M. Ceccarelli *et al.*, "Simulations in agricultural buildings: a machine learning approach to forecast seasonal energy need," *2022 IEEE Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, Perugia, Italy, 2022, pp. 116-120

The final published version is available online at:

https://doi.org/10.1109/MetroAgriFor55389.2022.9965083

Simulations in agricultural buildings: a machine learning approach to forecast seasonal energy need

1st Mattia Ceccarelli
Dep. of Agricultural and Food Sciences
University of Bologna
Bologna, Italy
mattia.ceccarelli5@unibo.it

2nd Alberto Barbaresi
Dep. of Agricultural and Food Sciences
University of Bologna
Bologna, Italy
alberto.barbaresi@unibo.it

3rd Giulia Menichetti
Department of Medicine
Harvard Medical School
Boston, MA, USA
gmenichetti@bwh.harvard.edu

4th Enrica Santolini
Dep. of Agricultural and Food Sciences
University of Bologna
Bologna, Italy
enrica.santolini2@unibo.it

5th Marco Bovo
Dep. of Agricultural and Food Sciences
University of Bologna
Bologna, Italy
marco.bovo@unibo.it

6th Patrizia Tassinari
Dep. of Agricultural and Food Sciences
University of Bologna
Bologna, Italy
patrizia.tassinari@unibo.it

7th Francesco Barreca
Dep. of Agriculture
University of Reggio Calabria
Bologna, Italy
fbarreca@unirc.it

8th Daniele Torreggiani
Dep. of Agricultural and Food Sciences
University of Bologna
Bologna, Italy
daniele.torreggiani@unibo.it

***Abstract*—A fast and reliable estimation of building energy need is essential in agricultural building design, nonetheless, a large number of simulations is required to obtain better energy saving solutions. The aim of this work is to understand if machine learning can substitute numerical simulations and speed up the building design process and assess the incidence of specific architectural elements. Supervised regression models has been trained and tested in a data-set of thousands simulations performed on a case-study agricultural building. Among the algorithms, the tree-based Extreme Gradient Boosting showed the best performance. A study on model explainability has been carried out using SHAP and features importance, which is fundamental to help academics and professionals devise better design strategies for both new constructions and retrofitting interventions.**

***Index Terms*—machine learning, building energy simulation, energy saving, ML explainability, food storage buildings**

I. Introduction

Europe, as most of worldwide regions, is facing a period of strong uncertainties related to the energy supply and costs. Building heating and cooling is one of the most energy consuming activity [1], [2]. Currently, several states are asking to change the indoor temperature sets reducing the consumption but, at the same time, the comfort. This solution can drive to remarkable results in the short period but must be considered as a temporary solution. In fact, an energy saving that reduces comfort cannot be considered as a real energy saving, in particular since it can bring to other problems related to the health of the building users, such as people or animals in livestock barns [3], and, in food storage buildings, to the food quality and safety.

Under this light, the process that will lead to more energy efficient buildings, both new and retrofitted, must be accelerated, made more systematic and cannot be limited to residential spaces or offices. The energy efficiency approach must be indeed extended also to buildings where indoor temperature plays a fundamental role even though the human presence is limited, such as greenhouse, food storage and ageing buildings [4], [5] and/or when nature based solutions are involved [6]. Since, in the latter, the preset temperatures are needed throughout the whole year, the indoor environment must be often controlled in both hot and cold seasons, adding a further variable to the design, considering that often the heating and cooling are performed by two different systems and several European standards in architecture are designed for cold seasons only.

In the last years, the use of computer energy simulations has helped to boost, optimise and fine-tune the energy efficiency in building design. Peculiar buildings - such as food storage buildings - take advantage of this approach since they can easily adopt active and passive solutions created for constructions with indoor environment requirements definitively different, such as residential spaces and offices [7]–[9]. During the design phase, computer software such as EnergyPlus [10] can simulate the whole building thermal behaviour and return the energy need, or energy consumption when a HVAC is installed. According to the details of model, the input variable, the complexity of the building, the process capacity of the computer and many other aspects, the simulations can take from seconds to minutes to run completely. This is definitely a short time if compared to the whole design and construction phase, on the other hand the modelling procedure requires a huge time and a single building can ask for dozen of simulations each of which needs modelling, analysing and adapting the model to the results of previous simulation results. Namely, a huge qualified human labour is required for a single building design [11]–[13].

Different approaches can be coupled to computer simulations to reduce the human labour, keeping or even improving the reliability of results. In fact, using codes to manage modelling and simulations (created by MatLab [14], Python and more) allows to add automation [15], [16], optimisation algorithms [17] and machine learning [18]. This approach drive to reduce drastically the required labour and even computational time, allowing to investigate and explore a countless number of building configurations and scenarios.

In this work, the Authors aim at testing a recent machine learning approach to a food storage building to assess the influence of architectural variables in both cold and hot seasons.

II. Materials and Methods

A. Energy Simulations

The tested approach is applied to a case study building located in the countryside of Bologna. The case study building (see Fig. 1) belongs to a wine-growing and wine-making farm and hosts both the wine-making phase and storage. The building is chosen as case study since can be considered representative of food storage building both for usage, dimensions and construction in central Italy. The building is symmetric on the two main axes and its measure are approximately $20m$ x $30m$ x $5-7m$. The structure is made by precast elements, the envelope does not respect any energy saving standard since the building was built decades ago. External walls are made with hollow concrete blocks, the floor is a 30-cm thick reinforced concrete slab, the rood is made by concrete and hollow bricks. Windows are single glazed and air permeability ensures a huge air infiltration.

Fig. 1. Case study building

To run the energy simulations, the software EnergyPlus was selected. The building has been modelled in OpenStudio and EnergyPlus and calibrated and validated in a previous work [19] and later updated to EnergyPlus 9.2 version. To assess different configurations and scenarios, the case study model has been modified using a MatLab code that automatised the building modelling, the energy simulations, and the result collection.

The architectural variables analysed in this work are visible in the Tab. I and their values are selected to investigate a wide range of building configurations. To ease the modelling phase, the input of the variables does not include materials but just theoretical values as better explained in a previous work [16]. The building orientation is inserted as one of the variables even though cannot be modified in retrofit interventions but can be a cost-less variable for new buildings. In the all models, the indoor temperature is set to be within 12° and 18°. The weather data have been collected on site in different years and reported in a weather file for the simulations to increase the accuracy of the study.

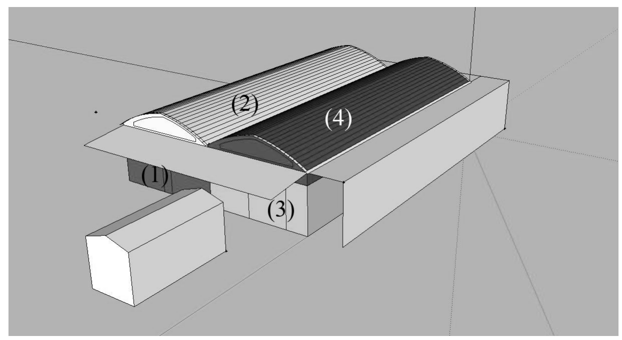

Fig. 2. Thermal zones of the case study model. The zone number 3 is the food storage area and is taken as reference for this study

At the end of the modelling process, the tested approach created more than 5000 models and the related EnergyPlus simulation returned nineteen results. Among those, the positive and negative energy loads to keep the temperature within the preset range are calculated. Those are respectively the heating and cooling energy need.

B. Models training and validation

The features and models selection procedure has been described in a previous work [18]. In this case, the dataset is divided in predictions for cooling and heating energy needs, which corresponds to hot and cold seasons. Two models—both *eXtreme Gradient Boosting* (XGB)—were trained and tested in predicting the total energy use of the same building configuration in the two seasons. Each model has been trained on a "train" set composed of 80% of data (4120 configurations) and validated on the remaining 20% (1030 configurations). For clarity, the 11 features chosen as input for the models are reported in Table I.

The distributions of energy use are shown in Figure 3. The plot shows that the energy used for cooling is much higher in general. This is due to the temperature range of the simulations and the effect of the sun which is positive for

TABLE I

LIST OF FEATURES USED BY *EnergyPLUS*. THE FIRST COLUMN REPORTS THE NAME OF THE VARIABLES, THE SECOND COLUMN SHOWS THE VARIABLE ABBREVIATIONS (USED HEREINAFTER), THE THIRD COLUMN SPECIFIES IF THE VARIABLE IS INSERTED BY THE USER (U) OR CALCULATED BY THE SOFTWARE (S), THE FOURTH COLUMN PROVIDES THE VARIABLE UNIT. THE FIFTH COLUMN SPECIFY IF THE FEATURE HAS BEEN SELECTED AS INPUT FOR MODEL REGRESSION.

Name	**Abbr**	**U/S**	**Unit**	**Selected**
wall resistance	wR	U	mK/W	✓
wall conductivity	wc	U	W/mK	x
wall density	wd	U	kg/m^3	✓
wall specific heat	wsh	U	J/(kgK)	✓
wall transmittance	Uw	S	W/(m^2K)	x
wall superficial mass	wsm	S	kg/m^2	x
wall attenuation	wa	S	-	x
wall thermal lag	wtl	S	hours	✓
roof resistance	rR	U	mK/W	✓
roof conductivity	wc	U	W/mK	x
roof density	rd	U	kg/m^3	✓
roof specific heat	rsh	U	J/(kgK)	✓
roof transmittance	Ur	S	W/(m^{2}K)	x
roof superficial mass	rsm	S	kg/m^2	x
roof attenuation	ra	S	-	x
roof thermal lag	rtl	S	hours	✓
orientation	o	U	degree	✓
air infiltration	ai	U	ACH[a]	✓
glaze transmittance	Ug	U	W/(m^{2}K)	✓

[a] Air Changes per Hour

heating and negative for cooling. It is also worth noting that the models have a lower density of data points to learn from for high-consuming configurations, which could worsen model performances.

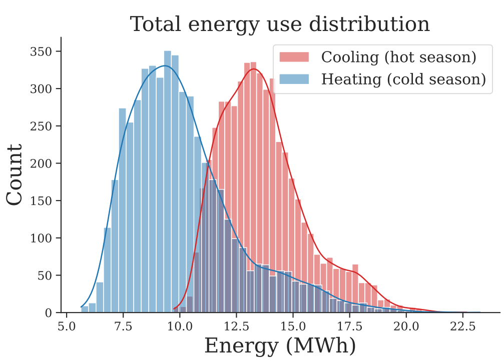

Fig. 3. Energy use distribution in the two data set. In blue, the distribution for the cold season and in red the distribution for the hot season

C. *SHAP explainability*

The impact of each feature on model output has been estimated using the python library *SHAP* (SHapley Additive exPlanations) [20]. *SHAP* is a game theoretic approach to interpret model output based on Shapley Values [21]. Shapley values are a system to distribute a reward in an n-persons game. Let's call $\nu(S)$ the *characteristic function* that maps subset of players into real numbers $\nu : 2^n \longrightarrow \mathbb{R}$. If S is a coalition of players, $\nu(S)$ is the total worth (or payoff) the coalition can obtain by collaboration. A Shapley value is the *fair* reward based of the contribution of each player to the coalition. The reward ϕ_i for player i can be computed as:

$$\phi_i(\nu) = \frac{1}{n!} \sum_{S \subseteq N \setminus \{i\}} |S|!(n-|S|-1)!(\nu(S \cup \{i\}) - \nu(S)) \quad (1)$$

where N is a set of n players and the sum extends for each subset S of N which doesn't contains player i. *SHAP* provide a model agnostic framework to compute features impact based on their contributions on the model output. *SHAP* values has been computed using the *TreeExplainer* for XGB [22] for each observation in the data set. When used to interpret model output, *SHAP* can provide features importance's for single observations and how features interact with each other. *SHAP* values has been computed for every configuration of the two datasets.

III. RESULTS

A. *Model Selection*

Figure 4 shows the results of the nested cross-validation on the four regression models tested. The XGB model outperforms the others in all metrics, except for computational time where a Linear Regression is faster. Considering the results, the XGB algorithm was selected and its predictions further explored through *SHAP*.

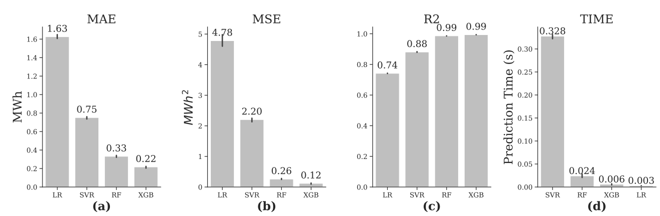

Fig. 4. (**a**) Average MAE; (**b**) average MSE; (**c**) average coefficient of determination R^2 and (**d**) average prediction time for the models, computed on each fold of the outer cross-validation.

In Figure 5 the fit diagrams between true and predicted value computed in the test set are shown.

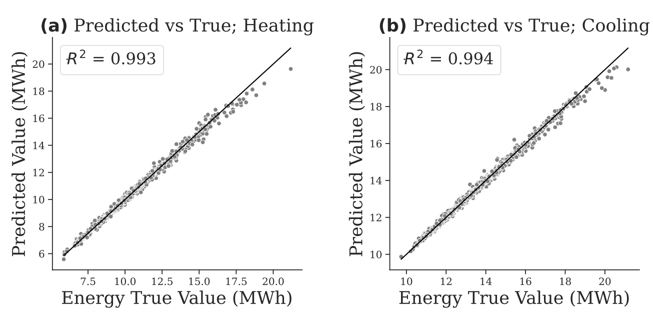

Fig. 5. Predicted vs True energy values for **(a)** Heating consume and **(b)** Cooling consume.

The model validation shows the high accuracy achieved by the regression predictions. As expected, a slight reduction of precision can be seen for high-energy-need configurations, probably due to the limited number of training data with high energy needs. Besides, considering the aim of most of the studies is to identify low-energy-need solutions, this reduction of the precision can be considered of minor importance for practical purposes. Figure 6 shows the influence of each feature on the model outputs, divided in heating and cooling.

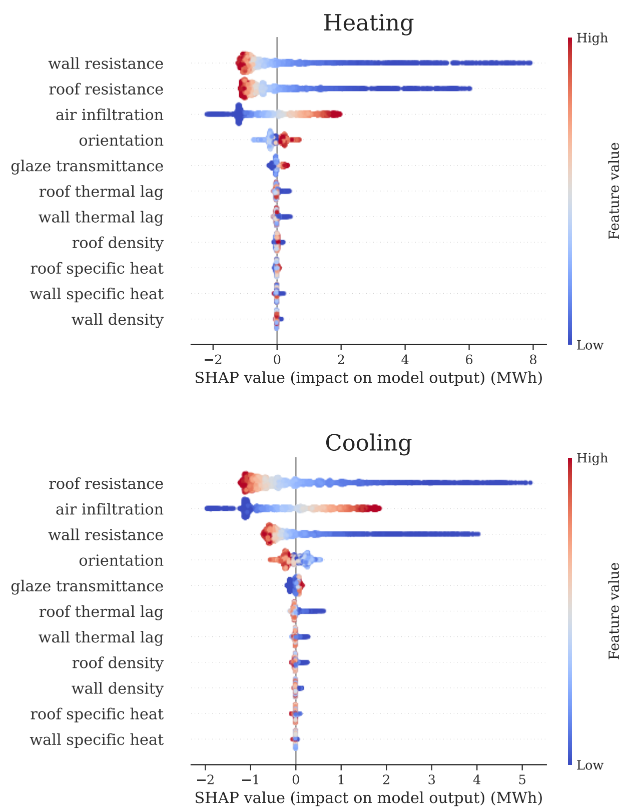

Fig. 6. SHAP values of each feature computed for each observations in the data set. On the top are shown the values for the heating model, on the bottom for the heating model. Every point represents a building configuration. The colours indicates the feature value qualitatively.

The variables are ranked on the basis of their average absolute SHAP values. In general, the two models judge *roof resistance, wall resistance* and *air infiltration* as the most influential features. The plots also show trends between the feature values and their corresponding impact: lower values of *roof resistance* and *wall resistance* have a positive impact (higher energy need) while high values have a negative impact (lower energy need). On the other hand, *air infiltration* has an opposite trend. It is worth noting that *wall* and *roof resistance* have a maximum negative impact on the energy needs about equal to $-1MWh$ (represented by the higher concentration of points around that value).

The first difference between the two models can be seen in the rankings of the variables. During heating, *roof resistance* is the first ranked feature, probably due to the effect of the sun. With the same reasoning, the reverse trend in *SHAP* values for *orientation* can be explained. Indeed, for heating energy, higher values of *orientation* correspond to a positive impact and lower values to negative impact instead. We observed the opposite behaviour for cooling energy. Another remarkable finding concerns the possibility to rank the investigated features according to their importance in the building energy need. *SHAP* can easily shows how much any feature affects the final result, allowing the personnel involved in the building design to focus more on the most important characteristics. Considering the rank can easily change even in the same building when some external factors change, e.g. weather data and/or thermostat settings, see [23], this result definitely helps to better drive the building design.

Under this light, a prediction model based on ML can be a useful tool to have fast and precise energy need predictions, eliminating the operator waiting time and avoiding energy simulation software. This method can be strongly needed when many feature configurations must be tested in a short amount of time.

IV. Conclusions

Today, accurate and fast prediction of the building energy need is a crucial matter in the path towards low-energy or near-zero-energy buildings. This work proved that an important advancement could come from the application of ML models. In fact, starting from the outcomes of several energy simulations on a case study building, three ML algorithms—SVR, RF, and XGB—were tested for the assessment of the energy need of the building under several configurations. The works confirm some findings of the previous paper [18]:

1) Fast computational time for predictions, much lower than what is required for an exact simulation;
2) Model validation shows a very high accuracy for XGB;

Moreover, it expands on results in model interpretation:

1) The models for cooling and heating change their *SHAP* values according to reasonable physical explanation, meaning that the models obtained information on environmental parameters from the data available.
2) The relative impact of each feature is also changed.
3) The method can be applied to both new and existing buildings. For the latter, the study of features importance can provide useful information for retrofit interventions.

This work demonstrated the efficacy of the proposed method that proved to be a valid alternative to the simulations and an additional tool that can integrate optimisation algorithms. We believe that it is possible to extend the results also for more complex buildings' configurations, including also geometrical characteristics. To achieve good precision, the proposed method needs an high number of simulations that are run anyway by the optimisation algorithms. Future works will consider an in-depth study of the amount of data necessary to achieve an acceptable regression. Further developments will investigate the application of ML models to different case studies and with the addition of more building features in order to test the ability of the models on a larger set of buildings and scenarios. Moreover, a more in-depth study on feature impact is planned.

References

[1] European Commission, "Energy Efficiency - Buildings," 2018.

[2] G. Ferrari, A. Pezzuolo, A. S. Nizami, and F. Marinello, "Bibliometric Analysis of Trends in Biomass for Bioenergy Research," *Energies 2020, Vol. 13, Page 3714*, vol. 13, p. 3714, jul 2020.

[3] A. Chiumenti, F. Da Borso, A. Pezzuolo, L. Sartori, and R. Chiumenti, "Ammonia and greenhouse gas emissions from slatted dairy barn floors cleaned by robotic scrapers," *Research in Agricultural Engineering*, vol. 64, no. 1, pp. 26–33, 2018.

[4] C. A. Campiotti, C. Bibbiani, A. Campiotti, E. Schettini, C. Viola, and G. Vox, "Innovative sustainable strategies in agro-food systems and in buildings for energy efficiency," *Innovative sustainable strategies in agro-food systems and in buildings for energy efficiency*, no. 2, pp. 79–96, 2016.

[5] A. Barbaresi, V. Maioli, M. Bovo, F. Tinti, D. Torreggiani, and P. Tassinari, "Application of basket geothermal heat exchangers for sustainable greenhouse cultivation," *Renewable and Sustainable Energy Reviews*, vol. 129, no. April, pp. 1–20, 2020.

[6] F. Fantozzi, C. Bibbiani, C. Gargari, R. Rugani, and G. Salvadori, "Do green roofs really provide significant energy saving in a Mediterranean climate? Critical evaluation based on different case studies," *Frontiers of Architectural Research*, vol. 10, pp. 447–465, jun 2021.

[7] F. Causone, R. Scoccia, M. Pelle, P. Colombo, M. Motta, and S. Ferroni, "Neighborhood Energy Modeling and Monitoring: A Case Study," *Energies*, vol. 14, no. 12, 2021.

[8] J. M. Felix Benitez, L. A. del Portillo-Valdés, V. J. del Campo Díaz, and K. Martin Escudero, "Simulation and Thermo-Energy Analysis of Building Types in the Dominican Republic to Evaluate and Introduce Energy Efficiency in the Envelope," *Energies*, vol. 13, no. 14, 2020.

[9] T. Alajmi and P. Phelan, "Modeling and Forecasting End-Use Energy Consumption for Residential Buildings in Kuwait Using a Bottom-Up Approach," *Energies*, vol. 13, no. 8, 2020.

[10] U.S. Department of Energy, "Energy Plus 8.0," 2013.

[11] S. Ferrari, F. Zagarella, P. Caputo, and G. Dall'O', "A GIS-Based Procedure for Estimating the Energy Demand Profiles of Buildings towards Urban Energy Policies," *Energies*, vol. 14, no. 17, 2021.

[12] S. Tsoka, K. Velikou, K. Tolika, and A. Tsikaloudaki, "Evaluating the Combined Effect of Climate Change and Urban Microclimate on Buildings' Heating and Cooling Energy Demand in a Mediterranean City," *Energies*, vol. 14, no. 18, 2021.

[13] A. Blumberga, G. Bazbauers, S. Vancane, I. Ijabs, J. Nikisins, and D. Blumberga, "Unintended Effects of Energy Efficiency Policy: Lessons Learned in the Residential Sector," *Energies*, vol. 14, no. 22, 2021.

[14] Mathworks, "Matlab R2017b," 2017.

[15] A. Barbaresi, M. Bovo, and D. Torreggiani, "The dual influence of the envelope on the thermal performance of conditioned and unconditioned buildings," *Sustainable Cities and Society*, vol. 61, pp. 1–20, oct 2020.

[16] A. Barbaresi, G. Menichetti, E. Santolini, D. Torreggiani, and P. Tassinari, "Two-Step Optimization of Envelope Design for the Reduction of Building Energy Demand," in *Building Simulation 2019*, (Rome), pp. 3055–3062, IBPSA, 2019.

[17] J. H. Holland, *Adaptation in natural and artificial systems*. Cambridge: MIT Press, 1975.

[18] A. Barbaresi, M. Ceccarelli, G. Menichetti, D. Torreggiani, P. Tassinari, and M. Bovo, "Application of Machine Learning Models for Fast and Accurate Predictions of Building Energy Need," *Energies 2022, Vol. 15, Page 1266*, vol. 15, p. 1266, feb 2022.

[19] A. Barbaresi, F. Dallacasa, D. Torreggiani, and P. Tassinari, "Retrofit interventions in non-conditioned rooms: calibration of an assessment method on a farm winery," *Journal of Building Performance Simulation*, vol. 10, no. January, pp. 91–104, 2017.

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 4765–4774, Curran Associates, Inc., 2017.

[21] L. S. Shapley, "17. A Value for n-Person Games," *Contributions to the Theory of Games (AM-28), Volume II*, pp. 307–318, may 2016.

[22] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 2522–5839, 2020.

[23] A. Barbaresi, M. Bovo, and D. Torreggiani, "The dual influence of the envelope on the thermal performance of conditioned and unconditioned buildings," *Sustainable Cities and Society*, vol. In press, 2020.