



Approximating optimal building retrofit solutions for large-scale retrofit analysis

Emmanouil Thrampoulidis ^{a,b,*}, Gabriela Hug ^a, Kristina Orehounig ^b

^a EEH-Power Systems Laboratory, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

^b Laboratory for Urban Energy Systems, Empa Dübendorf, Switzerland

HIGHLIGHTS

- Bottom-up approach for predicting near-optimal building retrofit solutions.
- Artificial neural networks ensure scalability for large-scale retrofit analysis.
- Retrofit model for residential buildings of various climatic regions.
- Trade-off between accuracy, computational cost, and ease of application.
- Near-optimal retrofit solutions for residential buildings in Geneva municipality.

ARTICLE INFO

Keywords:

Building retrofit
Surrogate model
Machine learning
Large-scale
Pareto-front

ABSTRACT

Coordinated large scale building retrofit is one of the most effective approaches not only to reduce the environmental footprint of existing buildings but also to assure that climate goals can be met. Existing bottom-up retrofit approaches become infeasible or lack in terms of accuracy and/or computational cost especially when the scale of application increases in size. Hence, the aim of this paper is to propose and develop a bottom-up approach for large-scale building retrofit with the use of a scalable building-level surrogate retrofit model. To ensure scalability and generalization to large areas, we use artificial neural networks. The developed approach consists of a set of artificial neural networks trained on building archetypes, aiming to predict near-optimal building retrofit solutions for all residential buildings in Switzerland. This is achieved with a small set of inputs, and within a short run time. In order to validate the developed approach, we compare it with a conventional mixed building simulation and optimization approach. The surrogate model computes near-optimal building retrofit solutions, with an average coefficient of determination of 0.91 and an average f1 score of 0.89. To illustrate the advantages of this model, we perform a large-scale retrofit analysis of a case study in Geneva municipality. Such a surrogate model is capable of providing insights into the overall retrofit potential of large areas and thus shall support reaching energy emission reduction targets of existing residential buildings.

1. Introduction

1.1. Background

The increased frequency of extreme weather phenomena, as well as the observed global warming, with their consequences, has strengthened the awareness of the need for a more sustainable built environment. In Europe, but also globally, the building sector is responsible for the most significant share of energy consumption and Green House Gas (GHG) emissions, accounting for almost 40 % and 36 %, respectively [1]. Thus,

there is a vital need to improve existing and new buildings' efficiency.

Energy targets for buildings have been established in the European Union and are described in the Energy Performance of Buildings Directive [2], accompanied by its amendment [3]. The latest recast [4] proposes new standards for energy performance and sets the roadmap until 2050. Similar targets have been set in Switzerland. Specifically, one of the main pillars of the Swiss energy strategy 2050 [5] is to improve the energy efficiency of buildings and increase renewable energy technologies' penetration. All new buildings have to be nearly net zero-energy buildings, this as an attempt to eliminate the increase of the environmental footprint of the building sector. However, this measure

* Corresponding author at: EEH-Power Systems Laboratory, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

E-mail address: Emmanouil.Thrampoulidis@empa.ch (E. Thrampoulidis).

Nomenclature	
<i>General Abbreviations</i>	
ANNs	Artificial Neural Networks
ASHP	Air Source Heat Pump
Bio boiler	Bio-mass boiler
GFA	Ground Floor Area
CHF	Swiss Franc
Class	Classifiers
CPU	Central Process Unit
GHG	Greenhouse Gas
GSHP	Ground Source Heat Pump
GWR	eidgenössische Gebäude- und Wohnungsregister(Federal Register of Buildings and Dwellings)
HDD	Heating Degree Days
MGT	Micro-Gas turbine
MFH	Multi-Family House
MILP	Mixed Integer Linear Programming
ML	Machine Learing
PF	Pareto Front
PV	Photovoltaics
Regs	Regressors
ReLU	Rectified Linear Unit
SFH	Single Family House
<i>U-Value</i>	Thermal transmittance
w2w	Window to Wall
<i>Variables</i>	
<i>n</i>	Number of samples
<i>p</i>	Pareto front number
<i>k</i>	Optimization objective index
<i>f</i>	Objective function
<i>i,j</i>	Pareto front point indices
<i>R</i> ²	Coefficient of determination
<i>R</i> ² _{building}	Per-building coefficient of determination
λ	Regularization parameter

does not address the environmental impact of existing buildings which is the majority of the building stock. Eventually, building retrofit, especially large-scale retrofit, is considered of highest importance to reduce the overall share of energy consumption and GHG emissions of the building sector. However, retrofit rates per year are still low, namely below 1 % in Switzerland [6]. In contrast, the New Energy policy scenario of the Energy Strategy 2050 aims at an average per year renovation rate of 1.9 %. Depending on the construction year and the building type this rate increases to 2.8 % for some years. Several barriers to building retrofit exist and are described in detail in [7].

Building retrofit is the process of adjusting a building after it has been constructed and occupied to improve its environmental footprint. For the building retrofit process to be successful a well-designed decision-making process must be applied. This usually refers to decisions on building envelope interventions, such as adding insulation to the roof, façade or ground floor, replacing windows with ones with better U-values, or energy system replacements. The latter pertains not only to replacing the heating or cooling system with a more efficient system, but also the possibility of installing storage or renewable technologies, such as photovoltaics (PV). The aim of building retrofit differs depending on the scale of application. When building retrofit is applied to a single building, we refer to individual building-level retrofit, and the main goal is to improve the building's energy efficiency while ensuring or improving indoor environmental comfort. On the other hand, when the building retrofit is applied to large building stocks, such as neighborhoods or municipalities, in a reasonable period of time, we refer to it as large-scale building retrofit. This process is a key component in designing and/or assessing energy strategies' interventions to achieve country-level or global energy targets.

1.2. The necessity for large-scale retrofitting

It could be disputed that, in contrast to individual actions, large-scale actions allow a much higher control in meeting national targets. This also applies for building retrofit. Even though building-specific retrofit is of high importance, it should be part of a coordinated large-scale retrofit plan. This is also suggested after a thorough review of the available retrofit decision support methodologies in [8]. Building-level retrofit might lead to piecemeal interventions which limit the potential available for reaching the energy targets at national and global levels [9]. On the other hand, with large-scale retrofit, we can benefit from: (i) a balanced responsibility and involvement of various stakeholders (e.g., municipality, building owners, and policymakers), (ii) a more systematic and effective approach to derive energy policies, strategies, and

incentives that can facilitate the broader adoption of retrofit solutions, (iii) the advantages of centralized-supply technologies, and (iv) the economies of scale. Centralized supply technologies or else district-level solutions follow the economic behavior of the economies of scale and thus have certain advantages compared to decentralized ones. Some of them are lower costs, when compared to many small decentralized systems, better efficiencies and emission profiles due to lower losses, more efficient operation and lower maintenance costs and easier integration of renewable energy technologies. However, depending on the studied objective, building level solutions might be preferable. Indeed, both building-level and district-level solutions should be evaluated before the decision process.

The boost of successful applications of machine learning (ML) in various fields, such as in image recognition, led to the application of ML to building retrofit. ML as a data-driven approach can offer specific characteristics that are desired in the retrofit process. Firstly, the generalization ability and secondly the reduced computational costs when used as a surrogate model. Moreover, ML solutions are more favored for easy-to-use web applications for early stage assessment of retrofitting measures. Therefore, ML-based retrofit solutions can contribute towards accelerating the adoption of retrofit measures, mainly by the building owners.

1.3. Related work

In the literature, there exist many approaches for individual building-level retrofit. Some of them are performed with the sole use of building simulation tools [10], but are limited to a small number of retrofit scenarios and often do not lead to the optimal solution. More advanced approaches combine building simulation with an optimization process to ensure optimal or near-optimal solutions. In the literature, a variety of methods are used for retrofit optimization [11]. The most dominant one is multi-objective optimization to ensure the satisfaction of several criteria that influence both the environment and the building owners (e.g. involved cost and GHG emissions). For instance, in [12], building simulation is coupled with a multi-objective optimization with the use of harmony search algorithm in order to generate near-optimal building envelope interventions. The objectives are minimizing the costs, the emissions and maximizing the comfort in the building. On the other hand, in [13] the multi-objective optimization problem is formulated as a mixed integer linear programming (MILP) in order to ensure optimality of the generated retrofit solutions.

With regards to large-scale building retrofit, two main approaches [14] can be used: top-down and bottom-up approaches.

1.3.1. Top-Down retrofit approaches

Top-down approaches are based on statistical methods and start by collecting and analyzing aggregated energy demand data, e.g., of a city, and then subdividing it by building type or by smaller areas. The main focus is to correlate energy demands with economic factors, such as income. Balaras et al. [15], developed a top-down country-level analysis to propose guidelines for retrofitting with the goal to reduce CO_2 emissions. Fracastoro and Serraino [16] developed a similar approach to, among others, evaluate the energy-saving potential in the case of building envelope interventions for large case studies.

TIMES [17] and NEMS [18] are commercially available energy models that also provide an econometric analysis following a top-down approach. The main disadvantage of such approaches is that they rely on standardized or historical data. This means that they have difficulties to capture the effect of new energy technologies or policies. Moreover, they lack in accuracy when a finer building analysis is needed. Recently, there has been a trend to utilize ML to overcome those challenges. A detailed review of such top-down approaches is given in [9].

1.3.2. Bottom-up retrofit approaches

Bottom-up retrofit approaches refer to the investigation of a small set of buildings and generalizing the results to large areas, i.e., districts, cities, and municipalities. Many studies focus on the development of representative buildings, called building archetypes, which are then used to extrapolate to larger scales. The reason for using building archetypes is to perform large-scale analysis with reasonable computational time and cost. Building archetypes are formally defined in [19] as “building definitions that represent a group of buildings with similar properties”. He et al. [20] used Energy Plus to derive energy demand profiles and employed multi-objective optimization to derive optimal building envelope interventions for building archetypes for the North East of England. In [21], the potential of energy and emission savings in five European countries was calculated with a bottom-up retrofit approach and the use of building archetypes. As admitted by the authors, the generalization of the derived archetypes is not high. However, the results show that a full intervention in the building envelope could lead to a reduction of up to 50 % of the energy consumption. Two hundred sixty-four regional building archetypes were used in [22] to perform a cost-effectiveness analysis of three retrofit scenarios for the Portuguese building stock. Streicher et al., in [23], used 326 building archetypes and some pre-defined retrofit measures to perform a techno-economic analysis of large-scale deep energy retrofit in the Swiss residential building stock. In order to account for the timing aspect of building retrofit, Streicher et al. extended their methodology in [6] with a dynamic scenario-based approach. Ascione et al. followed a very similar approach in [24] with the use of representative building archetypes. A more complete overview over the works using building archetypes is given in [19]. However, the main disadvantage of using building archetypes is that, usually, they cannot capture the uniqueness of each building in the area of investigation, and thus, the results might deviate significantly from reality. There have been attempts to reduce this gap. For instance, Eggimann et al. [25] proposed an upscaling spatiotemporal error assessment methodology to derive building archetypes for various countries. However, most retrofit studies with the use of building archetypes, arbitrarily generalize from individual building-level to large-scale level, without taking into account upscaling uncertainties and errors.

It could be argued that the ideal bottom-up large-scale retrofit approach would be to perform a detailed retrofit analysis for each building by also accounting for potential district-level solutions. However, such a building-by-building approach [26], would lead to very high computational costs, especially for city- and national-level case studies. For example, Wu et al. [27] developed a constrained multi-objective large-scale retrofit approach that decides which buildings should be retrofitted first, when and how. However, the authors admit that such an approach lacks scalability; the larger the investigation area,

the higher the computational cost and time needed to collect the required data and to run the process. Indeed, as observed in [28], most of the existing building retrofit toolkits, based on simulation and optimization tools, suffer from time consuming data input and long simulation run time.

There have been attempts to reduce the computational cost involved in building-by-building large-scale retrofit. For example, Heo et al. [29] reduced the computational cost by using a quasi-steady-state energy demand model and a Bayesian calibration model to refine input parameters based on measured data. Jennings et al. [30] also rely on steady-state calculations and a single objective MILP optimization process to develop a retrofit tool for large-scale retrofitting.

1.4. Novelty and contribution

There is still a need for a scalable individual building-level retrofit model. Scalable individual building-level models should be able to handle a large number of buildings in a reasonable time. Thus, the main challenge in bottom-up large-scale retrofitting is to identify generalizable retrofit patterns to guarantee the scalability of individual building-level retrofit models to a larger building stock.

This paper aims to overcome this challenge with the proposed bottom-up retrofit approach, which is based on ML. A mixed building simulation and optimization approach is used to generate the necessary retrofit data used to train the surrogate retrofit model. With the use of such a scalable building-level surrogate model, we can benefit from: (a) a good balance between accuracy and computational cost at the building level, (b) reduced computational cost for large area case studies – even at a national level - and (c) a small set of required inputs which simplifies the application of the proposed approach by non-experts.

This paper is an extension of the work presented in [31]. In contrast to the previous work, the proposed surrogate retrofit model is capable to predict near-optimal retrofit solutions for whole Switzerland. To achieve the scalability of the model for whole Switzerland we included several climatic data, such as heating degree days (HDD) and the average solar radiation, as input to the model. Moreover, building archetypes were used for a twofold purpose. Firstly, to keep a good balance between accuracy and computational cost and secondly, to ensure that we cover a large share of the Swiss residential building stock. Finally, the proposed surrogate model has a much more complex structure than the one developed in [31]. This is necessary to address the complicated problem of identifying near-optimal retrofit solutions independent of the climatic region. The surrogate model consists of more artificial neural networks (ANNs) which makes the training process more challenging but improves the predictions. The proposed model is finally applied to all the residential buildings of the considered Geneva municipality in order to calculate near-optimal retrofit solutions.

1.5. Structure

The remainder of the paper is structured as follows. In Section 2, we describe in detail how we have obtained the necessary building and retrofit data to train the surrogate model and the case study that has been used to showcase the developed model. The process followed to build up and train the retrofit surrogate model is given in Section 3, while Section 4 is dedicated to presenting and discussing the results. More specifically, the validation and test results are provided and the retrofit analysis of applying the model to a large area case study, the Geneva municipality, is given. The advantages and the limitations of the proposed model are also summarized in this section. Finally, Section 5 concludes the paper with a summary of the presented work and ideas for future research.

2. Data collection and case study

Deriving building retrofit solutions is very dependent on collected

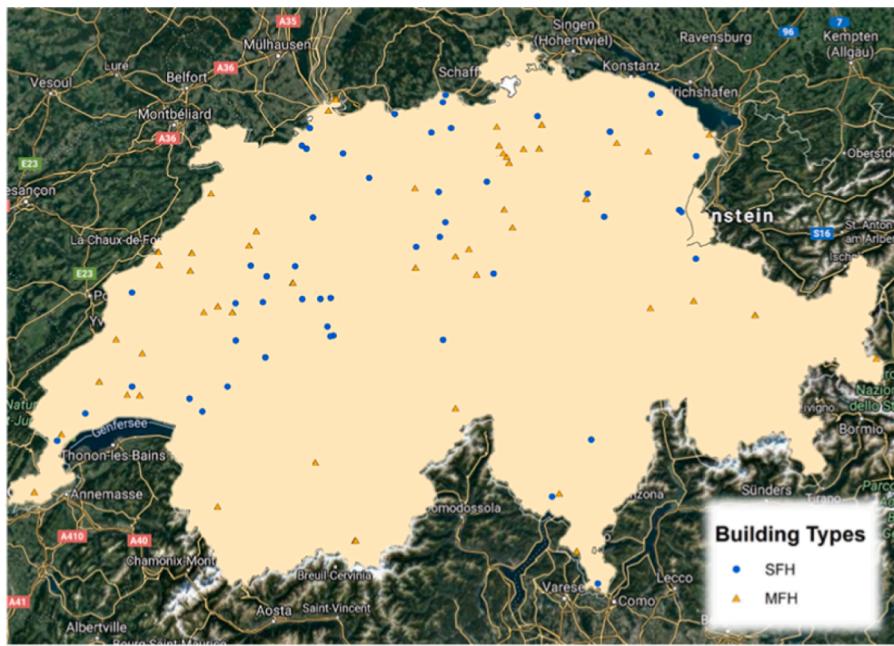


Fig. 1. Location of building archetypes (edited from [32]).

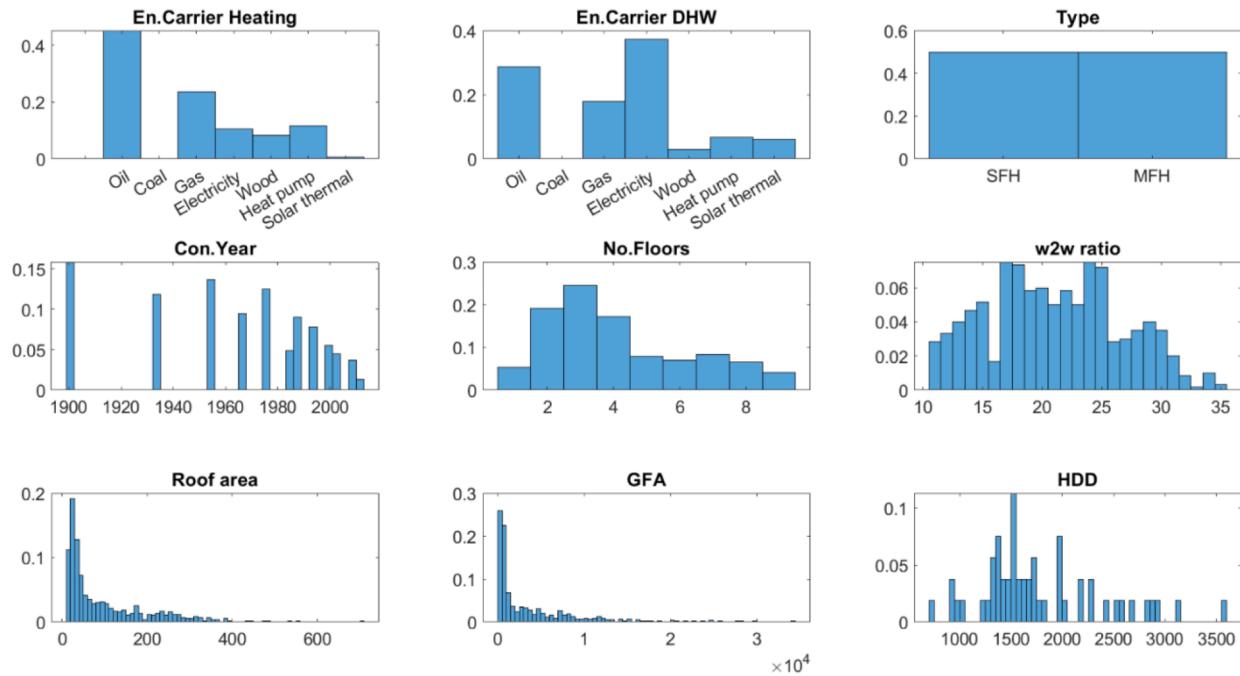


Fig. 2. Collected building information.

building data. Thus, one of the most important steps towards building up a surrogate retrofit model is to collect the necessary data. In this Section, the process of building data collection, the calculation of the respective retrofit data and finally the case study used to showcase the proposed approach are described in detail.

2.1. Data collection for building up the surrogate model

The data collection process is divided into three sub-tasks; the first is the selection of buildings and the respective building data, the second is the calculation of optimal building retrofit solutions, and the last step is

the selection of two buildings to validate the performance of the developed surrogate retrofit model.

2.1.1. Selection of building archetypes

One of the main characteristics of the developed surrogate model should be its applicability to the various climatic regions of Switzerland. To ensure this, three conditions have to be satisfied. Firstly, the selected buildings should be from all possible climatic regions in Switzerland. Secondly they should represent the whole Swiss residential building stock. Finally a trade-off has to be made between the data collection computational cost and calculating the respective retrofit solutions.

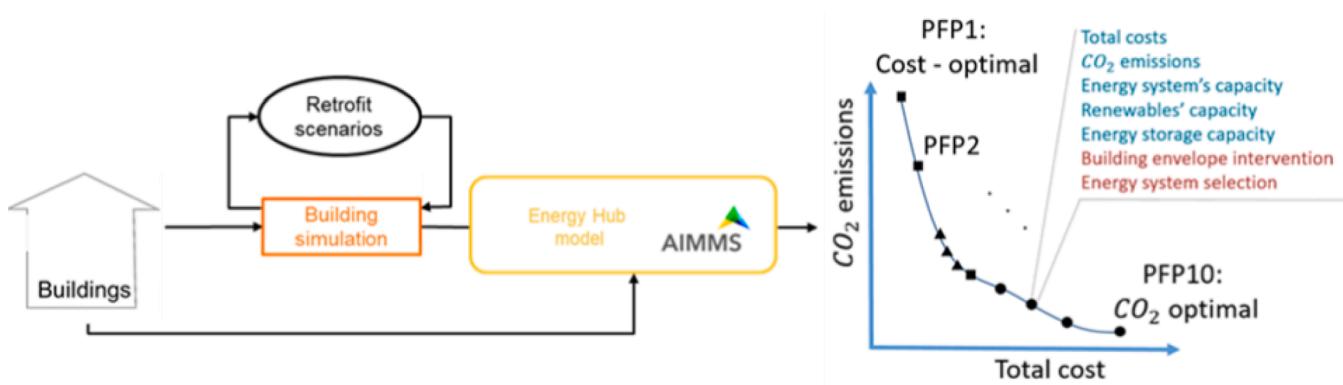


Fig. 3. Mixed building simulation and optimization model.

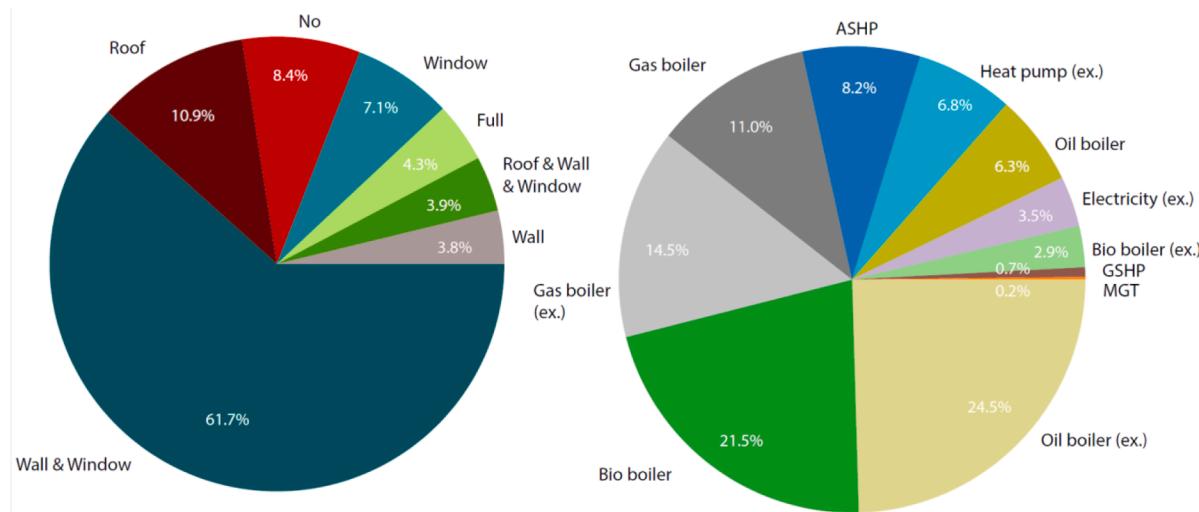


Fig. 4. Shares of building envelope interventions (left) and energy system selections (right) for all optimal retrofit solutions within the selected archetypes.

In order to satisfy those three conditions, we make use of building archetypes. More specifically, we used the building archetypes generated in [32]. A clustering algorithm was used to generate building archetypes of various building types based on building characteristics, such as the construction year, the energy demand and the building height. In this work, we focus on residential buildings, therefore 1000 archetypes, comprising of 500 single-family houses (SFH) and 500 multi-family houses (MFH) archetypes. In order to achieve a balance between accuracy and computational cost we use 600 of these residential building archetypes, with the same share of SFH and MFH as in the full set. Those 600 archetypes are meant to represent 92 % and 99 % of the Swiss residential building stock for SFH and MFH, respectively. The building archetypes are real buildings located in different places in Switzerland (see Fig. 1 for their geographical locations). In order to cover all the different climatic regions in Switzerland, we simulate the building archetypes for 53 climatic regions as defined in [33].

For the selected building archetypes we collect some building information that are easy to obtain or already known to the building owners or tenants. The building information is retrieved from a database provided by the Swiss federal statistical office [34]. The building information collected as well as their respective distribution are presented in Fig. 2. It can be seen that the majority of buildings is served by oil for space heating while domestic hot water is mainly provided with the use of electricity. Please note that for some buildings this has changed already, but is not included in the current version of the database. A large share of buildings are built before the year 2000 with the majority of them having between 2 and 4 floors. Moreover, in Fig. 2, the HDD of

the selected buildings are presented. The HDD, as defined by ASHRAE [35], are calculated with a heating threshold temperature of $T_{heating} = 12^\circ C$, and the formulas given as:

$$T_{a_{day}} = \frac{T_{max_{day}} + T_{min_{day}}}{2} \quad (1)$$

$$HDD = \sum_{day=1}^{365} \begin{cases} T_{base} - T_{a_{day}}, & T_{a_{day}} \geq T_{base} \\ 0, & T_{a_{day}} < T_{base} \end{cases} \quad (2)$$

2.1.2. Calculation of optimal retrofit solutions

The second step in the data collection process is to generate the optimal building retrofit solutions which will be used as target data to train the surrogate retrofit model. A mixed building simulation and optimization model depicted in Fig. 3, is used to generate the necessary target set for the selected buildings. All the necessary information and parameters used for building up the model can be found in [31], while more formulation details can be retrieved from [13].

The selected retrofit model consists of a building simulation tool, called CESAR-P [36], that is used to calculate the energy demand for different building envelope interventions. Those pertain to enhancements of the insulation of the walls or the roof, the replacement of windows with a lower U-Value and all potential combinations of these interventions. Full retrofit implies all the above mentioned interventions combined with the enhancement of the insulation of the ground floor. The generated demands are fed into a MILP optimization framework which generates for each building a Pareto front (PF), i.e., a set of

Table 1

Building input data of one single- and one multi-family house used for validating the prediction of the Pareto front.

Building type	Construction year class	Building height [m]	Floor area [m ²]	Roof area [m ²]	Window to wall ratio [%]	Energy Carrier (heating)	Energy Carrier (domestic hot water)
Single-family house	1979–1994	7.2	279.9	41.9	22	Oil	Oil
Multi-family house	1979–1994	14.4	726.5	108.9	26	Electricity	Electricity

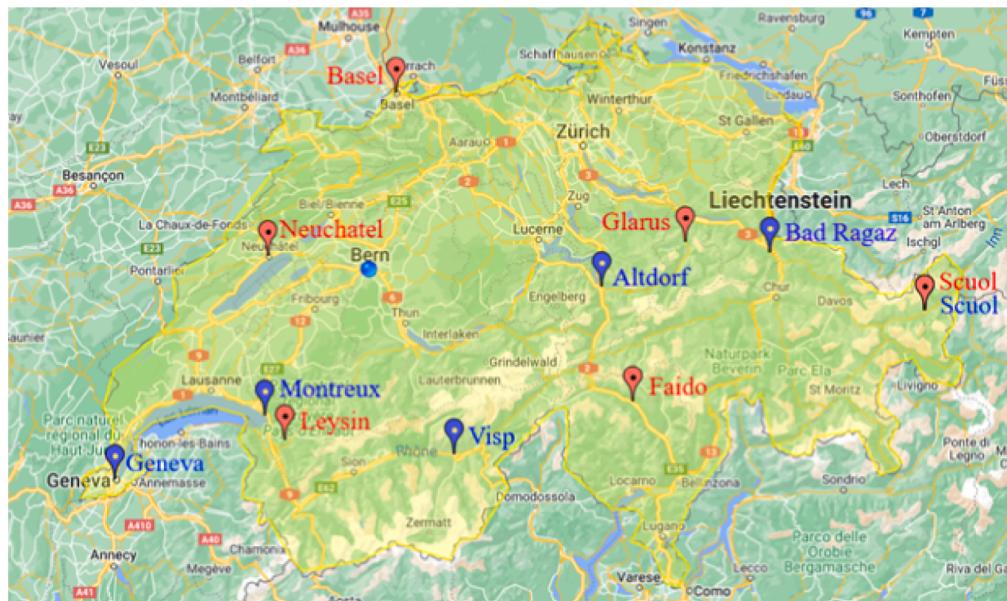


Fig. 5. Climatic regions considered for the validation of the two selected test buildings. The considered locations for the single-family house and the multi-family house are indicated with red and blue markers, respectively. [Generated with zeemaps.com].

optimal building retrofit solutions. The objectives are two, namely the minimization of the total cost and the CO₂ emissions. Ten optimal solutions are generated, from the cost optimal to the CO₂-optimal and some in between retrofit solutions. Each solution consists of nine retrofit dimensions: the objective values, the selection and the sizing of the energy system, the selection of the necessary building envelope intervention and the selection and sizing of renewables and storage. The considered renewable energy technologies are PV and solar thermal collectors while storage might be thermal and/or electrical. The energy system repository from which an energy system can be selected consists of boilers, served by oil, gas or biomass, air- or ground-source heat pumps, and micro gas turbines (MGTs). At this point, it should be mentioned that the already installed energy system might be a potential solution. For instance, it might be environmentally better to keep the already existing heat pump than installing a new heat pump or a biomass boiler. This can be explained from the fact that embodied emissions of systems and envelope interventions are taken into account. The lifetime of the existing energy system is considered to be equal to half of the respective new one.

The various building envelope interventions and the energy systems considered and their respective shares of all calculated retrofit solutions are depicted in Fig. 4. It can be seen that just above 60 % of the selected building envelope interventions for all PF solutions include both the enhancement of the insulation of the walls and the replacement of windows with more efficient ones. On the other hand, the enhancement of the insulation of the walls alone has the lowest share. Regarding the energy system selection, we observe that the use of the existing boilers, served by gas, oil or biomass, dominates with almost 42 % of the solutions. Following the existing boilers, biomass boilers, as a new system to

be installed, have a share of just above 21 % of all PF solutions. Finally, heat pumps have a share of 15.7 %.

2.2. Buildings selected for model validation

For the purpose of model validation for the developed surrogate model, we compare the results to the results from the mixed building simulation and optimization model for all the test buildings. More specifically, the PFs generated from the surrogate model and the mixed building simulation and optimization model are compared. We give selected results for one specific MFH and one specific SFH for some retrofit dimensions and for various climatic regions to showcase how the comparison is done.

The SFH was built some time between 1979 and 1994, has a height of 7.2 m and a total floor area of 279.9 m². For space heating, oil is used and thus the energy system installed is an oil boiler. The MFH is from the same construction year class, is 14.4 m high and has a total floor area of 726.5 m². A heat pump is used for heating and thus electricity is the energy carrier. The detailed building input data used are given in Table 1.

Since the surrogate retrofit model is capable of predicting near optimal retrofit solutions in all regions in Switzerland, we validate the accuracy of the PF for different climatic regions for those two test buildings. Specifically, we validate the surrogate model's test results for 6 weather climatic regions per building, namely for the regions shown on the map in Fig. 5. Red markers indicate the considered locations for the SFH, while the blue markers give the considered locations for the MFH. At this point it should be mentioned that those two buildings were selected because they provide a variety of building envelope

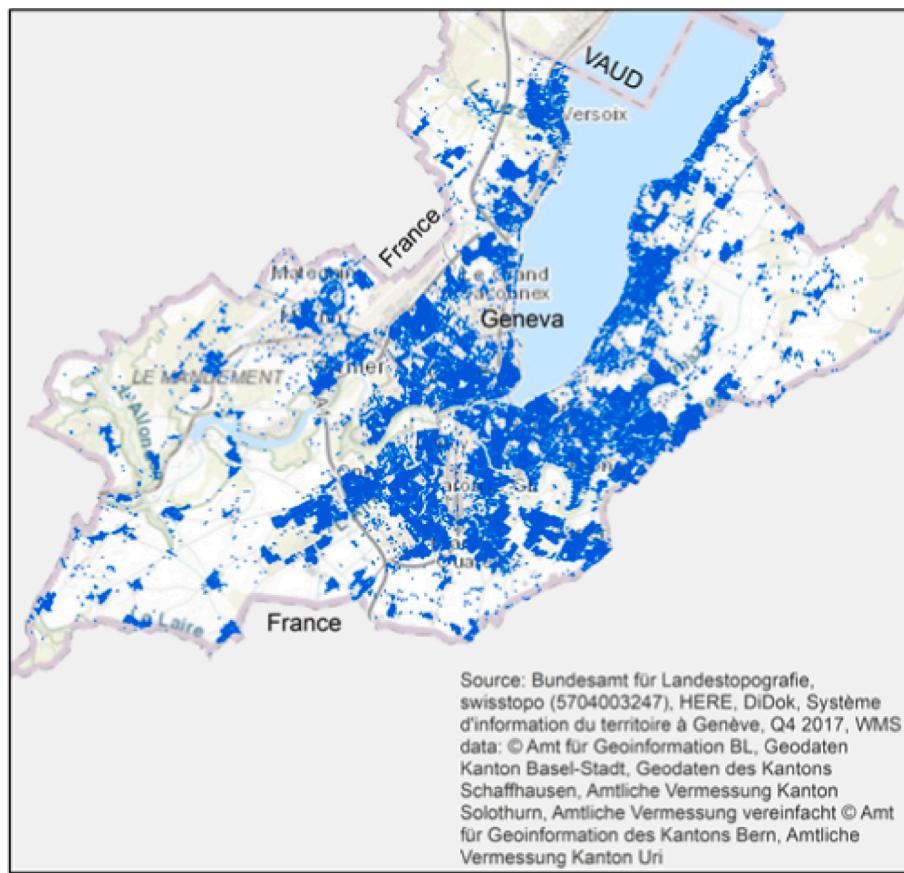


Fig. 6. Locations of the selected residential buildings to perform a large-scale retrofit analysis.

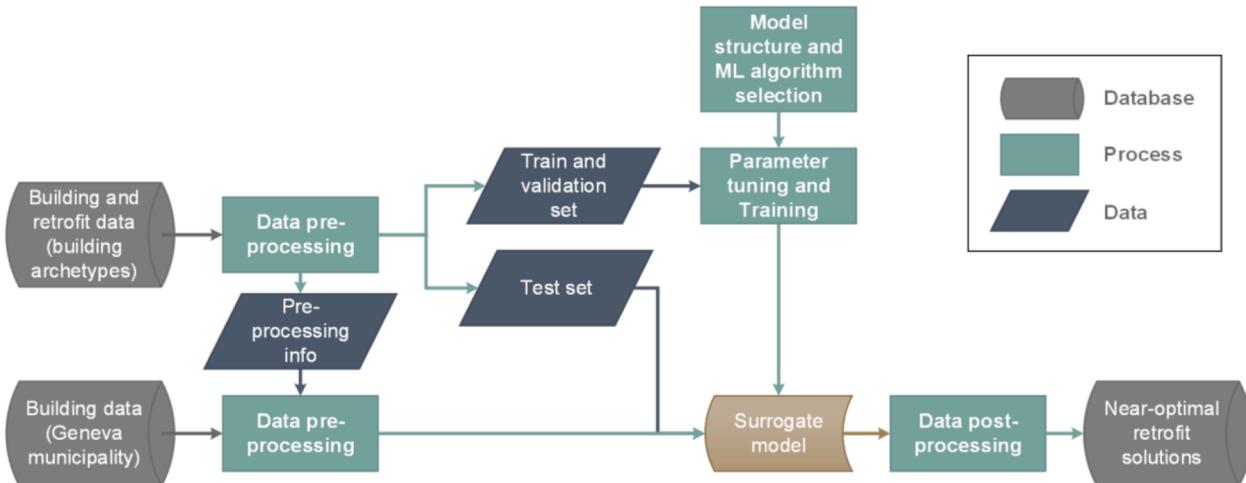


Fig. 7. Methodology for training, testing and validating the surrogate retrofit model.

interventions and energy system selections in the optimization results for the different regions. This makes the analysis of the performance of the surrogate model more interesting and meaningful.

2.3. Large-scale case study

One of the main advantages of the developed surrogate model is that it is scalable and once trained can be used to predict near-optimal retrofit solutions in large areas. To showcase this advantage, we have selected the Geneva municipality to calculate near-optimal solutions for

all residential buildings for which we were able to obtain building information. Out of the 43'881 residential buildings for which building data exist, we excluded the ones with wrong or inconsistent building data, i.e., no energy system or very low height. The final number of selected buildings is 36'639 for which the locations are depicted with blue dots in Fig. 6.

Three building databases were used to collect the necessary building information which serve as input to the surrogate retrofit model. More specifically, the Building and Apartment Registry (GWR) from the Swiss Federal Office for Statistics (BFS) for the years 2013 and 2017, as well as

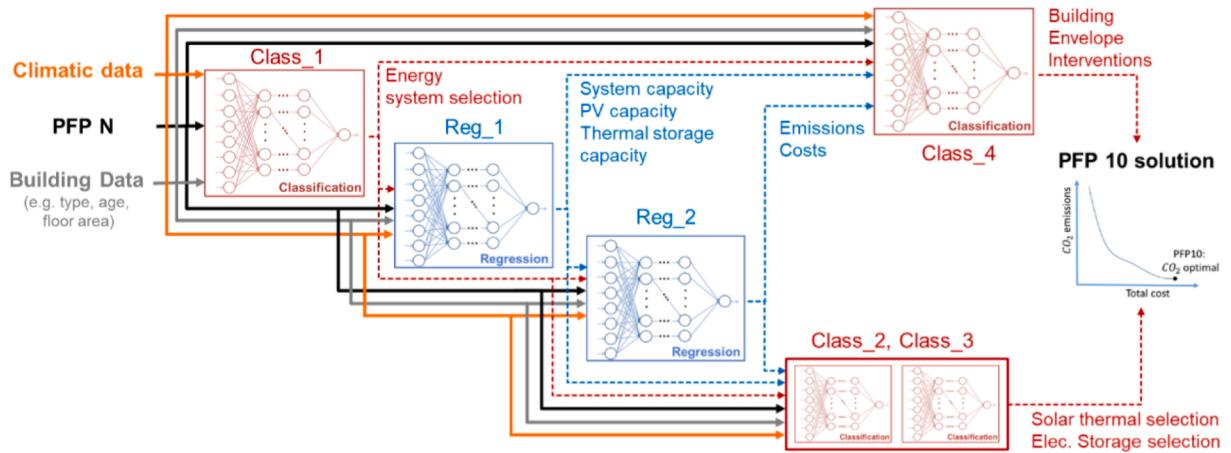


Fig. 8. First sub-model for predicting the CO₂ optimal solution, i.e. Pareto Front solution 10.

the 2.5D database from the Swiss Federal Office for Topography (swisstopo) were used. The two versions of GWR, i.e., the 2013 and the 2017 versions, were needed so as to ensure that we have both the latest available building information and the energy systems installed in the buildings. The 2.5D database is spatially joined with the GWR database so as to include the height of the buildings. The roof area is not provided in those databases and thus it was calculated separately. More specifically, we assume that the roof slopes of the selected residential buildings follow a uniform distribution and thus we sample from $U(0, 65^\circ)$ for each building's roof slope. The selected roof slope is then used to calculate the respective roof area, as:

$$\text{roof area} = \sec(\text{roof slope}) * \text{floor area} \quad (3)$$

The final roof area that can be used for PV installation is considered to be 60 % of the total roof area, as stated in [32].

3. Scalable building-level surrogate retrofit model

In this Section, the methodology followed to build up the scalable surrogate retrofit model is described in detail. The databases, the data and the processes used to develop the surrogate model are depicted in Fig. 7.

The building data of the archetypes and of the Geneva case study, both described in Section 2 are stored in databases. A data pre-processing step is necessary to ensure a reliable and smooth training and validation process. This data pre-processing step is followed by the definition of the model structure, the tuning of the ML parameters and the training process. Once the surrogate model is trained it can be applied to a set of buildings to predict near-optimal retrofit solutions right after a data post-processing step. In the following subsection, each of the processes is described in detail.

3.1. Data pre-processing

The data collected from the building archetypes and the Geneva municipality case study have to be pre-processed before used for generating the surrogate model. The pre-processing involves two steps. The first corresponds to the standardization of the building data so as to increase the accuracy of the trained ML models. The standardization of the input features pertains to a 25, 75 quantile normalization of continuous features and one-hot encoding of categorical features, as these have been performing the best in our studies. The second step of pre-processing refers to splitting up the input data of the building archetypes into a training, validation and testing set with 75 %, 15 % and 15 %, respectively. The training set is used for the learning process,

while the validation set is necessary to ensure a reliable learning process or in other words good generalization abilities to unseen buildings. The testing set, a set of unseen buildings during the training process, is used after the training process to validate the performance of the model. It should be mentioned that at least one representative building archetype from all potential climatic regions are included in the training and the testing set.

3.2. Model structure and ML algorithm selection

The developed surrogate model should be able to generate near-optimal retrofit solutions for residential buildings independent of their location in Switzerland. This implies that the target values are the nine retrofit dimensions for ten optimal retrofit solutions, resulting in a total of 90 variables, both continuous and categorical. To generate those target values, the main input features used are climatic data, e.g., HDD, mean temperature, and building data, e.g., building type, floor area, etc., as described in Section 2.1.

It should be mentioned that for some of the considered retrofit dimensions, namely the solar thermal collectors' capacity and the electric storage capacity, the prediction has to be transformed from a classification to a regression problem. The reason is that most of the generated optimal retrofit solutions do not suggest the use of solar thermal collectors and electric storage. Specifically, out of the 318'000 optimal retrofit solutions, 58 % suggest the use of solar thermal collectors, a percentage that falls to 28 %, if we account only the CO₂-optimal solutions predicted by the first sub-model. Concerning the selection of electric storage, the percentage is 15.9 % of all optimal retrofit solutions, and 1.9 % of all retrofit solutions except the CO₂-optimal ones. Since ML models perform poorly in predicting zero values, we transformed the problem of sizing those systems to the problem of deciding whether to install them or not which can be easier handled by a ML algorithm. In the few cases in which it is suggested to install a solar thermal collector or an electric energy storage, the MILP optimization framework is used to size the systems with all the other parameters being fixed.

The ML algorithm used for building up the surrogate model is feed-forward ANNs. There are mainly-two reasons for this choice. Firstly, ANNs generally perform well for both continuous and categorical data. Secondly, for this specific application they have been proven to work well [31] and outperform other ML algorithms, such as support vector machines and random forests [37].

The next, and maybe the most important task is to define the structure of the surrogate model. The structure of the surrogate model was derived from the target values. Those are the retrofit dimensions for each PF solution, a total of 90 variables, both continuous and

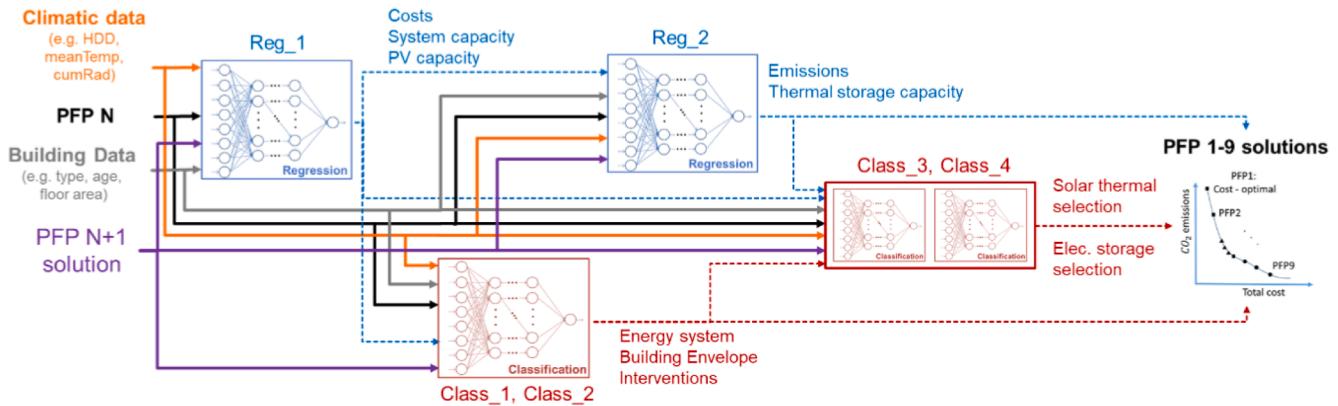


Fig. 9. Second sub-model for predicting the Pareto front solutions 1–9.

categorical. For the former, we use regressors (Regs) and for the latter we use classifiers (Class). Several steps were followed to obtain the structure which results in the best possible performance. Using the same structure of the surrogate building retrofit model for one specific climate region presented in [31] and training it to account for multiple climatic regions in Switzerland was not an option. The reason was that the performance of the model turned out to be poor. The next step is to predict each PF solution separately, or in other words, use one ANN which predicts the nine retrofit dimensions of one of the PF solutions. In total ten ANNs were used to predict the ten PF solutions. In our initial simulations, the best performing artificial neural network, by a significant margin, was the one predicting the CO₂-optimal retrofit solution. However, the overall performance, i.e. for the other PF solutions, was not satisfying. Before addressing this issue, we focus on the CO₂-optimal solution only and discuss how to predict each retrofit dimension for this solution. The best performing artificial neural network was the one predicting the selection of the energy system to be installed. This prediction was then used to enhance the input feature set to predict separately the rest of the retrofit dimensions. The ANNs performing well were fixed and the outputs are used as inputs to predict the rest of the retrofit dimensions. This process continues until all retrofit dimensions are predicted with a good accuracy. It should be mentioned that this process was done with randomized fixed number of layers and nodes. The rest of the hyper-parameters are optimized with grid search as soon as the structure of the sub-model is defined. The derived structure for the first sub-model predicting the CO₂-optimal solution is depicted in Fig. 8.

To improve the performance for the rest of the PF solutions, these are predicted with the use of a second sub-model. This model uses as input the output of the first sub-model that is the retrofit dimensions of the CO₂-optimal solution. In order to define the structure of the second sub-model, we follow a similar approach as with the first sub-model resulting in the structure depicted in Fig. 9. It consists of two Regs for predicting the continuous variables and four Class for each of the categorical variables. Each PF solution is predicted sequentially with the previously obtained PF solution used as input. For instance, in order to predict PF solution 7, we use as input also the PF solution 8.

3.3. Parameter tuning and training

ANNs are very powerful machines, yet they have to be trained properly so as to take advantage of their full potential. In order to ensure a reliable training process we have to properly select the ANNs' hyper parameters, the training parameters as well as the evaluation metrics.

With the structure of the two sub-models fixed, we need to tune the ANNs' hyper parameters so as to achieve the best possible performance. Trial and error with the use of grid search is used to derive the activation function, the number of layers and the nodes per layer. The rectified

linear unit function, in short ReLU, was used for the hidden layers, while the linear and the softmax functions were used for the Regs and the Class, respectively. The reason of selecting those activation functions is that they have been proven effective for a related problem [31]. Regarding the number of layers and nodes per layer, we have tested up to five layers with many different combinations of nodes per layer. The selected and best performing number of layers and nodes for all the ANNs are given in Section 4.1.2.

Fixing the training parameters is the next step in order to train the ANNs. The optimizer was selected to be Adamax, similar to the one used in [31], and the batch size was optimized for the Regs to 650 and for the Class to 100. The number of epochs was fixed to 2500. In order to avoid overfitting, we make use of L2 regularization and early stopping. The L2 regularization value is optimized for each ANN with grid search.

Each artificial neural network is trained sequentially by starting with the Class of the first sub model, or else by predicting the selection of the most environmentally friendly energy system. It should be mentioned that the ANNs of the second sub model are trained at once for all PF solutions since the target values are already known. The reason of performing the training at once and not re-training the existing ANNs every time with the new PF solution is to avoid the commonly named catastrophic forgetting of the ANNs.

ANNs training might be prone to bias due to the random initialization of some of the parameters at the start of the training process. To eliminate this bias and estimate the mean model performance we use k-fold cross validation. In this work, k was selected to be 5, as a good practice in ML. The 5-fold cross validation was repeated twice for each ANN to improve the estimate of the model's mean performance. This implies that each ANN was trained 10 times with the final performance being the mean performance of all those trials.

In order to evaluate the performance of both the training and the validation set we use the coefficient of determination (R^2). On the other hand, for the test set we use two different evaluation metrics, the R^2 calculated in the whole dataset and the per-building coefficient of determination $R_{building}^2$ as given by:

$$R^2 = 1 - \sum_{i=1}^n (y_i - \hat{y}_i) / \sum_{i=1}^n (y_i - \bar{y})^2 \quad (4)$$

$$R_{building}^2 = \sum_{i=1}^{n/10} \left(1 - \sum_{p=1}^{10} (y_{ip} - \hat{y}_{ip}) / \sum_{p=1}^{10} (y_{ip} - \bar{y}_i)^2 \right) \quad (5)$$

where n denotes the number of samples and p is the PF number. The main reason of not using $R_{building}^2$ for the training and validation set is the structure of the surrogate model. Since the CO₂-optimal solution is predicted separately from the other PF solutions it would be impossible

Table 2

Selection method and values of the hyper-parameters used to train the surrogate retrofit model.

Parameter	Value(s)	Selection method
Hidden layers/ nodes	1 / 400,300,200,100,50,25 2 / [400,200],[300,100],[200,100],[150,50], [100,25],[50,20] 3 / [4000,2000,500],[2000,1200,250], [1000,700,250],[700,250,60],[500,300,50], [100,50,25],[50,20,10] 4 / [2200,1200,750,250],[1000,700,500,250], [1200,700,250,25],[700,500,250,60], [500,300,150,50] 5 / [3000,2000,1200,250,50],[2200, 1000,700,500,250],[1000,700,500,250,60]	Grid search
Activation functions	ReLU (hidden layers) and Linear (regression), soft-max (classification)	Fixed
Cost functions	L2 loss (regression), categorical cross entropy & class weighting (classification)	Fixed
Regularization	$\lambda = [0.1, 0.01, 0.001, 0.0001]$	Grid search
Early stopping (patience)	50 epochs	Fixed
Evaluation metrics	Coefficient of determination R^2 (regression), f1 score (classification)	Fixed
Training function	Adamax (parameters as in [38])	Fixed
Epochs	4500 (regression), 2500 (classification)	Fixed
Batch size	50,100,300,600,1000,1500,2000	Grid search

Table 3

Grid search results for defining the hyper parameters of the surrogate model.

Parameter	Final Value	Sub-model / Artificial neural network
Hidden layers/ nodes	2 / [400,200] 2 / [100,50] 2 / [750,300] 4 / [1000,700,500,250] 5 / [3000,2000,1200,250,50] 2 / [750,120] 2 / [150,50] 2 / [1200,250] 2 / [100,25]	1 / classifier 1 1 / classifier 2–3 1 / classifier 4 1 / regressor 1 1 / regressor 2 2 / classifier 1–2 2 / classifier 3–4 2 / regressor 1 2 / regressor 2 1,2 / All regressors 1,2 / All classifiers 1,2 / All regressors 1,2 / All classifiers
Regularization	0.1 0.01	
Batch size	650 100	

to use this metric during training and validation. Moreover, this metric as a loss function would increase drastically the training time and thus make the learning process unfeasible. More specifically, it was calculated that this metric would lead to an increase of the training time from a couple of hours to days.

A summary of the hyper parameters selected for the training process are summarized in Table 2. The selection method is either fixed or performing a grid search to define the optimal selection based on the highest validation performance.

3.4. Data post-processing

The training of the surrogate model is followed by some post-processing steps to obtain the near-optimal PF. Since it is not realistic to have a thermal energy storage unit with a capacity of less than 1 kW we set all the output capacities below that value to be zero. Moreover, we can restrict the installed capacity of PV and/or solar thermal collectors to the maximum available roof area which is one of the input features to the model. It should also be mentioned that the capacity of the solar thermal collectors and the PV are given in m^2 .

Another important step of datapost-processing is de-standardization. It involves de-normalization and inverse encoding of the continuous and the categorical variables, respectively. The standardization information stored in the pre-processing step is used in order to obtain the actual values for the retrofit dimensions. The final step of post-processing is the application of a Pareto dominant filter to make sure that the surrogate model's solutions are Pareto dominant points. The filtering process is described in:

$$\begin{aligned} \text{A PFP } x_i \text{ dominates a PFP } x_j, \text{ with } i \\ \neq j, \text{ if and only if, for each objective } k = 1, 2 \quad f_k(x_i) < f_k(x_j) \end{aligned} \quad (6)$$

The filtering process might remove accurate predictions which are not Pareto dominant solutions. Therefore the evaluation of the model performance takes place before applying the Pareto dominant filter. However, for the PF validation process for the two test buildings the predicted non-dominant points have been removed.

4. Results and discussion

The prediction performance of the surrogate model at building level and the application of this model in a large area case study are presented and discussed in this section.

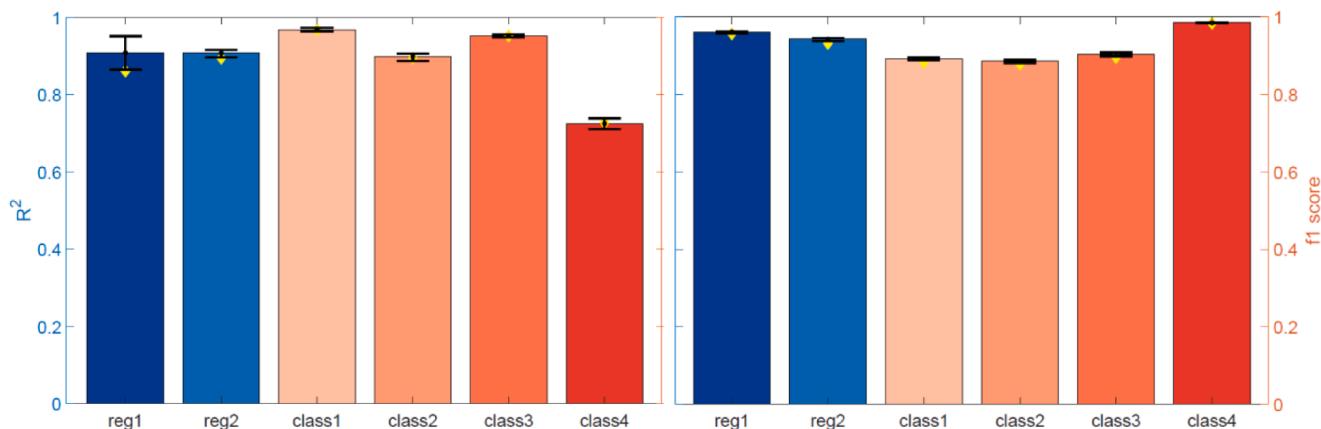


Fig. 10. Prediction performance of the first (left) and the second (right) sub-model of the building retrofit surrogate model. The error bars represent the twice repeated 5-fold cross validation errors and the yellow markers represent the performance on the test set. The retrofit dimensions predicted by each regressor or classifier can be seen in Fig. 8 and Fig. 9. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

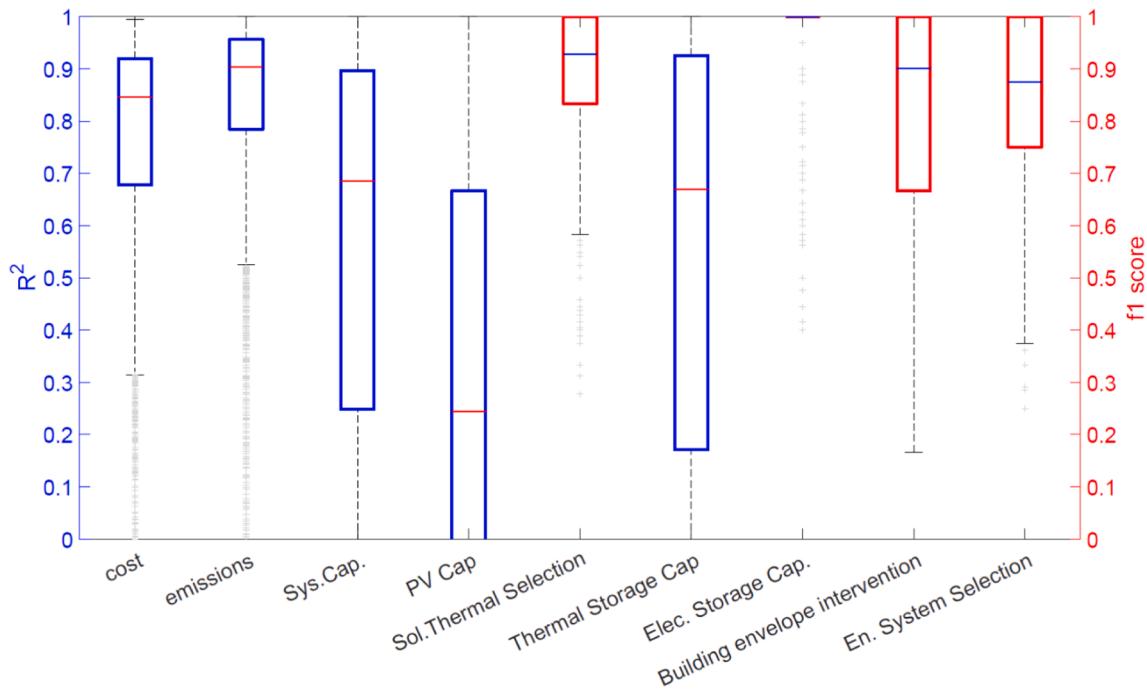


Fig. 11. Performance of the surrogate model with the use of the per-building coefficient of determination (metric 2).

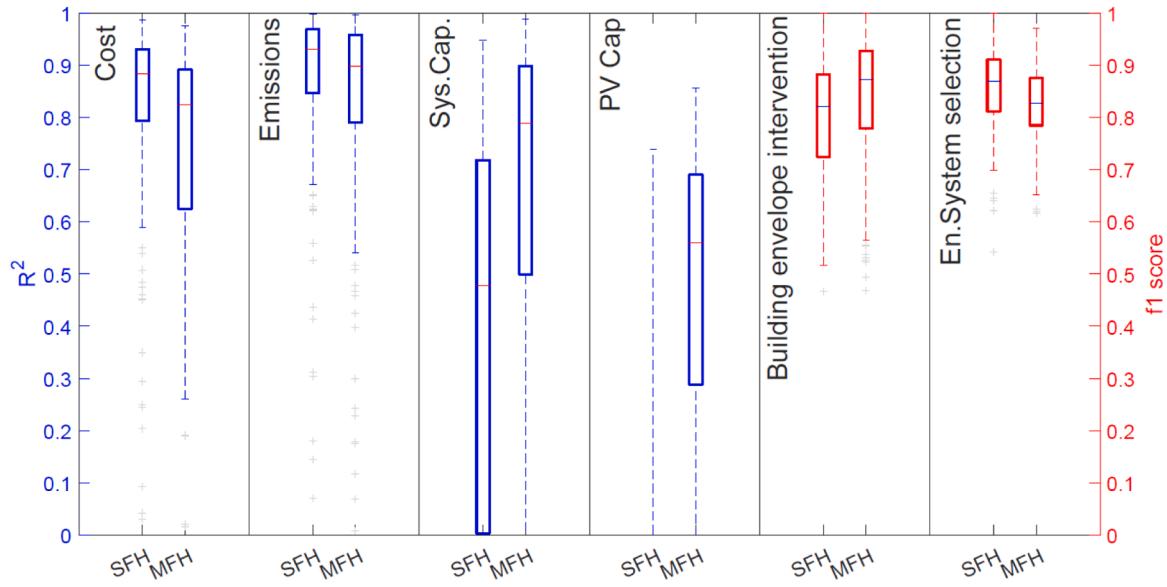


Fig. 12. Performance of the surrogate model (metric 2) per building type.

4.1. Building-level surrogate model performance

The surrogate model was developed to derive building-level near optimal retrofit solutions. The model's training performance is analyzed in detail in this subsection. More specifically, the validation and the test results are summarized and discussed.

4.1.1. Surrogate model's hyper-parameters

Parameters should be selected such that the surrogate model achieves the highest possible validation performance during training. As discussed in subsection 3.2, some parameters were fixed based on previous knowledge while others were selected with the use of grid search. In Table 3, the grid search results for each of the ANNs in the two sub-

models are presented. It is interesting to observe that the Regs of the first sub-model have a much higher capacity compared to the Regs of the second sub-model. For instance, Reg2 was selected to have 5 hidden layers which results in a set of training parameters, i.e., connection weights, of just below 8.8 million.

4.1.2. Validation and test performance

A two times 5-fold cross validation was used to better estimate the mean performance of the trained ANNs. In Fig. 10, the calculated validation and test errors are presented. The left bar graph represents the results of the first sub-model, which is the one predicting the CO₂-optimal solution. The right bar graph depicts the results of the second sub-model that predicts the rest of the PF solutions. The error bars

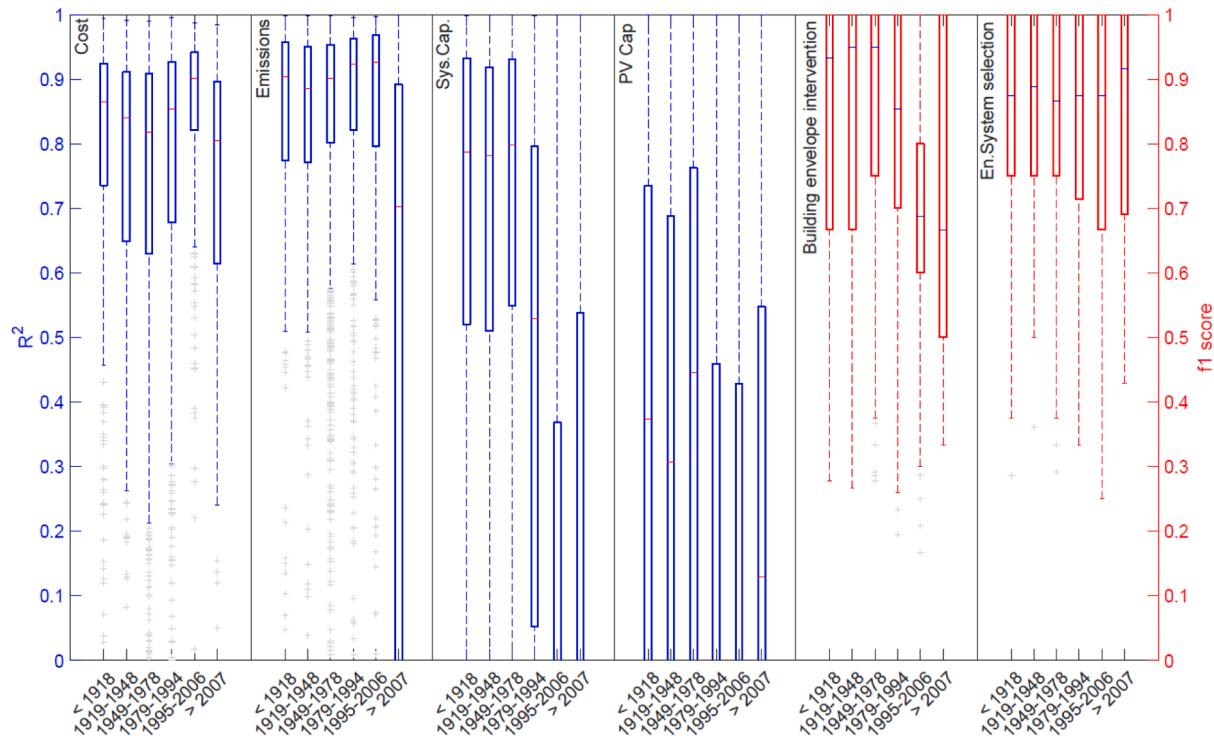


Fig. 13. Model performance based on construction year class.

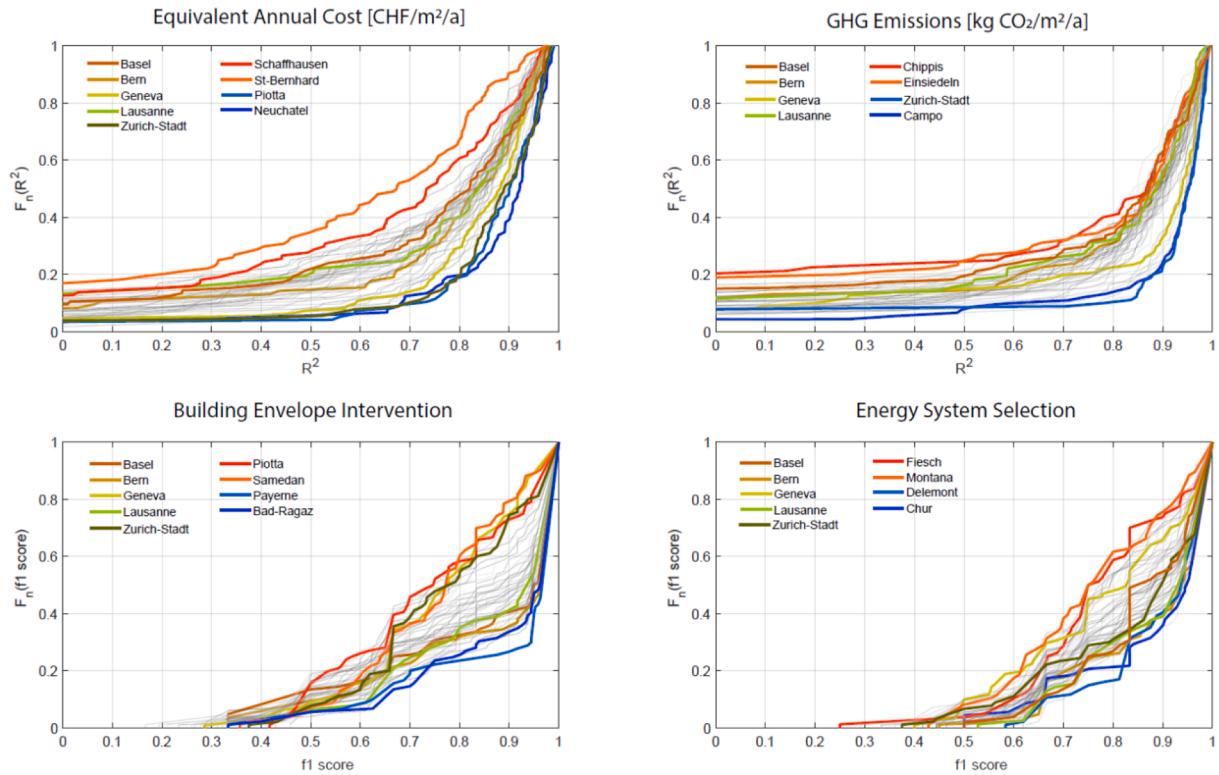


Fig. 14. Prediction performance based on climatic region based on the total cost, CO₂ emissions, the building envelope interventions and the energy systems selection (Starting from the top left and moving to the bottom right).

represent the validation error coming from the repeated 5-fold cross validation, while the yellow markers represent the test results. To ensure that an acceptable estimate of the performance of the ANNs is achieved,

these markers should lie within the range of the error bars. The categorical variables are colored in red and are evaluated with the f1 score, while the continuous variables, evaluated with the R^2 , are depicted in

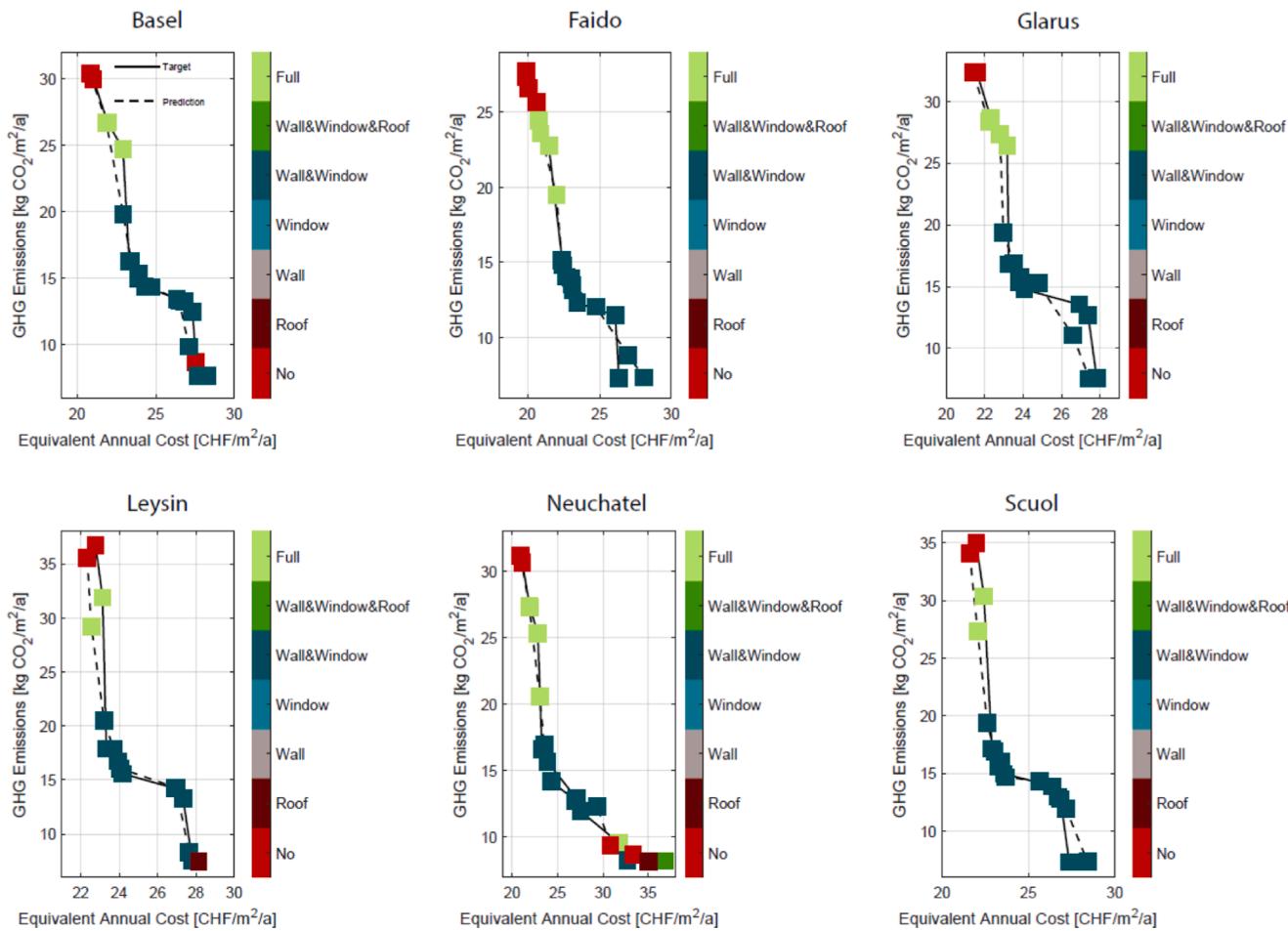


Fig. 15. Validation results for the building envelope intervention retrofit dimension for a single-family house for different climatic regions.

blue.

All the ANNs of the first sub-model achieve a very good performance with values above 0.85 for both R^2 and the f1 score. The only exception is the prediction of the building envelope intervention, i.e., Class 4, for which the performance is around 0.71. Reg 1, which is used to select the energy system, the PV and the solar thermal collectors, has, in contrast to all other ANNs, the highest validation error range. The Class test errors lie within the validation error bars, while the Regs test errors lie close to the lower error bar. The second sub-model's ANNs achieve a very good performance, namely above 0.88 for all predicted values. The range of the validation error is essentially negligible which implies a good approximation of the model performance.

To highlight the surrogate model's performance we compare the predicted PFs of two test buildings to the ones generated from the building simulation and optimization model. Additionally, to provide an overall picture on how well the surrogate model performs per test building we use $R^2_{building}$, as given in (5).

The overall per building- and per retrofit dimension performance is depicted in the boxplot chart in Fig. 11. The color pattern is the same as in Fig. 10. The box plots represent the 25 and 75 quantiles, and the inside line is the median, while the outliers are depicted with grey markers. The decision on installing or not installing an electric storage system is the best performing retrofit dimension with an f1 score approaching 1. The prediction of the objective values, the selection of a solar thermal collector, the building envelope intervention, and the energy system selection are also performing very well with medians above 0.85 for both the $R^2_{building}$ and f1 score. The prediction of the systems' capacities and the thermal energy storages' capacity have

medians which are just below 0.7, but the 25th percentiles fall below 0.3. The worst performing retrofit dimension is the sizing of the PV. However, the 75th percentile still reaches 0.68.

To investigate why the prediction of the PV capacity is not as high as expected we check the performance of the surrogate model per building type. It can be seen, in Fig. 12, that the surrogate model fails to predict PV capacity for SFH, while it performs considerably good for MFH with a median of around 0.58. Apparently, the majority of the solutions with a $R^2_{building}$ of less than 0.5 were SFHs having roof areas of less than 60 m^2 . Hence, as a conclusion, the surrogate model could not learn how to approximate the capacity of the installed PV given a building with a small roof area.

In Fig. 12, we also observe that the prediction of the energy system's capacity is much more accurate for MFH than SFH. The same pattern applies for the building envelope interventions but with a much smaller difference between the two building types. It should also be mentioned that the surrogate model predicts the building envelope interventions with a very good accuracy of more than 0.82 for both building types. In all other retrofit dimensions, the predictions for the SFH are slightly better compared to the MFH.

Interesting results can also be extracted from the model's performance per building construction year class. In Fig. 13, the prediction performances for 6 different construction year classes are depicted.

For all construction year classes the surrogate model predicts very accurately the objective values, i.e., costs and emissions, with accuracies greater than 0.8. The only exception is the prediction of the emissions for buildings constructed after the year 2007, for which the $R^2_{building}$ is 0.7. A similar pattern, but to a different extent, is observed for the energy

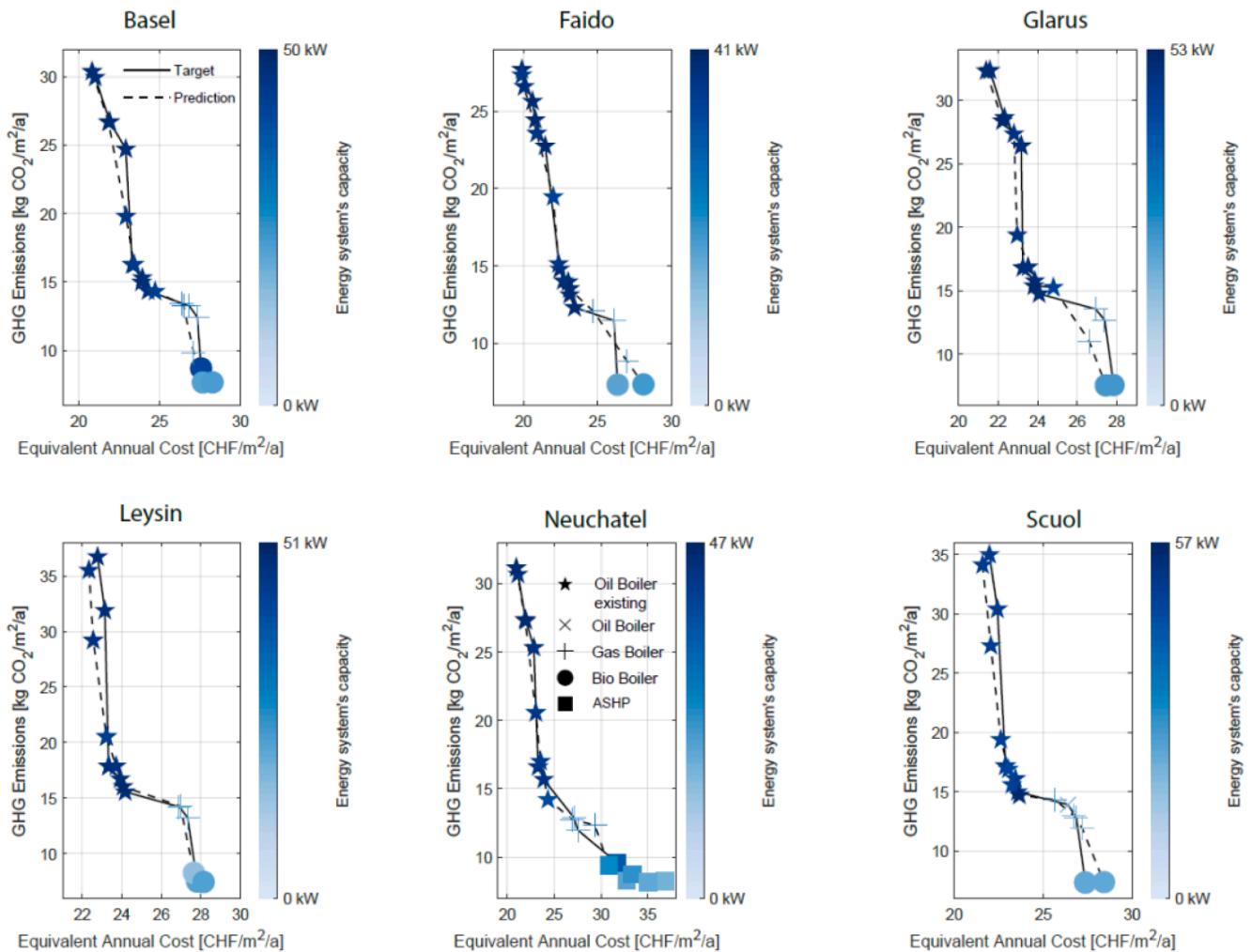


Fig. 16. Validation results for the energy system selection and sizing retrofit dimensions for a single-family house for different climatic regions.

system's capacity prediction and the building envelope interventions. The surrogate model cannot accurately predict the system capacity for buildings built after the year 1995. The PV sizing is more accurate for buildings built before 1978, while the prediction performance of the energy system selection is similar for all construction year classes.

Finally, we compute the model's performance for four retrofit dimensions, i.e., the objective values, the building envelope interventions, and the energy system selection, per climatic region. For better representing those four retrofit dimensions we generate the empirical cumulative distribution function (ecdf) of $R^2_{building}$ for the 53 climatic regions. The ecdfs are depicted in Fig. 14. The ecdf of the most-populated regions in Switzerland, i.e., Basel, Bern, Geneva, Lausanne and Zurich-city, the two worst and the two best performing regions are highlighted, while the rest are colored in grey. On the x axis we have the evaluation metric for each retrofit dimension and on the y axis the prediction performance chance.

With respect to the cost retrofit dimension, the surrogate model performs the worst for buildings located in Schaffhausen and St-Bernhard region. On the other hand, it performs the best for buildings located in Neuchatel and Piotta. More specifically, there is a 47 % chance that the $R^2_{building}$ of predicting the retrofitting cost for a building located in St-Bernard is higher than 0.7. This percentage reaches 90 % for a building located in Neuchatel or the city of Zurich. As far as the prediction of the emissions is concerned, the worst performance is observed for buildings located in Einsiedeln and Chippis with a 69 %

chance that the $R^2_{building}$ is higher than 0.7. The best performance is achieved for buildings in the city of Zurich and Campo region, located in the Italian-speaking part of Switzerland, with 91 % and 89 % chance that the $R^2_{building}$ is higher than 0.7, respectively.

Two regions in the southern part of Switzerland, Piotta and Samedan, are the worst performing in terms of predicting the building envelope intervention. The best performing regions are from central Switzerland, namely Payerne and Bad-Ragaz, with a 77 % chance that the f1 score is higher than 0.75. The surrogate model's prediction performance on buildings located in Geneva and Zurich is worse than the prediction performance of buildings located in Lausanne, Bern, and Basel. On the other hand, buildings located in the biggest cities are performing very well in terms of energy system selection with a chance of more than 76 % that the f1 score is higher than 0.7. The worst performing region is that of Geneva, with the respective chance being around 68 %. Prediction performance of buildings located in the canton of Valais, more specifically in Montana and Fiesch, is the worst in terms of energy system selection with 39 % and 42 % chance of having an f1 score higher than 0.8, respectively. The best f1 score is achieved with a chance of 78 % and 84 % for Delemont and Chur regions, respectively.

At this point, it should be mentioned that we could not observe any clear correlation between the climatic regions and the model's prediction performance. This implies that the model's prediction performance is not directly affected by the climate conditions of the building's location. The coldest climatic regions are Saas-Fee and Samedan, with

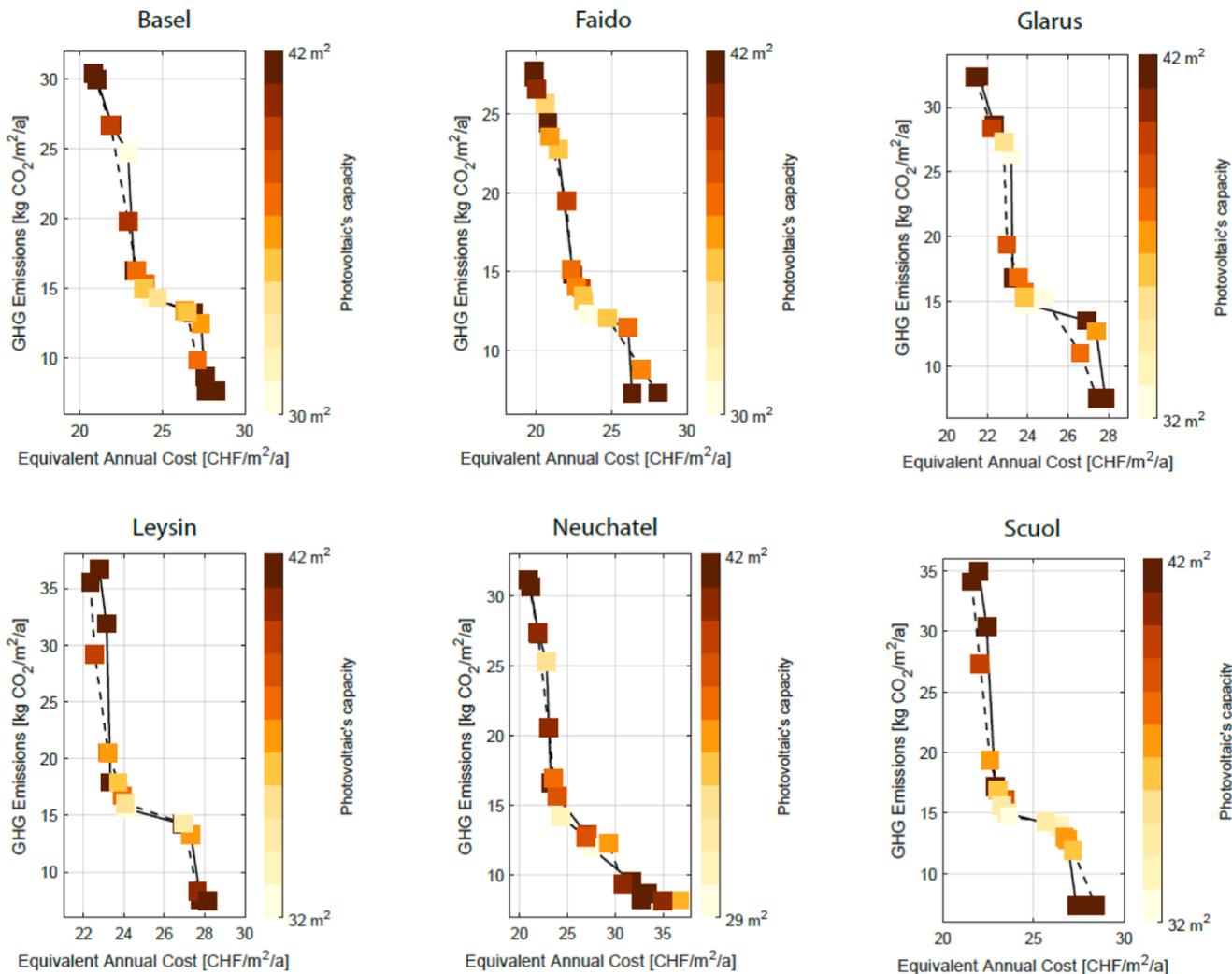


Fig. 17. Validation results for the sizing of the photovoltaics for a single-family house for different climatic regions.

HDD of 3148 and 3562, respectively, while the warmest are Cresciano and St.-Bernhard region with HDD of 941 and 700, respectively. The model's performance for buildings located in the coldest regions is the highest in terms of predicting the emissions and the lowest in terms of predicting the building envelope interventions. In terms of predicting the energy system selection, buildings located in the warmest region, i.e., St-Bernhard region, are one of the best performing ones with a 73 % chance that the f1 score is higher than 0.8. On the other hand, buildings located in the second warmest region, i.e., Cresciano, are the worst performing with a 46 % chance that the f1 score is higher than 0.8.

4.1.3. Validation on a test set of buildings

In this section, we present the validation results of the PF predictions of two test buildings, a specific SFH and a specific MFH, with the building characteristics described in Section 2.2. The PF predictions are validated for four retrofit dimensions, i.e., the building envelope interventions, the energy system selections and the PV selection and capacity.

In Fig. 15, a comparison between the surrogate model's and the mixed building simulation and optimization model's PFs, in terms of the building envelope intervention dimension, for the SFH is depicted. The comparison is performed for six different climatic regions. The axes represent the two objectives, namely the costs and the emissions. The solid lines are the PF lines generated from the mixed building simulation and optimization model while the dashed lines are the PF lines

generated from the surrogate model. The PF pattern is predicted with an acceptable accuracy. More specifically, the lowest average deviation in terms of cost is 0.31 CHF/m²/a and is observed for the Glarus region, while the highest one is 0.65 CHF/m²/a and is observed for the Neuchatel region. In terms of emissions, the lowest average deviation is 0.79 kg CO₂/m²/a for the Basel region, while the highest is 1.55 kg CO₂/m²/a for the Leysin region.

The marker colors represent the various building envelope interventions. A perfect matching in selecting the building envelope intervention is achieved for Faido, Glarus, and Scuol regions. For the Basel region there is one mismatch on the ninth PF. The target solution is no building envelope intervention, while the surrogate model predicts an enhancement of the insulation of the walls and the replacement of windows. One mismatch is also observed for the Leysin region for the CO₂-optimal solution. The worst performing climatic region is Neuchatel, for which there is a mismatch on the three CO₂-least emitted solutions. Overall, we can claim that we achieve a very good matching of selecting the necessary building envelope measure for all climatic regions and PF solutions.

A similar plot from another retrofit dimension, i.e., the energy system selection and sizing, is presented in Fig. 16. The markers represent the energy systems installed and the marker colors depict the capacity of the systems. For all the climatic regions, the dominantly selected energy systems are the existing oil boiler, a new gas boiler and a biomass boiler. The only exception is Neuchatel region for which an ASHP is selected as

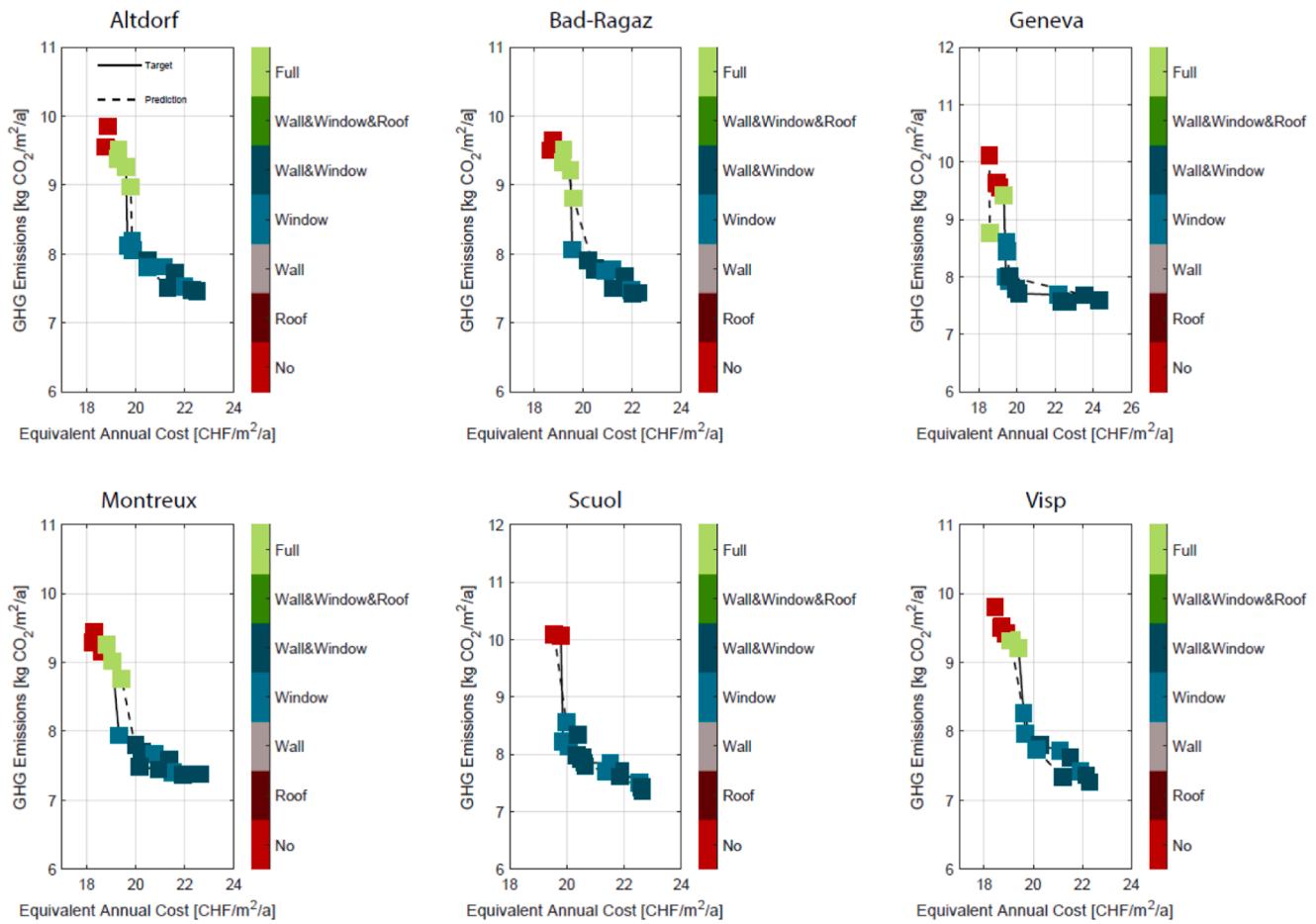


Fig. 18. Validation results for the building envelope intervention retrofit dimension for a multi-family house for different climatic regions.

the CO₂-optimal solution which is perfectly predicted by the surrogate model. A new oil boiler is selected in only one PF solution, the seventh, for the Scuol region. However, the surrogate model for this specific Pareto-front solution suggests to keep the existing oil boiler. A perfect matching for all the PF solutions is observed for Glarus and Neuchatel.

As far as the energy system sizing is concerned, the surrogate model is capable of sizing the selected systems quite accurately, with the cost-optimal solutions heated by larger systems than the CO₂-optimal solutions. For around 65 % of the calculated solutions the surrogate model underestimates the sizing of the system. The worst predictions are observed for Leysin with an average deviation for all PF solutions of 3.6 kW. In contrast, the best performing region for this building is Scuol with an average deviation of just above 0.8 kW.

To ensure that the predicted set of solutions constitute a PF we need to pass the solutions through the Pareto dominant filter. The filter eliminates one solution from both the Glarus and Leysin climatic region results. This implies that all predicted PFs for this building have ten dominant solutions, as is the case in the target PFs, except for those two regions for which the PFs consist of nine dominant solutions.

In Fig. 17, one more retrofit dimension, namely the sizing of the PV on the rooftop, is validated. The color map represents the size of the PV that is suggested to be installed on the rooftop of the SFH. The maximum available area for PV is just below 42 m² and the lowest area covered with PV is 29 m² for the Neuchatel region. The highest deviation in sizing the PV is observed for PF number 3 with a deviation of 9 m², while a perfect matching is observed for all cost-optimal solutions and all CO₂-optimal solutions except of the Neuchatel region which has a deviation of 5 m². The region with an almost zero average deviation in the PF is the Neuchatel region. On the other hand, the highest average deviation

is 2.6 m² for the Scuol region.

A similar procedure is followed for the MFH with the building information described in detail in Section 2.2. In Fig. 18, the PFs comparison for the building envelope dimension is presented. The range of the objective values is much smaller compared to the SFH. The building envelope interventions selected for all climatic regions are limited to no intervention, full intervention, replacement of windows, and enhancement of the insulation of the wall combined with the replacement of windows.

Similarly to the SFH, the PF trend is approximated quite well. This can be proven by the fact that the highest average deviations of the cost and emissions objective value is 0.79 CHF/m²/a and 0.43 kg CO₂/m²/a, respectively, both for the Geneva region. The lowest average deviation in terms of cost is 0.16 CHF/m²/a, while the lowest average emission deviation is the 0.17 kg CO₂/m²/a, both observed for the Scuol region. A perfect matching of building envelope interventions is observed for the climatic regions Altdorf and Bad-Ragaz. A mismatch in one PF solution is observed for the Geneva and the Montreux region. A two PF solution mismatch is observed for the other two regions, i.e., Scuol and Visp. The worst mismatch is the one of the PF point 2 for the Geneva region. More specifically, the target implies no building envelope intervention while the prediction implies a full building envelope intervention.

The validation of the PFs for the MFH in terms of energy system selection and sizing is depicted in Fig. 19. The energy systems involved in the PFs are the existing heat pump for the cost-optimal solutions, an ASHP, and a biomass boiler for almost all the CO₂-optimal solutions, except for the Geneva region. A perfect matching of the energy system selections is observed for the Scuol region. For the rest of the climatic regions, we have one mismatch per region, namely with the prediction

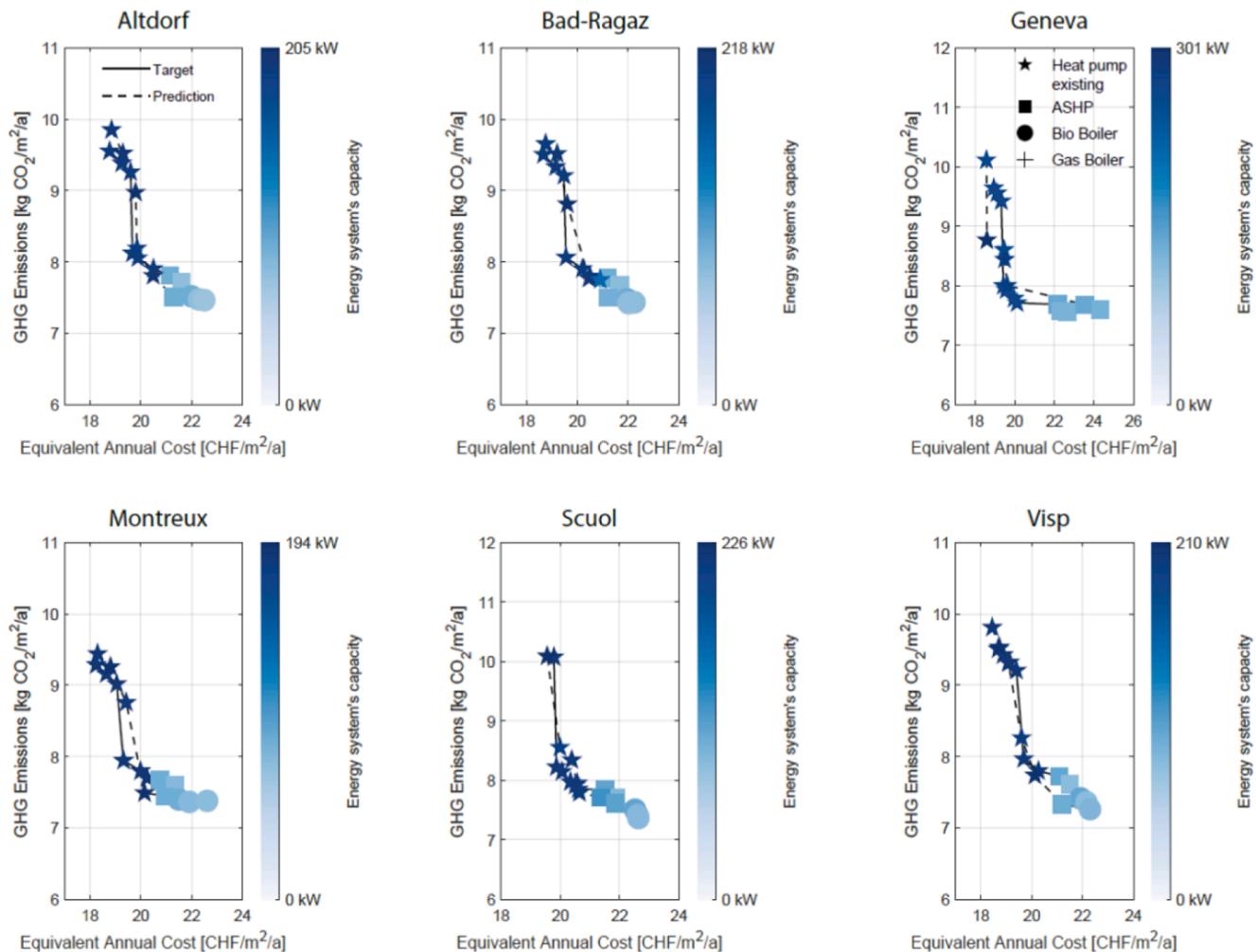


Fig. 19. Validation results for the energy system selection and sizing retrofit dimensions for a multi-family house for different climatic regions.

of choosing the existing heat pump instead of a new ASHP.

Regarding the energy system sizing, in contrast to the SFH, the surrogate model tends to overestimate the sizing of the installed systems with a percentage of up to 57 %. The lowest average capacity deviation is 0.6 kW for the Scuol region while the highest is 8.7 kW for the Altdorf region.

In Fig. 20, the final validation for the prediction of the PV that are suggested to be installed is presented. The maximum available roof area that can be covered by PV is almost 109 m². The lowest coverage of PV is 63 m² for the Montreux region for the second least expensive PF solution. The lowest average capacity deviation with less than 1 m² is observed for the Montreux, Scuol, and Visp regions. The highest deviation is just below 4 m² for the Bad-Ragaz region.

4.2. Large-scale retrofit analysis

One of the main advantages of the developed building-level surrogate retrofit model is the reduced computational cost. Specifically, after training, it can be used in a short period of time to compute near-optimal retrofit solutions for very large areas with many buildings, such as cities or municipalities. To showcase this advantage we used a large-scale case study that covers the municipality of Geneva. More specifically, we predict for the residential buildings described in Section 2.3 the near-optimal PFs. Those solutions can be used by a variety of stakeholders to form a strategy to motivate and incentivize building owners to retrofit their buildings. Moreover, it can be used by building owners to get an insight on the many different ways they could retrofit their building and

how they could benefit from it.

In Fig. 21, the near-optimal retrofit solutions for the residential buildings of the Geneva municipality are presented. More specifically, the distribution of the respective costs and emissions as well as the shares of the selected building envelope interventions and energy systems are shown. From the normalized cost histogram, we observe that almost 54 % of the retrofit solutions come with a cost of 19.5–25.5 CHF/m²/a. As for the emissions, 60 % of the retrofit solutions emit 7.6–12 kg CO₂/m²/a. The pie chart for the building envelope interventions shows that the enhancement of wall insulation combined with replacement of windows is the most dominant solution with 35 %, while the enhancement of the insulation of the roof comes second with almost 34 %. The least selected building envelope intervention is the enhancement of the insulation of the wall with just 0.7 %. It is interesting to note that around 11 % of the predicted retrofit solutions imply no change to the building envelope. Finally, from the energy system selection pie chart, we can see that 49 % of the retrofit solutions suggest the use of the existing system while for the rest a change is suggested. The most often selected new energy system to be installed is a biomass boiler with a share of just above 31 %.

The derived near-optimal retrofit solutions from Geneva municipality can also be plotted in the form of a map to get an overall understanding of the proposed retrofit measures. In Fig. 22, the near optimal retrofit solutions, which are the closest to the target of 10 kg CO₂/m²/a set in the Energy strategy 2050, for 1419 residential buildings, located in Cologny and Vandoeuvres, are given. Four maps are plotted for four retrofit dimensions, namely the objective values, i.e., the equivalent

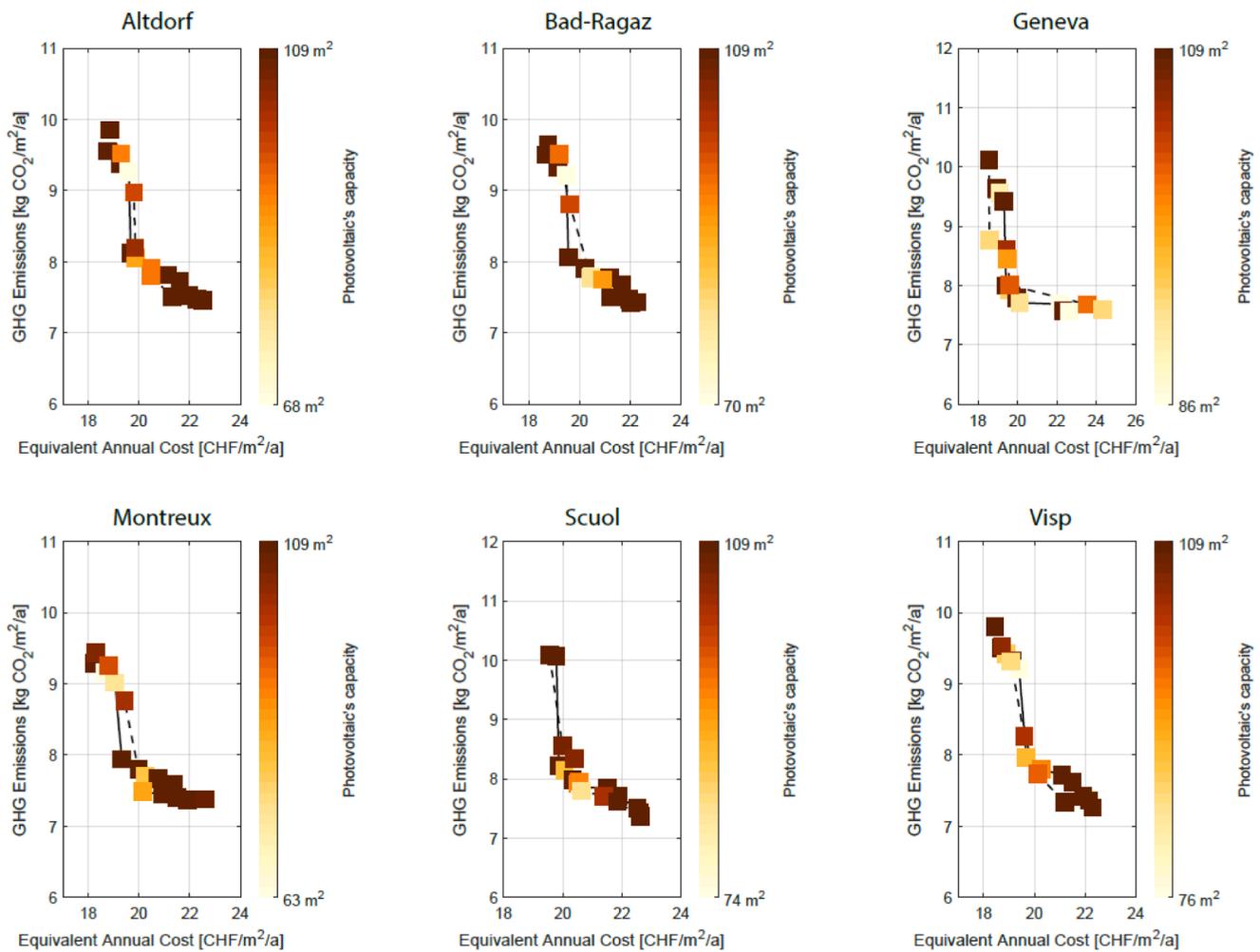


Fig. 20. Validation results for the sizing of the photovoltaics for a multi-family house for different climatic regions.

annual costs and CO_2 emissions, the building envelope interventions and energy system selection. The emission's target of $10 \text{ kg CO}_2/\text{m}^2/\text{a}$ is achieved in this region with a mean value of $9.9 \text{ kg CO}_2/\text{m}^2/\text{a}$ and a standard deviation of $0.6 \text{ kg CO}_2/\text{m}^2/\text{a}$. The respective equivalent annual costs have a mean value of $23 \text{ CHF}/\text{m}^2/\text{a}$ and a standard deviation of $6.7 \text{ CHF}/\text{m}^2/\text{a}$. To achieve this emission's target, the enhancement of the insulation of the walls and replacement of windows with more efficient ones is necessary for almost 43 % of the buildings in this region. Moreover, around 41 % of the buildings have to have a better insulated roof, while no change to the building envelope intervention accounts for only 5 %. Concerning the energy system selection, the most dominant solutions in order to be close to the specified emission's target is the use of ASHPs and biomass boilers with percentages of 35 % and 25 %, respectively. Finally, around 25 % of the buildings in the region can keep their existing system without missing the emission's target.

4.3. Advantages and limitations of the proposed surrogate model

The proposed large-scale retrofit approach, which uses a scalable building-level surrogate model, has certain advantages and limitations when compared to the conventional building retrofit approaches.

The main advantage of the developed approach, compared to the mixed building simulation and optimization one, is that it is easily scalable. This implies that once the building-level model is trained it can be used to predict near-optimal retrofit solutions for large areas with many buildings in a very short period of time. The prediction time of the

near-optimal retrofit solutions for the whole Geneva municipality, hosting 36'639 residential buildings, was 8 minutes and 36 seconds, in an Intel Xeon CPU E5-2680 running at 2.5 GHz. Based on [31], the mixed building simulation and optimization model needs 3.5 minutes to calculate the optimal retrofit solutions for just one building. This implies that for the number of buildings in question for the Geneva municipality we would need almost 90 days to predict the PF solutions if no parallelization is used.

Furthermore, the proposed approach offers a trade-off between scalability and accuracy in the prediction of almost all the retrofit dimensions. It is observed that the accuracy is higher when the performance is measured in a set of buildings rather than at the building-level. At large scale, the developed surrogate model achieves an average coefficient of determination for all the continuous retrofit dimensions of 0.91. Respectively, for the categorical retrofit dimensions an average f1 score of 0.89. However, the prediction performance is reduced for some retrofit dimensions to values of around 0.85 depending on the building type and the construction year, if the evaluation of the model is done per building.

Another advantage of the proposed approach is that it can be applied by non-expert people after it is trained by experts. This can be explained from the fact that the input set consists of building information which is already known or easily accessible from tenants or building owners. Moreover, there is no technical expertise needed in terms of tuning parameters and running simulation or optimization models. Eventually, such a scalable building-level retrofit model can be used from various stakeholders, independent of their expertise and the level of it, so that

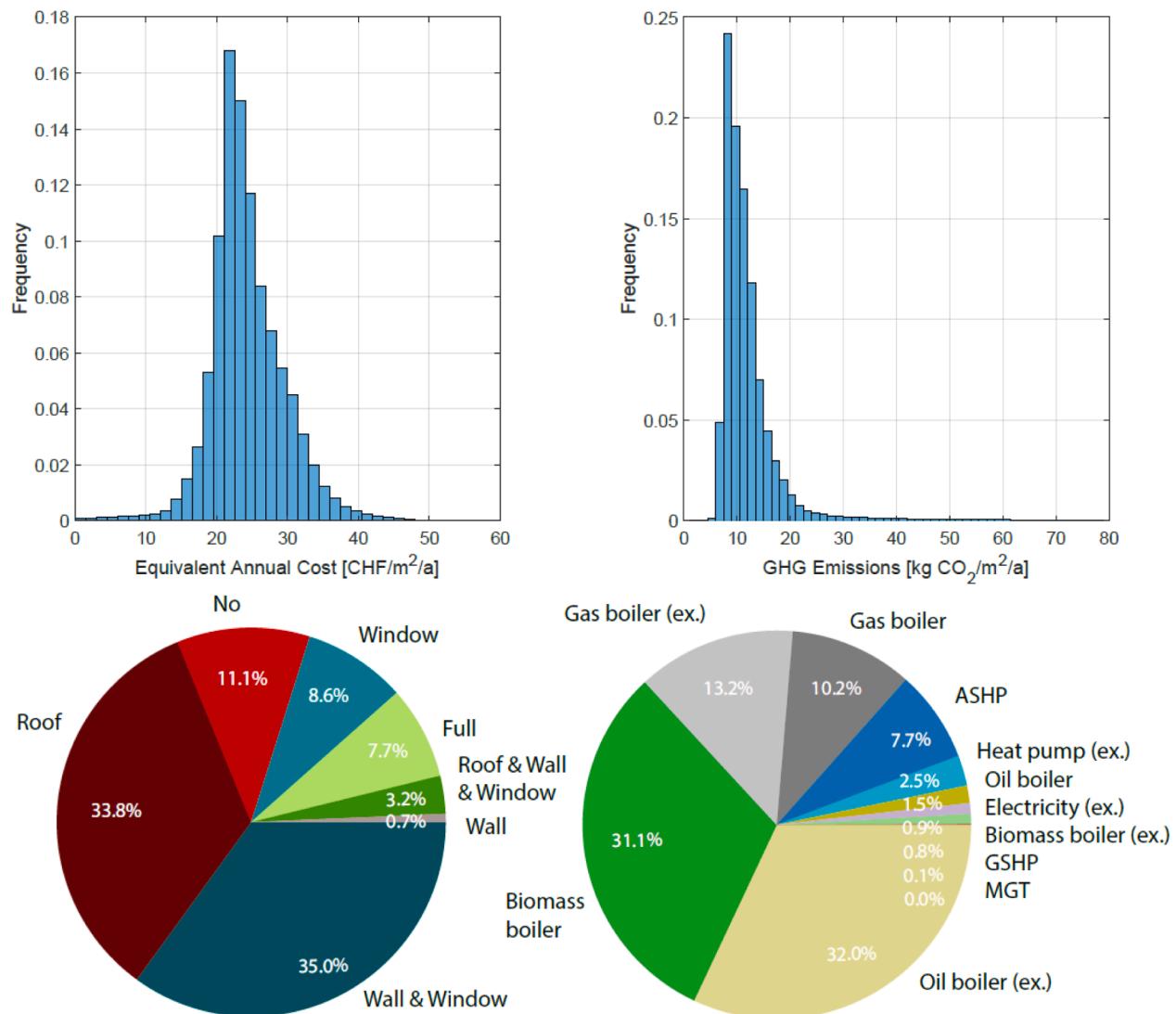


Fig. 21. Distribution of the costs and the emissions (top) and the shares of building envelope interventions and energy systems selections (bottom) for the near-optimal retrofit solutions for Geneva municipality.

they get a first insight of the potential of building retrofit and the benefits it can offer.

There are certain limitations of the proposed approach. The most important is that it fails to predict the PV capacity of SFH, or else residential buildings with small roof areas, if the per building coefficient of determination is used as an evaluation metric. For such buildings the MILP should be used to correctly size the PVs. Moreover, in the case that a solar thermal collector or an electrical storage is selected to be installed, the capacity should be calculated separately with the use of the MILP. However, the computational cost of running the MILP to obtain those values is negligible and it doesn't negate the benefits of using the proposed surrogate model.

Moreover, the proposed model is prone to errors imposed by the use of building archetypes, or else by the clustering process used to derive the building archetypes. Specifically, this error is incorporated on the training data and it has nothing to do with the process of training and building up the proposed surrogate model. This implies that in the presence of more accurate building data the whole methodology can be applied again as is.

Finally, the proposed approach is limited to the Swiss residential building stock. However, the same methodology could be applied to any

other building type, e.g., offices and commercial buildings, and any other country. The only prerequisite is the availability of building data to train the ANN models.

5. Conclusions

Building retrofit is one of the most promising approaches towards reducing the environmental footprint of buildings. However, in order to achieve the overall energy targets set by the European Union for 2050 there is a vital need towards a coordinated large-scale retrofit plan. This implies that retrofitting at building-scale is not sufficient if there is no plan for retrofitting at larger scale, i.e., cities and municipalities. Existing approaches for large-scale retrofitting, both bottom up and top down, lack in terms of accuracy or computational cost as the scale of case study increases.

The goal of this work was to develop a bottom-up large scale retrofit approach that is capable of providing near optimal retrofit solutions for all buildings in various climatic regions. The proposed approach, based on a scalable building-level surrogate model, provides a very good balance between accuracy and computational cost. The surrogate model consists of two sub-models, one predicting the CO₂optimal solution,

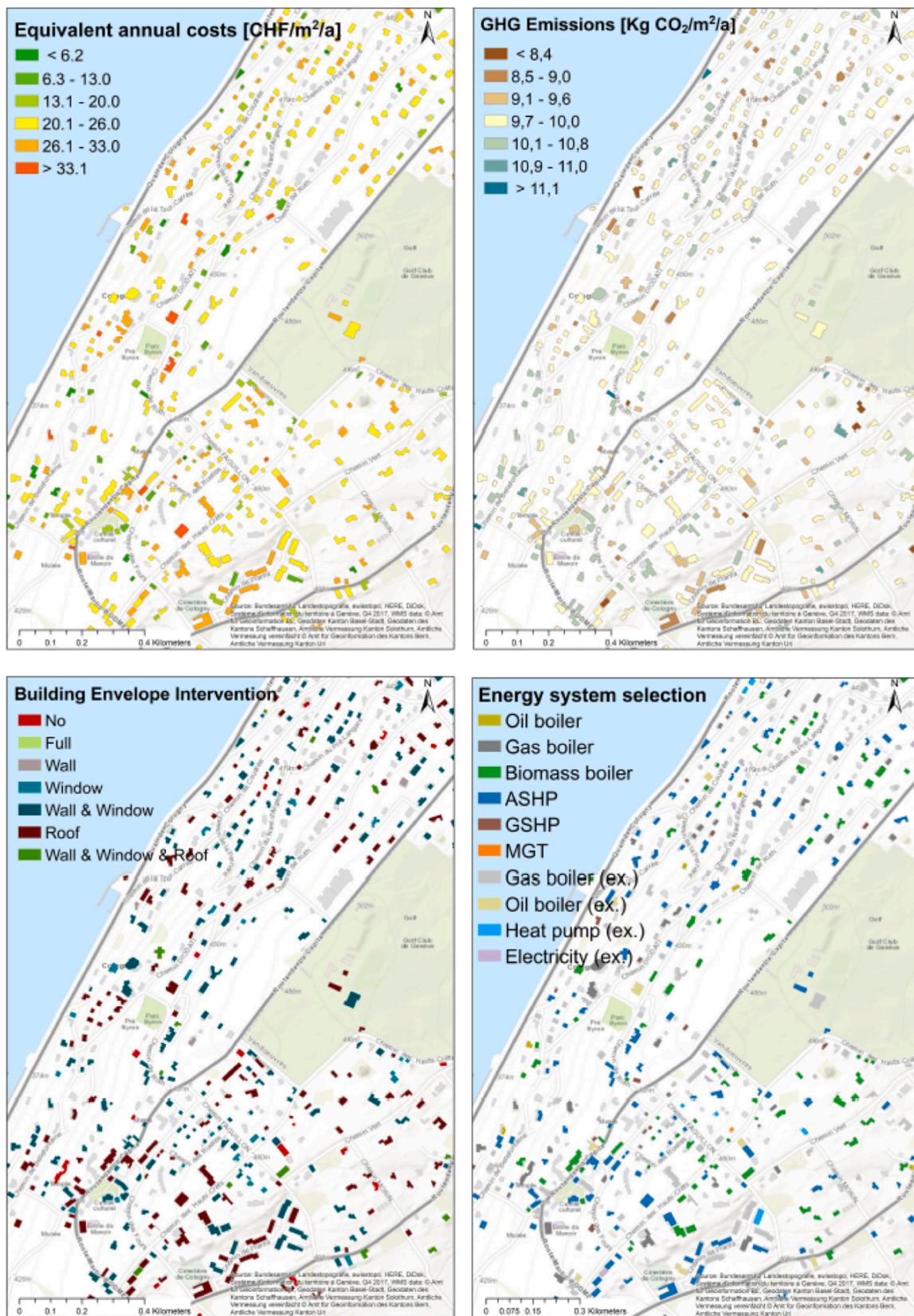


Fig. 22. The closest to the target of $10 \text{ CO}_2/\text{m}^2/\text{a}$, as set in the Energy strategy 2050, near-optimal retrofit solutions calculated by the surrogate model for Cologny and Vandoeuvres in Geneva municipality. The retrofit dimensions depicted are the necessary building envelope interventions (bottom left) and energy system selections (bottom right), and the respective total costs (top left) and CO_2 emissions (top right).

while the other predicts the rest of the Pareto front solutions. Each of the sub models consists of several artificial neural networks that sequentially coordinate to predict all the different retrofit dimensions involved in each Pareto front solution. The prediction performance varies depending on the retrofit dimension, the building type, and the climatic region investigated. Overall, the surrogate model achieves very good performance for all the retrofit dimensions with both average coefficient of determination and f1 scores of above 0.8. On the other hand, when the surrogate model is evaluated per building the best performance is achieved for the objective values, i.e., cost and emissions, the solar thermal selection, the energy system selection and the building envelope intervention with per building median coefficient of determination values of above 0.85. However the model fails to predict the capacity of the selected photovoltaics for single family houses, while the respective prediction performance for MFH achieves an f1 score of 0.58.

As soon as the model is trained it can be used to predict near-optimal retrofit solutions for many buildings in a very short period of time compared to the conventional building retrofit approaches. Moreover, the set of inputs required to run the model is small and the data is easy to obtain or already known by the building owners. Even though the accuracy of the developed model is not and cannot be higher than the conventional mixed building simulation and optimization model, it can be used to obtain first insights into the potential of building retrofit for large areas. Moreover, the retrofit dimensions that are not predicted very accurately, could be verified and determined with a retrofit optimization model. This can be done in a short period of time since many variables of the optimization problem can be fixed with the values predicted from the surrogate model. For instance, some design variables, i.e., the building envelope intervention and the energy system selection, could be fixed with high confidence that the predictions are accurate.

Training the surrogate model with data coming from real buildings from different climatic regions, and not from building archetypes, would improve the generalization ability and the performance of the model at building level. The building archetypes used in this study cover a certain percentage of the whole residential building stock. By using a much larger set of real buildings as sample set would make the surrogate model more robust against new buildings. However, in order to achieve an acceptable training performance the set of real buildings would have to be very large and this would increase significantly the computational cost of collecting the necessary optimal building retrofit solutions. To define the necessary set of real buildings to be used, active learning could be exploited. Active learning would give information on how large the data set should be and also which buildings have to be taken into account in the training process so as to achieve the best possible performance.

As a future work to further improve the training performance of the surrogate model several steps could be followed. Firstly, a sensitivity analysis study can be applied to find out which of the input variables are influencing the most the variance of the output. This would also help to further reduce the input features used for training or add some others which are not included in this study. Secondly, a customized loss function, i.e., a per building loss function, could be used rather than the mean squared error. This could potentially lead to a better model performance even though it would also increase significantly the computational cost of the training. Finally, we could try to improve the performance of the different retrofit dimensions by adding more input features. For instance, using more inputs such as roof slope and orientation might improve the prediction of the sizing of the photovoltaic panels.

To conclude, such a surrogate model can be used in combination with the mixed building simulation and optimization approach in a hybrid format so as to benefit from both worlds. The surrogate model is capable of predicting building-level retrofit solutions within a short time. On the other hand, the mixed building simulation and optimization approach ensures optimality of the derived results and it can also provide district-level solutions. Marrying those two approaches in

retrofit analysis for large areas would be very beneficial.

CRediT authorship contribution statement

Emmanouil Thrampoulidis: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing – original draft. **Gabriela Hug:** Writing – review & editing, Supervision, Project administration. **Kristina Orehoung:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgements

This research project was partially financially supported by the Swiss Innovation Agency Innosuisse and was part of the Swiss Competence Center for Energy Research SCCER FEEB&D.

References

- [1] Energy performance of buildings directive; n.d. https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en (accessed June 9, 2022).
- [2] Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings OJ L 153 2010.
- [3] Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC Text with EEA relevance 2012.
- [4] Directive of the European parliament and of the council on the energy performance of buildings (recast) 2021.
- [5] Energieperspektiven 2050 n.d. <https://www.bfe.admin.ch/bfe/en/home/policy/energy-strategy-2050/documentation/energy-perspectives-2050.html> (accessed January 22, 2019).
- [6] Streicher KN, Berger M, Panos E, Narula K, Soini MC, Patel MK. Optimal building retrofit pathways considering stock dynamics and climate change impacts. Energy Policy 2021;152:112220. <https://doi.org/10.1016/j.enpol.2021.112220>.
- [7] Zhang H, Hewage K, Karunathilake H, Feng H, Sadiq R. Research on policy strategies for implementing energy retrofits in the residential buildings. J Build Eng 2021;43:103161. <https://doi.org/10.1016/j.jobe.2021.103161>.
- [8] Nielsen AN, Jensen RL, Larsen TS, Nissen SB. Early stage decision support for sustainable building renovation – A review. Build Environ 2016;103:165–81. <https://doi.org/10.1016/j.buildenv.2016.04.009>.
- [9] Deb C, Schlueter A. Review of data-driven energy modelling techniques for building retrofit. Renew Sustain Energy Rev 2021;144:110990. <https://doi.org/10.1016/j.rser.2021.110990>.
- [10] Mora TD, Righi A, Peron F, Romagnoni P. Cost-Optimal measures for renovation of existing school buildings towards nZEB. Energy Procedia 2017;140:288–302. <https://doi.org/10.1016/j.egypro.2017.11.143>.
- [11] Nguyen A-T, Reiter S, Rigo P. A review on simulation-based optimization methods applied to building performance analysis. Appl Energy 2014;113:1043–58. <https://doi.org/10.1016/j.apenergy.2013.08.061>.
- [12] Mostavi E, Asadi S, Boussaa D. Framework for Energy-Efficient Building Envelope Design Optimization Tool. J Archit Eng 2018;24. [https://doi.org/10.1061/\(ASCE\)AE.1943-5568.0000309](https://doi.org/10.1061/(ASCE)AE.1943-5568.0000309).
- [13] Wu R, Mavromatis G, Orehoung K, Carmeliet J. Multiobjective optimisation of energy systems and building envelope retrofit in a residential community. Appl Energy 2017;190:634–49. <https://doi.org/10.1016/j.apenergy.2016.12.161>.
- [14] Krarti M. Optimal Design and Retrofit of Energy Efficient Buildings, Communities, and Urban Centers. Elsevier; 2018. 10.1016/C2016-0-02074-0.
- [15] Balaras CA, Gaglia AG, Georgopoulou E, Mirasgedis S, Sarafidis Y, Lalas DP. European residential buildings and empirical assessment of the Hellenic building stock, energy consumption, emissions and potential energy savings. Build Environ 2007;42:1298–314. <https://doi.org/10.1016/j.buildenv.2005.11.001>.
- [16] Fracastoro GV, Serraino M. A methodology for assessing the energy performance of large scale building stocks and possible applications. Energy Build 2011;43:844–52. <https://doi.org/10.1016/j.enbuild.2010.12.004>.
- [17] IEA-ETSP | Times n.d. <https://iea-etsap.org/index.php/etsap-tools/mod-el-generators/times> (accessed June 9, 2022).

- [18] NEMS - National Energy Modeling System: An Overview n.d. [https://www.eia.gov/analysis/pdfs/0581\(2009\)index.php](https://www.eia.gov/analysis/pdfs/0581(2009)index.php) (accessed June 9, 2022).
- [19] Reinhart CF, Cerezo DC. Urban building energy modeling – A review of a nascent field. *Build Environ* 2016;97:196–202. <https://doi.org/10.1016/j.buildenv.2015.12.001>.
- [20] He M, Brownlee A, Lee T, Wright J, Taylor S. Multi-objective Optimization for a Large Scale Retrofit Program for the Housing Stock in the North East of England. *Energy Procedia* 2015;78:854–9. <https://doi.org/10.1016/j.egypro.2015.11.007>.
- [21] Mata E, Kalagasisidis AS, Johnsson F. Contributions of building retrofitting in five member states to EU targets for energy savings. *Renew Sustain Energy Rev* 2018; 93:759–74. <https://doi.org/10.1016/j.rser.2018.05.014>.
- [22] Palma P, Gouveia JP, Barbosa R. How much will it cost? An energy renovation analysis for the Portuguese dwelling stock. *Sustain Cities Soc* 2022;78:103607. <https://doi.org/10.1016/j.scs.2021.103607>.
- [23] Streicher KN, Mennel S, Chambers J, Parra D, Patel MK. Cost-effectiveness of large-scale deep energy retrofit packages for residential buildings under different economic assessment approaches. *Energy Build* 2020;215:109870. <https://doi.org/10.1016/j.enbuild.2020.109870>.
- [24] Ascione F, Bianco N, Stasio CD, Mauro GM, Vanoli GP. Addressing Large-Scale Energy Retrofit of a Building Stock via Representative Building Samples: Public and Private Perspectives. *Sustainability* 2017;9:940. <https://doi.org/10.3390/su9060940>.
- [25] Eggimann S, Vulic N, Rüdisüli M, Mutschler R, Orehonig K, Sulzer M. Spatiotemporal upscaling errors of building stock clustering for energy demand simulation. *Energy Build* 2022;258:111844. <https://doi.org/10.1016/j.enbuild.2022.111844>.
- [26] Mastrucci A, Marvuglia A, Leopold U, Benetto E. Life Cycle Assessment of building stocks from urban to transnational scales: A review. *Renew Sustain Energy Rev* 2017;74:316–32. <https://doi.org/10.1016/j.rser.2017.02.060>.
- [27] Wu Z, Wang B, Xia X. Large-scale building energy efficiency retrofit: Concept, model and control. *Energy* 2016;109:456–65. <https://doi.org/10.1016/j.energy.2016.04.124>.
- [28] Lee S, Hong T, Piette M, Taylor-Lange S. Energy retrofit analysis toolkit for commercial buildings: A review 2015;89:1087–100. <https://doi.org/10.1016/j.energy.2015.06.112>.
- [29] Heo Y, Augenbroe G, Graziano D, Muehleisen RT, Guzowski L. Scalable methodology for large scale building energy improvement: Relevance of calibration in model-based retrofit analysis. *Build Environ* 2015;87:342–50. <https://doi.org/10.1016/j.buildenv.2014.12.016>.
- [30] Jennings M, Fisk D, Shah N. Modelling and optimization of retrofitting residential energy systems at the urban scale. *Energy* 2014;64:220–33. <https://doi.org/10.1016/j.energy.2013.10.076>.
- [31] Thrampoulidis E, Mavromatidis G, Lucchi A, Orehonig K. A machine learning-based surrogate model to approximate optimal building retrofit solutions. *Appl Energy* 2021;281:116024. <https://doi.org/10.1016/j.apenergy.2020.116024>.
- [32] Murray P, Marquant J, Niffeler M, Mavromatidis G, Orehonig K. Optimal transformation strategies for buildings, neighbourhoods and districts to reach CO₂ emission reduction targets. *Energy Build* 2020;207:109569. <https://doi.org/10.1016/j.enbuild.2019.109569>.
- [33] Remund J, Müller SC, Schilten C, Rihm B. The use of Meteonorm weather generator for climate change studies. 2010.
- [34] Swiss Geoportal. GeoAdminCh n.d. <https://map.geo.admin.ch> (accessed October 3, 2022).
- [35] ASHRAE 90.1 n.d. <https://www.ashrae.org/technical-resources/bookstore/standards-90-1> (accessed October 3, 2022).
- [36] Orehonig K, Fierz L, Allan J, Eggimann S, Vulic N, Bojarski A. CESAR-P: A dynamic urban building energy simulation tool. *J Open Source Softw* 2022;7:4261. 10.21105/joss.04261.
- [37] Thrampoulidis E. Emulation of energy optimization models via machine learning towards the design of a building energy consultant. Master thesis. ETH Zurich, 2017.
- [38] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. ArXiv14126980 Cs 2014.