearly perfect symmetry between the East–West and South-East–South-West pairs is notable, a consequence of the geometric symmetry of the building about the North–South axis. The ratio between the most severe scenario (South, 5 m) and the least severe one (East, 15 m) is approximately 27, illustrating the categorical dominance of the southern orientation during the heating season.

APPENDIX B
THE INPUT VARIABLES OF THE MODEL

Table B.1 describes the eleven physical and operational variables sampled for the generation of the database, together with their variation ranges.

Table B.1: The description of the model input variables.

Code	Description	Unit
E2	Interior insulation thickness (exterior walls)	cm
E3	Exterior insulation thickness (exterior walls)	cm
E6	Exterior insulation thickness (stairwell walls)	cm
E7	Interior insulation thickness (roof)	cm
E8	Exterior insulation thickness (roof)	cm
E10	Insulation thickness (ground floor slab)	cm
E11	Thermal transmittance of the windows (North)	$W/(m^2K)$
E19	Solar factor of the glazing (North)	–
E31	Air change rate (typical floors)	h^{-1}
E32	Air change rate (ground floor/corridor)	h^{-1}
G4	Heating active on the ground floor (binary variable)	–

Note: The variable codes (E2, E3, . . . , G4) correspond to the labels used in the *Pléiades* simulation model, which includes a wider set of descriptive building parameters. The table above presents exclusively the eleven variables that were sampled (varied) for the generation of the database. The other parameters of the model — such as the geometric dimensions, the number of storeys, the window-to-wall ratios and the thermal properties of the façades with orientations other than North — were kept at fixed values for all the simulations, which is why the numbering of the codes is not continuous.

APPENDIX C
THE HYPERPARAMETER CONFIGURATION

This appendix details the hyperparameter search space and the optimal configurations selected for the two models.

C.1 The search space

The hyperparameter optimisation was carried out with `Keras Tuner`, through the `RandomSearch` strategy, exploring the space described in Table C.1.

Table C.1: The hyperparameter search space.

Hyperparameter	Range / values
Number of hidden layers	1 – 10
Number of neurons per layer	20 – 500
Activation function	SELU (fixed)
Weight initialiser	`lecun_normal`
Learning rate	$10^{-4} - 3 \times 10^{-2}$
Optimiser	Adam
Maximum number of epochs	100
Stopping strategy	early stopping

C.2 The optimal configurations

Table C.2 compares the resulting optimal configurations for the two models.

Table C.2: The optimal hyperparameter configurations for the two models.

Hyperparameter	MLP without shading	MLP with shading
Hidden layers	9	1
Neurons per layer	80	20
Learning rate	0.001	0.01
Optimiser	Adam	Adam
RMSE (test)	2.1047	0.5977
CV(RMSE)	0.96%	0.27%
R^2	0.9991	0.9999

The striking difference in complexity between the two models (9 layers versus a single one) constitutes indirect evidence of the informational value of the shading descriptors: explicitly providing this information eliminates the need for a deep architecture to infer it implicitly.

APPENDIX D
THE DATA PREPROCESSING PIPELINE

The data preprocessing was carried out through a `ColumnTransformer` that applies differentiated transformations depending on the variable type, according to Table D.1.

Table D.1: The transformations applied in the preprocessing pipeline.

Transformation type	Targeted variables	Purpose
Indicator transformation	The 8 glazing parameters and glazing ratios	Capturing abrupt façade changes
Standardisation (z-score)	Numerical variables (insulation, ventilation)	Balanced contribution in the optimisation
One-hot encoding	*Type de mur* (categorical)	No artificial ordering
Passthrough	*Chauffage RDC* (binary)	Kept unchanged

This unified approach ensures a consistent scaling of all the input variables, preventing the optimisation process from being dominated by variables with large numerical scales and contributing to the stability of the training process.

APPENDIX E
THE PYTHON IMPLEMENTATION

This appendix presents the essential code fragments that implement the methodology described in the thesis: the data preprocessing transformers, the definition of the neural network architecture and the hyperparameter search procedure. The code is shown selectively, keeping the methodologically relevant components and omitting the auxiliary data loading and visualisation routines. The implementation uses the `TensorFlow/Keras` libraries for the neural network, `Keras Tuner` for the hyperparameter optimisation and `scikit-learn` for the preprocessing.

E.1 Preprocessing transformers

The preprocessing described in Section 5.1.2 relies on two custom transformers, compatible with the `scikit-learn` interface (`BaseEstimator`, `TransformerMixin`). The first, `IndicatorTransformer`, generates binary indicators that signal the absence of windows on a façade (a glazing ratio equal to zero). The second, `ConditionalStandardScaler`, applies z-score standardisation while ignoring "placeholder" values, so that they do not distort the computed mean and standard deviation.

```python
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.base import BaseEstimator, TransformerMixin


class IndicatorTransformer(BaseEstimator, TransformerMixin):
    """Generates a binary indicator: 1 if the glazing
    ratio is 0 (facade without windows), otherwise 0."""
    def __init__(self):
        pass

    def fit(self, X, y=None):
        return self

    def transform(self, X):
        X = X.copy()
        Y = pd.DataFrame()
        for column in X.columns:
            Y[f'{column}_indicator'] = np.where(X[column] == 0, 1, 0)
        return Y

```

```python

class ConditionalStandardScaler(BaseEstimator, TransformerMixin):
    """Z-score standardisation that ignores placeholder values,
    so as not to distort the mean and standard deviation."""
    def __init__(self, placeholder_value):
        self.placeholder_value = placeholder_value
        self.scalers = {}

    def fit(self, X, y=None):
        for col in X.columns:
            if (X[col] != self.placeholder_value).all():
                self.scalers[col] = StandardScaler()
                self.scalers[col].fit(
                    X.loc[X[col] != self.placeholder_value, col]
                    .to_numpy().reshape(-1, 1))
        return self

    def transform(self, X):
        X_transformed = X.copy()
        for col in X.columns:
            if (X[col] != self.placeholder_value).all():
                mask = X_transformed[col] != self.placeholder_value
                X_transformed.loc[mask, col] = self.scalers[col].transform(
                    X_transformed.loc[mask, col]
                    .to_numpy().reshape(-1, 1)).flatten()
        return X_transformed
```

Listing E.1: The custom preprocessing transformers (`Custom_Transformers.py`).

E.2 Building the transformation pipeline

The transformers are integrated into a `ColumnTransformer` that applies the standardisation of the input variables. A separate stage standardises the target variable (the heating demand), so that the network's predictions can subsequently be brought back to the real physical scale through the inverse transformation.

```python
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler

# Pipeline for the target variable (the heating demand)
num_labels_pipeline = Pipeline([
    ('std_scaler', StandardScaler()),
])
full_pipeline_std_scaler_labels = ColumnTransformer([
    ("num", num_labels_pipeline, num_labels_attribs),
])
chauffage_labels_prepared = \
    full_pipeline_std_scaler_labels.fit_transform(chauffage_labels_df)

# Pipeline for the input variables
full_pipeline_std_scaler = ColumnTransformer([
    ("num", StandardScaler(), num_attribs),
])
chauffage_prepared = full_pipeline_std_scaler.fit_transform(chauffage)
```

Listing E.2: Building the standardisation pipeline for the input variables and for the target variable.

E.3 Defining the MLP architecture

The network architecture is defined through a `build_model` function parametrised by the `hp` object of `Keras Tuner`, which allows the automatic exploration of the hyperparameter space. The number of hidden layers, the number of neurons per layer, the learning rate and the optimiser are selected from the sets of values described in Appendix C. The activation function is SELU, with the `lecun_normal` initialiser, as substantiated in Section 3.3.

The input layer has a size equal to the number of variables: 11 for the model without shading descriptors, and 27 for the model that explicitly includes the shading information. This difference reflects the encoding strategy of the 15 shading scenarios described in Section 6.3.

```
import tensorflow as tf
from keras.models import Sequential
from keras.layers import Dense

def build_model(hp):
    n_hidden = hp.Choice("n_hidden",
                         values=[1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
    n_neurons = hp.Choice("n_neurons",
                          values=[20, 40, 60, 80, 100, 200, 500])
    learning_rate = hp.Choice("learning_rate",
                              values=[1e-4, 1e-3, 1e-2, 3e-2])
    optimizer = hp.Choice("optimizer", values=["sgd", "adam"])

    if optimizer == "sgd":
        optimizer = tf.keras.optimizers.SGD(
            learning_rate=learning_rate, momentum=0.9, nesterov=True)
    else:
        optimizer = tf.keras.optimizers.Adam(learning_rate=learning_rate)

    model = Sequential()
    # input_dim = 11 (without shading) or 27 (with shading)
    model.add(Dense(n_neurons, input_dim=11, activation="selu"))
    for _ in range(n_hidden):
        model.add(Dense(n_neurons, activation="selu",
                        kernel_initializer="lecun_normal"))
    model.add(Dense(1))
    model.compile(loss="mse", optimizer=optimizer, metrics=["mse"])
    return model
```

Listing E.3: The MLP model building function, parametrised for the hyperparameter search. The `input_dim` value is 11 for the model without shading and 27 for the one with shading.

E.4 The hyperparameter search

The hyperparameter optimisation uses the `RandomSearch` strategy of `Keras Tuner`, with the objective of minimising the error on the validation set (`val_mse`). The training of each candidate configuration uses early stopping (`EarlyStopping`) to prevent overfitting, as well as the adaptive reduction of the learning rate (`ReduceLROnPlateau`) when the validation error stagnates. Reproducibility is ensured by fixing the random seeds.

```python
import keras_tuner as kt
from sklearn.model_selection import train_test_split

# Fixing the seeds for reproducibility
seed = 42
np.random.seed(seed)
tf.random.set_seed(seed)

# Splitting into training and validation sets
X_train, X_valid, y_train, y_valid = train_test_split(
    chauffage_prepared, chauffage_labels_prepared,
    test_size=0.2, random_state=42)

# Defining the RandomSearch tuner
random_search_tuner = kt.RandomSearch(
    build_model,
    kt.Objective("val_mse", direction="min"),
    max_trials=10,
    overwrite=False,
    directory="Grid_search_50k",
    project_name="my_rnd_search",
    seed=42)

# Callbacks: early stopping and learning rate reduction
early_stopping_cb = tf.keras.callbacks.EarlyStopping(patience=10)
lr_scheduler = tf.keras.callbacks.ReduceLROnPlateau(
    factor=0.5, patience=5)

# Launching the search
random_search_tuner.search(
    X_train, y_train,
    epochs=100,
    validation_data=(X_valid, y_valid),
    callbacks=[early_stopping_cb, lr_scheduler])

# Retrieving the best model and the optimal hyperparameters
best_model = random_search_tuner.get_best_models(num_models=1)[0]
best_hp = random_search_tuner.get_best_hyperparameters(num_trials=1)[0]
print(f"Optimal hyperparameters: {best_hp.values}")
```

Listing E.4: Configuring and launching the random hyperparameter search with `Keras Tuner`.

E.5 Evaluation on the test set

Finally, the optimal model is evaluated on the independent test set. The predictions are brought back to the real physical scale through the inverse of the standardisation transformation, and the performance is quantified through the RMSE, R^2 and CV(RMSE) indicators, presented in Section 6.2.

```python
from sklearn.metrics import mean_squared_error, r2_score

# Prediction on the test set
final_predictions = best_model.predict(X_test_prepared)
```

```python

# Bringing back to the real physical scale (inverse transformation)
scaler_labels = full_pipeline_std_scaler_labels \
    .transformers_[0][1].named_steps['std_scaler']
predictions_real = scaler_labels.inverse_transform(
    final_predictions.reshape(-1, 1))
y_test_real = scaler_labels.inverse_transform(
    y_test_prepared.reshape(-1, 1))

# Computing the performance indicators
r2_real = r2_score(y_test_real, predictions_real)
rmse_real = np.sqrt(mean_squared_error(y_test_real, predictions_real))
cv_rmse = (rmse_real / np.mean(y_test_real)) * 100

print(f"R^2: {r2_real:.4f}")
print(f"RMSE: {rmse_real:.4f}")
print(f"CV(RMSE): {cv_rmse:.2f}%")
```

Listing E.5: Evaluating the model on the test set and computing the performance indicators on the real physical scale.

E.6 The computation of the economic indicators

The economic analysis presented in Chapter 9 is based on a set of constants that reflect the conditions in Romania for the period 2025–2026 and on two computation functions: the conversion of the energy savings into financial terms and the computation of the net present value. The constants are grouped by category (geometric, energy prices, insulation costs and financial parameters), ensuring full transparency of the assumptions used.

```python
# --- Geometric constants (from the reference building model) ---
A_UTILA = 754.8          # total usable floor area [m2]
AMPRENTA_L = 20.0        # footprint length [m]
AMPRENTA_l = 10.0        # footprint width [m]
INALTIME = 11.5          # building height [m]
WWR = 0.28               # window-to-wall ratio

# --- Energy prices in Romania (2025-2026) ---
PRET_GAZ = 0.31          # natural gas (capped, VAT incl.) [lei/kWh]
PRET_ELECTRIC = 1.30     # electricity (post-liberalisation) [lei/kWh]
RANDAMENT_CENTRALA = 0.92  # gas boiler efficiency [-]
                           # (electric resistance heating has efficiency =
    1)
CURS_EUR = 5.0           # indicative exchange rate [lei/EUR]

# --- Cost of exterior thermal insulation in Romania (2025-2026) ---
COST_IZOLARE_MIN = 60    # minimum cost (material + labour) [EUR/m2]
COST_IZOLARE_MED = 90    # average cost [EUR/m2]
COST_IZOLARE_MAX = 120   # maximum cost [EUR/m2]

# --- Parameters for the financial analysis (NPV) ---
RATA_ACTUALIZARE = 0.05  # discount rate [-]
ORIZONT_ANI = 20         # lifetime of the investment [years]
```

Listing E.6: The physical and economic constants used in the analysis (Romania, 2025–2026).

The first function converts the energy savings (expressed in kWh/year) into financial savings, taking into account the efficiency of the heating system. For gas heating, the boiler efficiency (0.92) increases the required primary energy consumption, while for electric resistance heating the efficiency is unity.

```python
def economie_bani(kwh, pret, randament=1.0):
    """Computes the annual savings in lei and euros.
    kwh: the energy savings [kWh/year]
    pret: the energy price [lei/kWh]
    randament: the efficiency of the heating system [-]
    """
    lei = kwh * pret / randament
    eur = lei / CURS_EUR
    return lei, eur
```

Listing E.7: The function converting the energy savings into financial savings (lei and euros).

The second function implements the computation of the net present value (NPV), according to equation (9.3) of Chapter 9. It discounts the future savings to their present value, while also allowing the inclusion of an annual energy price growth rate.

```python
def calcul_npv(investitie, economie_anuala, ani, rata, crestere_pret=0.0):
    """Computes the net present value (NPV).
    investitie: the initial investment I0 [EUR]
    economie_anuala: the savings in the first year [EUR/year]
    ani: the analysis horizon N [years]
    rata: the discount rate r [-]
    crestere_pret: the annual energy price growth rate [-]
    Formula: NPV = -I0 + sum [ E_t / (1+r)^t ]
    """
    npv = -investitie
    for t in range(1, ani + 1):
        economie_t = economie_anuala * (1 + crestere_pret)**t
        npv += economie_t / (1 + rata)**t
    return npv
```

Listing E.8: The function computing the net present value (NPV) over a multi-year horizon.

These functions, applied to the actual values extracted from the simulations, generated all the economic indicators presented in Chapter 9: the annual savings, the payback period, the sensitivity analysis to the energy price and the net present value over 20 years.

The complete code, including the routines for loading the databases, generating the plots and performing the SHAP analysis, is available in the code repository associated with the thesis.

BIBLIOGRAPHY

[1] Yue Pan, Mingya Zhu, Yifan Lv, Yujie Yang, Yiming Liang, Rui Yin, and Xiaolei Yuan. Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. *Advances in Applied Energy*, 10:100135, 2023.

[2] Liang Zhang, Jin Wen, Yanfei Li, Jianli Chen, Yunyang Ye, Yangyang Fu, and William Livingood. A review of machine learning in building load prediction. *Applied Energy*, 285:116452, 2021.

[3] Martina Ferrando et al. Urban building energy modeling (ubem) tools: A state-of-the-art review of bottom-up physics-based approaches. *Applied Energy*, 306:117787, 2021.

[4] U. Ali et al. A review of urban building energy modeling (ubem): Approaches, challenges and future directions. *Energy and Buildings*, 250:111273, 2021.

[5] Xavier Faure, Tim Johansson, and Oleksii Pasichnyi. The impact of detail, shadowing and thermal zoning levels on urban building energy modelling (ubem) on a district scale. *Energies*, 15(4):1525, 2022.

[6] Chujie Lu, Sihui Li, and Zhengjun Lu. Building energy prediction using artificial neural networks: A literature survey. *Energy and Buildings*, 262:111718, 2022.

[7] Kadir Amasyali and Nora M. El-Gohary. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81:1192–1205, 2018.

[8] Paul Westermann and Ralph Evins. Machine learning as a surrogate model for early-stage heating demand optimization. In *Proceedings of SimBuild 2024, IBPSA-USA*, 2024.

[9] Doyun Lee and Steven Jige Quan. Surrogate modeling strategy for urban building energy simulation in early-stage urban design: A case study of energy-efficient neighborhood design in seoul. In *E3S Web of Conferences*, volume 436, page 01016, 2023.

[10] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[11] George E. Karniadakis et al. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021.

[12] Amber Jiayu Su and Timur Dogan. Voxel shader – an efficient shadow calculation approach for urban building energy modeling. *Energy and Buildings*, 332:115957, 2026.

[13] Anna Laura Pisello, John E. Taylor, Xiaoqi Xu, and Franco Cotana. Inter-building effect: Simulating the impact of a network of buildings on the accuracy of building energy performance predictions. *Building and Environment*, 58:37–45, 2012.

[14] Yilong Han, John E. Taylor, and Anna Laura Pisello. Exploring mutual shading and mutual reflection inter-building effects on building energy performance. *Applied Energy*, 185:1556–1564, 2017.

[15] Danny H. W. Li and S. L. Wong. Daylighting and energy implications due to shading effects from nearby buildings. *Applied Energy*, 84(12):1199–1209, 2007.

[16] Ziling Jiang, Jian Ge, Kang Zhao, and Jianya Xiao. Evaluating the impact of shading from surrounding obstructions on buildings' net energy in different climates. *Energy and Buildings*, 2025. în curs de publicare.

[17] Dong Hyuk Yi, Sebin Choi, Deuk-Woo Kim, and Sungmin Yoon. Quantifying urban building shading effect for data-driven building energy modeling. *Building and Environment*, 2025. în curs de publicare.

[18] Bruno Peuportier and Isabelle Blanc-Sommereux. Simulation tool with its expert interface for the thermal design of multizone buildings. *International Journal of Solar Energy*, 8(2):109–120, 1990.

[19] Bruno Peuportier. Bancs d'essais de logiciels de simulation thermique. In *Journée SFT-IBPSA, Outils de simulation thermo-aéraulique du bâtiment*, La Rochelle, France, 2005.