MACHINE LEARNING DRIVEN OPTIMIZATION OF A HYBRID ELECTRICAL AND THERMAL SYSTEM

M. Dallapiccola, F. Trentin, G. Barchi, C. Dipasquale, R. Fedrizzi, D. Moser
Institute for Renewable Energy, EURAC Research, 39100 Bolzano, Italy;
Phone: +39 0471 055 617; Fax: +39 0471 055699; E-mail: mattia.dallapiccola@eurac.edu

ABSTRACT: With the diffusion of electric heating and cooling devices, coupling the electric and thermal systems in the residential sector is becoming attractive and could help to increase photovoltaic penetration. The heating and cooling needs of buildings correspond to an important component of the total energy consumption of the residential sector. Thus, it is important to properly design the thermal and electric systems accounting of the interactions from the first phases of the design process. In the design phase, detailed models implemented in dynamic simulation tools can be used for the sizing process of system components, but they hardly can be adopted in optimization algorithms due to the computational time required for each simulation. This is particularly true for multi-objective optimization algorithms, where usually a wide number of simulations is required. In this work, TRNSYS was used to train a machine learning model that is used in a multi-objective optimization with the final goal of improving the design of the thermal system and optimizing the KPIs of a coupled photovoltaic plus battery system.

Keywords: PV System, Simulation, Sizing, Storage, System Performance

1 INTRODUCTION

In Europe, the residential sector accounts for 27% of the total energy consumptions and more than half is required for heating and cooling and domestic hot water preparation. Due to the diffusion of heat pumps and other electric devices, the coupling of the thermal and electric systems in hybrid systems is becoming more and more attractive and could be an important driver for the future installations of photovoltaics systems in the residential sector. The benefits of coupling photovoltaics system with heating and cooling devices was discussed in many studies [1], [2], [3]. Considering also the effect of climate changes [4], [5], in the recent years there is a growing interest related to cooling applications [6], [7]. Advanced controls to improve self-consumption were also investigated in many articles such as [8] and [9], however in many of these studies the size of the system components was usually considered fixed by the authors and not optimized. In many previous analysis, it was demonstrated how it is possible to apply optimization techniques for the design and sizing of photovoltaics and battery systems [10], [11], [12] leading to improved system design. In these studies, related to design optimization, the electric demand was often considered as an input for the optimization algorithm and the interaction between the thermal side and electric side of the system was not considered. This approach is acceptable if the designer is only interested in the optimization of the electric system. However, the authors recognized that when designing a hybrid system, it is important to include from the design process the interactions between the electric and thermal side of the system, optimizing together the size of the electric and thermal components and not as two separate systems. But even if dynamic simulation software are more precise than grey or black box models [13], their complexity could result in a process which is too computationally expensive especially if included in an optimization algorithm. One possible solution to reduce the computational time is to simplify the models of the building and of the system as in [14], [15]. In this paper we propose an approach similar to the one presented in [16], where machine learning was introduced to decrease the computational effort. The machine learning model is trained to substitute the dynamic simulation in the optimization and unlock the possibility to use a multi-objective optimization algorithm. Finally, the dynamic model is used to check the results of the selected optimal configurations with a short-time step simulation. The method presented in this paper is applied to one of the reference cases of the Horizon 2020 project HYBUILD [17] that focuses on innovative hybrid electrical-thermal storage systems for the residential sector.

2 METHOD

As described in the Introduction, the design of a complex hybrid thermo-electrical system using a dynamic model, may be complex and time-consuming. For this reason, the aim of this work is to propose an approach which is based on a multi-objective optimization able to design in the different parts of the system in a single process. As previously mentioned, the present methodology is applied to one of the reference case of the Horizon 2020 HYBUILD project, however it could be also adopted to different systems with a larger number of design variables. In this section we briefly present the project context and the hybrid system considered in the case-study. Then we introduce the proposed method defining the design variables, the objective functions and the adopted machine learning (ML) model.

2.1 Hybrid system case study

The electric-thermal system considered in this paper is related to one of the reference building in the HYBUILD project. This European project has the aim to integrate advanced electric and thermal storage systems in residential building able to cooperate in order to not only fulfill the electrical or thermal demand of the final user but also to maximize the use of renewable and reduce the CO_2 emissions.

Due to this purpose, the project proposes two different hybrid systems suitable to use in two different European climates, i.e. Mediterranean and Continental. In this work we focus only on the system developed for the Mediterranean area. In particular, the building considered is a Single-Family House (SFH) built in the period 1980-1990 and placed in the city of Athens that was selected as a reference location. The building is a two-floors house with a net area of 100 m^2. According to the purpose of the HYBUILD project, the aim of the Mediterranean system is the covering of space cooling demand and to fulfill the requirement for the domestic hot water needs. **Figure 1** shows a simplified scheme of the hybrid system composed by the following sub-systems:

- Solar thermal collector: the solar thermal collector sub-system is composed by Fresnel concentrating solar collector, buffer tank and the hydraulic equipment that allows the connection of the two components. The buffer tank is used for different purposes: to feed up the adsorption chiller, to charge the DHW tank and, although this last option for the sake of clarity is not reported in **Figure 1**, to ensure heating to the building exploiting the solar source when it is available.
- Adsorption module: the adsorption machine uses as input the heat coming from the buffer tank to produce chilled water that is sent at the condenser side of the compression chiller rejecting heat through a dry cooler.
- Compression chiller and PCM: the compression chiller is mainly devoted to the latent storage charge. If the operation constraints of the adsorption machine are not satisfied and at the same time the compression chiller operation is needed, there is the possibility to bypass the adsorption part and to connect directly the dry cooler to the condenser side of the compression chiller. There is also the possibility to bypass the latent storage and connect directly the compression chiller to the distribution circuit.
- Distribution circuit: it is composed of fan coils that are used to ensure the desired indoor temperature.
- Domestic hot water circuit: DHW circuit guarantees the correct flow of hot water to the users in the apartment. To exploit solar source also for DHW purposes, the DHW tank can be charged using the buffer tank. An electrical resistance placed inside the DHW tank is considered as a back-up element.
- Electric system: a photovoltaic system composed by multi-crystalline modules connected in series, an inverter connected to the external grid and an electric battery considered ideal.

The building and the systems described in the previous section, were modelled in TRNSYS [18], where the climate data are taken from the Meteornom database [19]. Each component of the system was modelled by standard (when available) or non-standard TRNSYS types with performance maps given directly by industrial partners or expertise in the modelling of the single component. The timestep used for the simulations is one minute, minimum requirement to well approximate the real operation of the studied system considering its complexity, the high number of components involved and, consequently, the high number of possible operational modes. Due to the complexity of the model and the simulation timestep, each simulation takes about an hour.

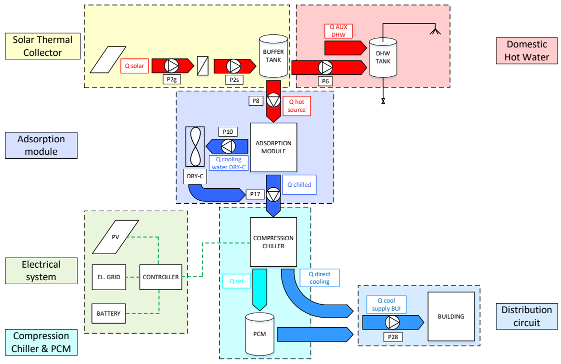


Figure 1: simplified scheme of the hybrid electric-thermal system used as case-study. The scheme represents the Mediterranean reference case of the HYBUILD project.

2.2. Optimization and machine learning model

As reference to the system in **Figure 1**, since the building was already defined, as well as the size of most of the other components, we focused on the design of four components of the system. For each of these components, we select a sizing variable which are:

1. Number of solar thermal modules of the solar thermal system
2. Number of modules for the PV system
3. Number of battery cells for the electric storage system
4. Diameter of the tank for the domestic hot water storage

To obtain an optimized system from both the electric and thermal point of view, a multi-objective optimization algorithm was selected for the analysis. The considered four cost functions are:

1. Maximization of self-consumption (SC), defined as the ratio between the self-consumed energy and the total energy produced by the photovoltaic system
2. Maximization of self-sufficiency (SS), defined as ratio between the self-consumed energy and the total energy consumed
3. Minimization of the cumulative electric energy consumed by the system (i.e. sum of the electrical consumptions of the compression chiller, dry cooler and circulating pumps.)
4. Minimization of system cost

Considering the computational burden of the TRNSYS model and its significant increase when the optimization is also performed, the idea introduced in this work is to adopt a machine learning model that could substitute the TRNSYS model in the optimization process. Thus, the dynamic model was used to create a dataset for the training of the machine learning model able to estimate three relevant indicators describing the performance of the system using as inputs the design variables. The machine learning model was then used in the optimization algorithm to size the components of the system reducing the computational time required. Finally, the white box model (e.g. TRNSYS model) was used to check the results in a dynamic simulation.

3 IMPLEMENTATION

This section describes the procedure to pass from the TRNSYS model to a machine learning model and how to run the optimization to size the components. A summary of the steps is represented in Figure 2.

3.1 Creating the dataset

To prepare the data required for the training of the machine learning model, a set of parametric simulations were done with TRNSYS. The parameters of the configurations simulated are the four design variables that we selected as optimization variables: the number of solar thermal modules installed, the capacity of the photovoltaic system, the capacity of the battery and the diameter of the domestic hot water storage. The idea was to create a database of inputs-outputs based on simulations to be used in the training phase of the machine learning model. A scheme of the workflow is reported in **Figure 2**. To create the set of configurations to be simulated an interval of acceptable values was defined for each design variable. **Table I** reports the intervals used.

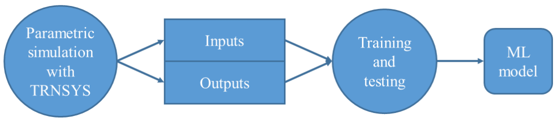


Figure 2: workflow of the training process

Table I: Intervals

	Min	Max
Number of ST modules	1	3
Number of PV modules	0	30
Number of battery cells	0	1000
DHW storage diameter [m]	0.4	0.8

The statistical technique Latin hypercube sampling [20] was used to obtain the set of configurations to be simulated with TRNSYS. Latin hypercube sampling is a widely used technique to originate a quasi-random set of parameters sampling N different sets of parameters between the given intervals. In this case, we used 150 samples. To obtain the final dataset, TRNSYS was coupled with Python [21] to automatically run all the simulations. Since the Mediterranean reference case focuses on cooling, the simulations and the reported KPIs refer to the cooling season (period from beginning of May to beginning of October). The output of this process is a database of 150 rows (different configurations) where the columns are the inputs (design variables) and outputs of the TRNSYS simulations in terms of cost functions (SC, SS, E_{tot}). Initial costs of the plant are not included in this phase since they can be fully determined once the size of the components is known. The complete process lasted about 6 days. The dataset obtained was then divided in two sub-sets corresponding to the input (size of the components) and output arrays (SC, SS, E_{tot}) to be used in the machine learning training phase. Since no nans data or outliers were identified in a first analysis, the only pre-processing step done was the use of the Standard Scaler implemented in scikit-learn [22]. According to the documentation, the Standard Scaler removes the mean and scales the features to unit variance. This step is needed to scale all inputs and avoid that large order of magnitude inputs dominate over smaller scale inputs in the training process. The formula used in the scaling process is:

$$z = (x - u)/s \qquad (1$$

where z is the scaled value, x is the original value, u is the mean of the feature and s is the standard deviation.

3.2 Training and testing

Between the different algorithm tested, the best performance were obtained with a Random Forest Regressor [23]. The basic idea behind Random Forest is to build multiple decision trees and obtain a more stable and accurate model averaging the predictions of all the trees. To significantly improve the performances, one regressor per target was trained with the MultiOutputRegressor function implemented in scikit-learn. Since the dataset is not huge, the 10-fold cross validation method was selected for the validation of the model. Thus, the dataset was divided in 10 subsets and the model trained 10 times. For each training, 9 folds were used for training and the remaining one for testing. Finally, the accuracy of the model was estimated as the average score of the 10 tests. The most important advantage of using K-fold cross validation is that it gives a more stable estimation of the accuracy of the model to predict never seen data with respect to other methods such as R^2.

3.3 Optimization

The algorithm used for the optimization of the design variables is the NSGA-II, a widely used multi-objective evolutionary algorithm [24]. This algorithm was chosen for its set-up simplicity and because it allows the user to optimize simultaneously all objectives returning a set of optimal solutions that can be analysed at the end of the optimization process. In the optimization process, a mask was applied on the design variables to obtain only integer values for the number of ST modules, PV modules and battery cells. The diameter of the domestic hot water storage was considered as a continuous variable. All the steps related to optimization were implemented with the pymoo [25] Python library using the default settings, a population of 100 individuals and a maximum of 200 generations. For each individual, the algorithm calculates the target functions SC, SS and the total consumed energy with the machine learning model and calculates the initial cost of the configuration. To calculate the initial cost of each solution, the following assumptions were considered:

- cost of ST modules = 400 €/m^2
- cost of PV modules = 1500 €/kWp
- cost of battery = 670 €/kWh
- cost of the DHW storage = 2000 €/m^3.

The initial cost of the system is calculated as the sum of each component cost obtained multiplying the unit cost by the size.

4 RESULTS AND DISCUSSION

As discussed in the previous paragraph, the accuracy of the machine learning model to predict the output of TRNSYS was tested with a 10-fold cross validation. The average accuracy of the model was estimated equal to 96.4±2.1% and was considered acceptable for the goals of this work. **Figure 3** represents the goodness of the prediction with respect to the values obtained with TRNSYS. Since the accuracy of the model depends on the

goodness of the training process, the training and testing phase were repeated using an increasing percentage of the data available randomly sampled from the initial database with the goal of investigate the impact of the number of configurations used in the training phase on the accuracy. As reported in **Figure 4**, 60% of the initial database is enough to train a model with a good accuracy. In this case, it means 90 simulations: the ratio between the number of simulations required and the number of design variables is thus equal to 22.5, confirming the value found in [16]. However, the following pages refers to a model trained using the entire available database. The trained model was finally used in the multi-objective optimization algorithm to predict at each step the self-consumption and self-sufficiency indexes and the total energy consumed by the system for the considered period. The result of the optimization is a set of Pareto-optimal solutions represented in **Figure 5**. At this point the designer can choose among the final set of solutions those that he considers more relevant and proceed with further evaluations.
The authors selected only the solutions with SC and SS greater than 40% to avoid undersized or oversized PV systems. The filtered set was then reordered by initial cost in ascending order. Finally, the solutions reported in **Table II** were selected and simulated with TRNSYS. They are highlighted in green in **Figure 5**. This step is needed to verify with a dynamic simulation that the optimized configurations of the plant satisfy the comfort requirement of the building. In fact, the system was optimized using as target functions the cumulative indexes of self-consumption, self-sufficiency and consumed energy but there is no guarantee that the plant works as expected even for a small timestep simulation.

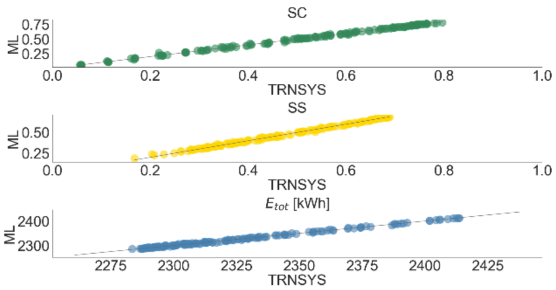


Figure 3: machine learning model (ML) predictions compared with TRNSYS simulations. The plot confirms the goodness of the training process and the absence of outliers.

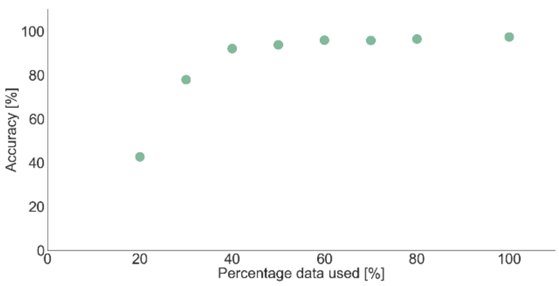
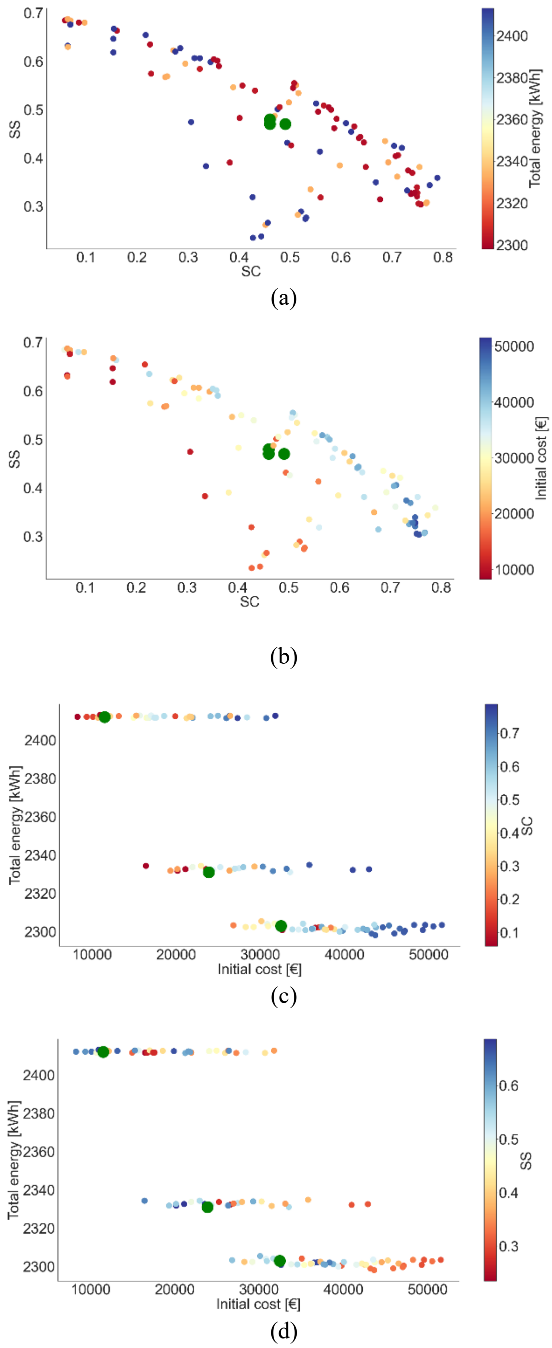


Figure 4: accuracy of the model compared to the percentage of data used in the training phase. 60% of the dataset could be enough to obtain an accurate model. The configurations selected for testing were simulated and the results compared to those obtained with the machine learning model. The comparison is reported in Table III. Considering that the approach presented in this work is intended as a support for the designer in the first phases of the design process of a complex hybrid system, results presented in Table III were considered acceptable in terms of predictions of the indicators. Moreover, the dynamic simulations confirmed that the cooling set point of the internal air temperature is guaranteed for all the three configurations for the entire cooling period. Also, the DHW demand was always satisfied.



Figures 5: Set of optimal configurations: due to the number of objective functions multiple plots are needed to represent the solution. The solutions selected for the dynamic simulations were highlighted in green.

Table II: final configurations selected for detailed dynamic simulation

Configuration	1	2	3
Number of ST modules	1	3	2
Number of PV modules	12	13	19
Number of battery cells	117	205	106
DHW storage diameter [m]	0.50	0.58	0.79
SC [%]	0.46	0.49	0.46
SS [%]	0.48	0.47	0.47
E_{tot} [kWh]	2412	2303	2331
Initial costs [€]	11465	32438	23831

Table III: comparison of the objective functions obtained with TRNSYS and ML. Values confirms that the ML model can estimate well the KPIs selected.

Configuration 1	SC	SS	E_{tot} [kWh]
TRNSYS	0.46	0.44	2437
ML	0.46	0.47	2412
Configuration 2			
TRNSYS	0.50	0.52	2368
ML	0.49	0.48	2303
Configuration 3			
TRNSYS	0.32	0.50	2371
ML	0.29	0.50	2333

Comparing the computational time required for the optimization process with and without the adoption of machine learning techniques, it is possible to highlight the great advantage of the approach used in this work. In fact, according to the number of simulations required, using directly TRNSYS in the optimization algorithm would have required more than 2 years to end the process. On the contrary, the approach with machine learning can be completed in less than 7 days, considering the time required for training, optimization and simulation of the most interesting solutions selected by the designer. Additionally, computational time could have been reduced even more as shown in **Figure 4**.

5 CONCLUSIONS

In this paper, the authors demonstrated how it is possible to use machine learning to speed up the optimization process in the first phases of the design of a hybrid thermal and electric system of a residential building. In some cases, when the number of design variables or the number of objective functions is large, using a simplified or a black box model could be the only solution reasonable from the computational point of view. The advantage of using machine learning in this kind of optimization problems is that it could replace TRNSYS in the calculation of the target functions reducing drastically the computational time. This step unlocks the possibility to use a multi-objective optimization algorithm which typically requires a great number of iterations to solve a complex problem. The result of the optimization is a set of optimal solutions that the designer can further evaluate to refine the design process. The approach used in this work demonstrate that relevant KPIs such as SC and SS, often needed in the design phase of a residential PV system, can be estimated well with machine learning. This opens the possibility to optimize together the PV plus battery system and the thermal side of the plant (potentially also the building characteristics could be included as design variable). On the other side, the dynamic model is required to create the training set and at the end of the optimization to check with a small simulation timestep that the comfort requirements of the buildings are satisfied. In the next years, where the installation of photovoltaics coupled with electric heating or cooling devices is expected to increase, including this approach in the first phases of the design process could lead to an improvement of the performances and economic outcome of hybrid systems.

7 ACKNOWLEDGEMENTS

The HYBUILD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 768824.

8 REFERENCES

[1] F. Aguilar, S. Aledo, and P. Quiles, "Experimental analysis of an air conditioner powered by photovoltaic energy and supported by the grid," *Applied Thermal Engineering*, vol. 123, pp. 486–497, 2017.
[2] J. Salpakari and P. Lund, "Optimal and rule-based control strategies for energy flexibility in buildings with PV," *Applied Energy*, vol. 161, pp. 425–436, 2016.
[3] F. Aguilar, D. Crespí-Llorens, and P. Quiles, "Environmental benefits and economic feasibility of a photovoltaic assisted heat pump water heater," *Solar Energy*, vol. 193, pp. 20–30, 2019.
[4] E. Zanetti, M. Aprile, D. Kum, R. Scoccia, and M. Motta, "Energy saving potentials of a photovoltaic assisted heat pump for hybrid building heating system via optimal control," *Journal of Building Engineering*, vol. 27, p. 100854, 2020.
[5] L. Langer and T. Volling, "An optimal home energy management system for modulating heat pumps and photovoltaic systems," *Applied Energy*, vol. 278, p. 115661, 2020.
[6] L. W. Davis and P. J. Gertler, "Contribution of air conditioning adoption to future energy use under global warming," *Proceedings of the National Academy of Sciences*, vol. 112, no. 19, pp. 5962–5967, 2015.
[7] M. Santamouris, "Cooling the buildings–past, present and future," *Energy and Buildings*, vol. 128, pp. 617–638, 2016.
[8] H. S. Laine, J. Salpakari, E. E. Looney, H. Savin, I. M. Peters, and T. Buonassisi, "Meeting global cooling demand with photovoltaics during the 21st century," *Energy & Environmental Science*, vol. 12, no. 9, pp. 2706–2716, 2019.
[9] M. Watt, S. Partlin, M. Oliphant, I. MacGill, and T. Spooner, "The value of PV in summer peaks," 2004.
[10] M. Lovati, G. Salvalai, G. Fratus, L. Maturi, R. Albatici, and D. Moser, "New method for the early design of BIPV with electric storage: A case study in northern Italy," *Sustainable Cities and Society*, vol. 48, p. 101400, 2019.
[11] K. Ndwali, J. G. Njiri, and E. M. Wanjiru, "Multi-objective optimal sizing of grid connected photovoltaic batteryless system minimizing the total

life cycle cost and the grid energy," *Renewable Energy*, vol. 148, pp. 1256–1265, 2020.
[12] Y. Ru, J. Kleissl, and S. Martinez, "Storage size determination for grid-connected photovoltaic systems," *IEEE Transactions on sustainable energy*, vol. 4, no. 1, pp. 68–81, 2012.
[13] V. Harish and A. Kumar, "A review on modeling and simulation of building energy systems," *Renewable and sustainable energy reviews*, vol. 56, pp. 1272–1292, 2016.
[14] L. Magnier and F. Haghighat, "Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network," *Building and Environment*, vol. 45, no. 3, pp. 739–746, 2010.
[15] "Horizon 2020 HYBUILD, grant agreement N° 768824." http://www.hybuild.eu/.
[16] S. A. Klein, W. A. Beckman, J. W. Mitchell, N. A. Duffie, and et al., *TRNSYS 17, transient system simulation program*. University of Wisconsin, Madison, WI, USA, 1979.
[17] "Meteonorm." http://www.meteonorm.com/.
[18] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 42, no. 1, pp. 55–61, 2000.
[19] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.
[20] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[21] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[22] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.
[23] J. Blank and K. Deb, "pymoo: Multi-objective Optimization in Python," *IEEE Access*, pp. 1–1, 2020.