COMPARISON BETWEEN TWO GENETIC ALGORITHMS MINIMIZING CARBON FOOTPRINT OF ENERGY AND MATERIALS IN A RESIDENTIAL BUILDING

Richard Gagnon[1], Louis Gosselin[1], Sumee Park[2], Sebastian Stratbücker[2], Stéphanie Decker[3]

[1] Département de génie mécanique, Université Laval, Québec City, Québec, Canada, G1V 0A6.

[2] Fraunhofer Institute for Building Physics IBP, Standort Holzkirchen, Valley 83626, Germany.

[3]Centre de Ressources technologiques Nobatek, Anglet 64600, France.

Article accepté pour publication dans : Journal of Building Performance Simulation, Volume 12, 2019.

Abstract

The emergence of building performance optimization is recognized as a way to achieve sustainable building designs. In this paper, the problem consists in minimizing simultaneously the emissions of greenhouse gases (GHG) related to building energy consumption and those related to building materials. This multi-objective optimization problem involves variables with different hierarchical levels, i.e. variables that can become obsolete depending on the value of the other variables. To solve it, NSGA-II is compared with an algorithm designed specifically to deal with hierarchical variables, namely sNSGA. Evaluation metrics such as convergence, diversity and hypervolume show that both algorithms handle hierarchical variables, but the analysis of the Pareto front confirms that in the present case, NSGA-II is better to identify optimal solutions than sNSGA. All the optimal solutions are made of buildings with wooden envelopes and relied either on heat pumps or on electrical heaters for proving heating.

Keywords: Hierarchical variables; NSGA-II; building performance optimization; heating systems; building envelope

Nomenclature

Variables

$CO2_E$	CO_2 equivalent due to the building energy consumption [ton of CO_{2eq}]
$CO2_M$	CO_2 equivalent due to the building materials and heating systems [ton of CO_{2eq}]
$C(P^{(t)})$	Metric of convergence [-]
d_i	Euclidean distance [-]

Env	Wall envelope type [-]
E_{tot}	Total energy consumption of the heating system [kWh]
f_k	Values of the kth objectives function [ton of CO_{2eq}]
$F^{(t)}$	Non-dominated set of solution from generation t [-]
Gen	Number of generation before stopping the evolution process [-]
HE	Heat emitter type [-]
HS	Heating system type [-]
HV	Hypervolume (metric of diversity and convergence) [-]
I	Carbon intensity of energy [CO_{2eq}/kWh]
$\dot{m}_w$	Maximum water flow rate [kg/hr]
P^*	Pareto optimal point
$P^{(t)}$	Population from generation t [-]
PF	Penalty function [-]
Pop	Population size [-]
P_c	Crossover probability [-]
P_m	Mutation probability [-]
q_{aux}	Heat pump auxiliary power [kW]
q_{bo}	Boiler heating system capacity [kW]
q_{er}	Electrical heating system capacity [kW]
q_{he}	Heater rated capacity [kW]
q_{hp}	Heat pump heating system capacity [kW]
RSP	Percentage of time of respected set point condition [-]
t	Generation number [-]
T_a	Zone ambient temperature [°C]
T_{aux}	Enabling temperature for auxiliary electric power unit [°C]
T_{sp}	Set point temperature for unoccupied hours [°C]
T_z	General set point temperature [°C]
T_w	Water temperature at radiator inlet [°C]
Win	Windows type [-]
WWR	Windows to wall ratio [-]
Yr	Number of year [-]

Greek letters

α	Monthly average outdoor temperature increase rate [-]
δ	Number of hours of the simulation period [hr]
Δ	Metric of diversity [-]
Δt	Time step [hr]

Subscripts

E	East
N	North
S	South
W	West

Acronyms

BEM	Building energy modeling
BPO	Building performance optimization
CO_{2eq}	Equivalent emissions of carbon dioxide
CLT	Cross-laminated timber wall
DSW	Double stud wall
FSS	Faure sequence sampling
GAs	Genetic algorithms
HSS	Hammersley sequence sampling
HVAC	Heating, ventilating and air conditioning
LFW	Light frame wooden wall
LHS	Latin hypercube sampling
PSO	Particle swarm optimization
SBX	Simulated binary crossover
sGA	Structured genetic algorithm
sNSGA	Structured non-dominated sorting genetic algorithm

1 Introduction

Building energy modeling (BEM) combined with building performance optimization (BPO) has been an active field of research over the last few years. Evins reviewed 74 studies among the most significant in this domain (Evins 2013). The urgent need to design greener and yet affordable buildings has contributed to the emergence of this body of work. The challenges and opportunities related to the integration of such optimization tools were analyzed (Attia et al. 2013). The authors reported the main following obstacles: the low trust of professionals in the results, the long computational time and the lack of standard systematic approach.

According to literature, genetic algorithms (GAs) are the most commonly used method for multi-objective building optimization (Evins 2013). They are particularly attractive to solve problems with discrete and continuous variables, and identify a set of non-dominated solutions. They have been compared to other algorithms in many studies where their outperformance has been proven (Brownlee, Wright, and Mourshed 2011; Bichiou and Krarti 2011; Tuhus-Dubrow and Krarti 2010). Bichiou and Krati have compared three algorithms for the optimization of envelopes and HVAC systems in residential buildings. Among them was a GA, a particle swarm optimization (PSO) technique and a sequential search algorithm. The GA and PSO required less computational time than the last one in order to obtain optimal solutions. Moreover, the GA has shown better performances when more than 10 variables were considered in the

optimization.

Former GAs exhibited some deficiencies such as the lack of elitism, the convergence to local optimum or the lack of genetic diversity. These issues were addressed with the development of the Non-dominated Sorting Genetic Algorithm (NSGA-II) (Deb et al. 2002). It is used extensively in energy and building studies as reviewed by Attia et al. (2013) and it is one of the most efficient GAs according to the following comparison studies (Deb et al. 2002; Zitzler, Deb, and Thiele 2000).

Despite of their advantages, authors claimed that GAs can perform~~s~~ poorly when they have to solve deceptive problems, in which case the algorithm can mislead the search to some local optima rather than the global optimum (Chen et al. (2008) and Manso and Correia (2013)). When this situation occurs, the set of solutions become homogeneous and the poor genetic diversity makes it hard to move toward another region of the search space. It can lead to a premature convergence of the optimization process. The presence of hierarchical variables enhances the probability to face a deceptive problem and can reduce the efficiency of typical GAs (Dasgupta and McGregor 1993; Molfetas 2006; Dasgupta 1994; Tiwari and Roy 2002), but no building optimization studies have addressed that aspect. Figure 1 shows a schematic representation of this type of variables. In this example, the variable a can take 3 different discrete values (e.g., the type of heating system which can be chosen among a list). This selection activates a specific set of inherited low-level variables: $\{a_{11}, a_{12}, a_{13}\}$, $\{a_{21}\}$ or $\{a_{31}, a_{32}\}$ (e.g., variables associated to a specific type of heating system). Inherited variables can be discrete or continuous. It happens that realistic building multi-objective optimization problems typically involve hierarchical variables, but the building community has not addressed the implications of such a feature. As can be seen, it is still unclear whether a particular algorithm should be used to achieve a satisfactory convergence rate and a true non-dominated set of solutions when dealing with such variables in typical BPO problems.

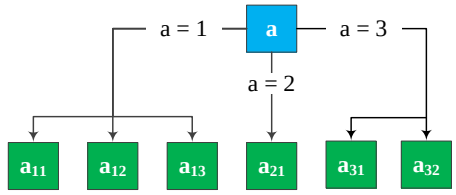

Figure 1: Schematic representation of an example of hierarchical variables. In blue, a high-level variable, in green, the inherited low-level variables.

In an early stage of design optimization, several choices of sub-systems should be considered including their inherited variables. However, standard GAs do not usually understand the link existing between a high-level design variable (e.g. sub-system selection such as the type of heating system) and the low-level design variables (e.g. the inherited variables belonging to a specific sub-system such as the design parameters related to the specific type of heating system). Dasgupta and McGregor (1993) have developed a more robust algorithm for such cases: a structured Genetic Algorithm (sGA). Based on the theoretical approach of sGA, the present study proposes a modified NSGA-II implemented in DEAP[1] named sNSGA. The main goal of this study is to compare the performance of sNSGA and NSGA-II for a building optimization problem with hierarchical variables. The algorithms are applied to a relevant and timely multi-objective optimization problem related to the energy and environmental performance of a research building facility (Strachan, Svehla, et al. 2016). Since the optimal solutions returned by both algorithms are also valuable to the industry and research community, their analysis is briefly presented at the end of the paper.

2 Relevant studies

The first usages of BPO were oriented toward the optimization of one set of variables at a time: e.g. architecture variables, daylighting variables, natural ventilation strategies, variables related to heating, ventilating and air conditioning (HVAC) systems, managing of energy storage and renewable energy, etc. (Evins 2013). Despite the fact that it is an efficient way to optimize systems, this sequential approach does not consider the potential synergies between different set of variables and narrows down the capacity of designing greener and at the same time cheaper buildings. On the contrary, holistic approaches consider simultaneously the variables of several systems in the problem formulation and resolution.

2.1 Multi-objective optimization without hierarchical variables

Hamdy, Hasan, and Siren (2011) have considered five types of heating systems combined with envelope parameters, i.e. insulation thickness, windows type, building tightness and shading type. Their main goal was to minimize the investment cost and the carbon dioxide equivalent emissions. In a continuation of this work, PV panels were included in order to achieve a nearly zero energy design (Hamdy, Hasan, and Siren 2013). Verbeeck and Hens (2007) used GAs to minimize the life cycle cost, energy and

[1] DEAP is an evolutionary framework freely available in python designed for a fast prototyping of new algorithms

environmental impacts of low energy dwellings in Belgium. The methodology was based on a two-stage framework. At first, the building envelope was optimized by minimizing the 3 objectives functions. Based on the net energy demand of every scenario, the energy systems were added to every scenario in order to determine the new values of the 3 objectives. In another study (Bichiou and Krarti 2011), the HVAC systems were represented as discrete variables. For example, the algorithm could select an air conditioning system combined with a furnace or an air conditioning system combined with electric resistances. Wright and Farmani (2001) have simultaneously optimized the envelope, the HVAC system size and the control strategy using a GA. The intention was to reduce the energy, cost and thermal discomfort.

In the studies presented above, the energy systems were represented as "black boxes", i.e. that the detailed composition of the systems were not fully implemented and interconnected with the building model. In these studies (Hamdy, Hasan, and Siren 2011a, 2013), the space-heating energy demand is determined in a first stage without specifying the heating system. The best solutions were then combined with 4 heating system options and depending of the selection, new primary energy consumption and life cycle cost were determined. There were no hierarchical variables related to the specific heating systems that were optimized simultaneously.

2.2 Multi-objective optimization with hierarchical variables

A whole building design optimization problem is obviously composed with hierarchical variables (Wang 2005; Wang, Zmeureanu, and Rivard 2005). In the early 1990s, a new approach was developed to consider hierarchy between variables, namely sGA (Dasgupta and McGregor 1993). This algorithm makes a distinction between the chromosome and the phenotype. The chromosome is represented by a set of strings, each one representing a variable. It contains all the possible variables. The information transferred to the evaluation function comes from the phenotype. Only the active genes of the chromosome are transferred to the phenotype. The other "dormant" genes are stored in the chromosome and will be activated only when the higher level variable value is meaningful for them. This strategy allows preserving the quality of the part of a chromosome that was dormant (Parmee 1998).

Although sGA allows to explore several design paths, some deficiencies and limitations are revealed (Parmee 1998): the mutation probability P_m is imbalanced between the different design paths and the crossover operator applied to two designs may damage them if they are represented by different high-level variables. If P_m and the crossover probability P_c are low, it will reduce the chance of disrupting good

combinations; however, it may trap the search to local optima without exploring all the design paths. To overcome those limitations, several solutions are proposed: a variable mutation approach according to the variables level, a hybrid mutation approach with variation after a certain number of generations and a crossover restriction applied to the high-level variables. The result algorithm is named GAANT (GA combined with ant colony search) that combined the most successful solutions together.

The conceptual framework developed by Dasgupta and McGregor (1993) was adapted by Rafiq, Mathews, and Bullock(2003) to explore several structural designs of building construction. Traditionally, the major structural dimensions and grid layouts are determined by specialists and the remaining parameters optimized in a later phase according to specific objectives (cost, speed of construction, etc.). This restrictive approach in terms of design options was overcome with sGA. It has been used successfully to explore many types of frame and their sub-components (Rafiq, Mathews, and Bullock 2003). The algorithm was implemented on a platform named DPRO, on which one had the capacity to enable or disable a part of the gene. Wang et al. have proposed an object-oriented framework for the optimization of green buildings using sGA (Wang, Rivard, and Zmeureanu 2005). The high-level variables were the type of walls whereas the lower level variables were related with their detailed composition. Other variables like the orientation, the shape of building and the windows ratio are enabled independently of the choice of the high-level variables. The framework has been validated using mathematical functions for single objective and multi-objective optimization.

The literature shows that sGA is capable of representing hierarchical variables but the actual benefits of using such an approach to solve BPO problems involving hierarchical variables are unclear. In fact, although this type of variables is abundant in BPO, there is still a lot of unknown as to how to treat them and how they affect the final solutions. This study addresses this specific issue.

2.3 Introduction to genetic algorithms

GAs have been used to solve single and multi-objective optimization problems (Tuhus-Dubrow and Krarti 2010; Hamdy, Hasan, and Siren 2013; Palonen, Hasan, and Siren 2009; Evins 2015; Gossard, Lartigue, and Thellier 2013; Caldas and Norford 2003). They are particularly attractive to optimize non-differentiable functions (Magnier and Haghighat 2010). The optimization parameters (e.g., population size, crossover probability, mutation probability) can have an influence on the efficiency and the performance of the method. However, it is difficult to apply common rules to set the values of those

parameters (Alajmi and Wright 2014; Wright and Alajmi 2016). Most of them influence the speed of convergence and the design space covered in an opposite way. Different studies argued that the population size and the number of generations are the most important parameters (Seshadri 2006; Evins 2010; Goldberg, Deb, and Clark 1991). The initial population should contain a rich set of solutions, which means that a higher number of design variables demands a bigger population.

3 Structured Non-dominated Sorting Genetic Algorithm

In this paper, a structured non-dominated sorting genetic algorithm was developed to solve multi-objective building design problems with hierarchical variables. It is a modified version of NSGA-II and its development was inspired from the following studies: Dasgupta and McGregor (1993), Molfetas (2006), Wang, Rivard, and Zmeureanu (2005). The algorithm which we named sNSGA is presented in this section. It has the capacity to deal with structured variables and is specifically adapted to the case of BPO. Figure 2 shows a schematic representation of a genetic algorithm and distinctions between sNSGA and NSGA-II are detailed in the next sections.

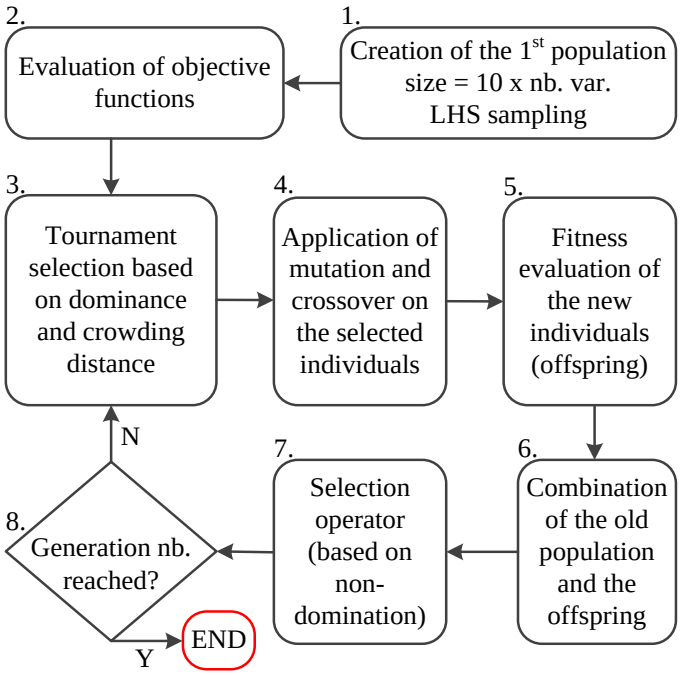

Figure 2: Schematic representation of a genetic algorithm.

Most of the steps are processed from a script written in Python (using DEAP framework) except for steps 2 and 5 that are related to the evaluation of the objective functions, which are processed by TRNSYS in our case (energy simulation software). The parameters defining each individual in step 1 are transferred to TRNSYS to perform an energy simulation. The results are sent back to the script for the evaluation of the fitness functions. Steps 3 to 8 describe the genetic evolution procedure. During this process, the fitter individuals will reproduce and create an offspring population ready for a new evaluation. The evolution stops after a pre-defined number of generations *Gen*, as defined in Table 4. Based on the thorough analysis of Section 5, it will be shown that the value of *Gen* that was chosen was sufficient to provide a good convergence to the optimal solutions.

3.1 Initial population

There are several techniques to initialize the first population and each one can affect the convergence of the optimization algorithm. As BPO requires a lot of time-consuming simulations, convergence speed is a critical characteristic to consider. For a typical GA, a uniformly distributed random population is first created. Preechakul and Kheawhom (2009) have compared this technique with a Latin hypercube sampling (LHS), a Faure sequence sampling (FSS) and a Hammersley sequence sampling (HSS) for 9 test problems. It was demonstrated that LHS, FSS and HSS are more effective in terms of speed of convergence and diversity of optimal solutions. Therefore, the present sNSGA uses LHS to create the first population with a population size (*Pop)* equal to 10 times the number of variables. The population size is based on the work of Evins (2016, 2010), Wright and Alajmi (2016) and Hamdy, Hasan, and Siren (2011b).

3.2 Description of the evolution process

Tournament selection

In step 3, a tournament selection based on crowding distance and dominance is used to generate two lists of individuals of size *Pop/2* ready for reproduction. This operation is applied identically as for the NSGA-II.

Simulated binary crossover

From the two lists generated above, pairs of individuals are selected for the mating process. The NSGA-II uses the genetic operator named simulated binary crossover (SBX) in the case of real-coded GAs

(Agrawal, Deb, and Agrawal 1995). SBX was designed to simulate the behavior of a binary coded GA with one-point crossover. The sNSGA proposes a modification to the SBX. If both individuals have the same high-level values, only the independent variables and the inherited low-level variables are subject to mating. The independent variables referred to those without any relation to the high-level variables. The other variables keep their original value. If both individuals have different high-level values, only the independent variables are subject to mating. In this paper, P_c is set to 0.9, a typical value recommended by Deb et al. (2002).

Polynomial mutation

The mutation is an important operator in GAs in order to explore new regions of the design space. A high P_m value enhances the exploratory power of the algorithm while reducing the speed of convergence. NSGA-II uses a genetic operator named polynomial mutation. The operator is applied systematically to the resulting individuals from the mating process. The sNSGA proposes the following modification: the only variables that can mutate are the independent variables, the high-level variable and the inherited low-level variables. The mutation probability is defined by $P_m = 1/x$ with x being the number of variables available for mutation, based on Deb et al. (2002).

As can be seen above, the different parameters of the algorithms (i.e., population size, crossover rate, mutation rate, number of generations) were all selected based on current best practices and recommendations from literature. It is worth to mention that different techniques have also been developed to "fine-tune" the parameters of evolutionary algorithms (see, for example, López-Ibáñez et al. (2016) and Hutter et al. (2011)), but these techniques were not used explicitly here since the information available in literature was deemed sufficient for the purpose of selecting the proper algorithm parameters.

3.3 DEAP, an evolutionary framework for rapid prototyping of new algorithm

Many optimization tools are available in order to support BPO. Attia et al. (2013) have sorted 12 tools in order of importance from an exhaustive review of 165 publications in the domain. The results show that GenOpt and Matlab are the two most commonly used. With similar features as the one included in GenOpt and Matlab, DEAP is an open-source platform designed for a fast prototyping of new algorithms (Rainville et al. 2012). New toolboxes, genetic operators and algorithms can easily be created or the users can use the ones from a database. Moreover, it is compatible with SCOOP, a Python module allowing concurrent parallel programming on various environments. It allows speeding up the optimization process which can

be crucial when the number of design variables is high. For all those reasons, DEAP was selected in order to implement the sNSGA and NSGA-II.

4 Case study

The algorithms NSGA-II and sNSGA were applied to a case study in order to compare their performance for building design optimization. The optimization problem was carefully chosen based on a current need expressed to our team by the industry and by policy makers to identify the best design tradeoffs between initial CO_2 and operation CO_2 for residential buildings and in particular to develop appropriate methods to do that. In a review of 95 case studies of residential buildings, Chastas et al. (2017) came to the conclusion that the embodied carbon emissions of residential building reflects a share between 9% and 80% of the total life cycle impact. Sartori and Hestnes (2007) showed that design of low-energy buildings is beneficial for life cycle energy demand but increases the embodied energy. In contexts involving high share of renewable energy, Lessard et al. (2017) have concluded via an exhaustive LCA that contributions from materials can account for more than 50% for the environmental impact, the other part coming from the energy consumption (Lessard et al. 2017). Therefore, there is a great need to optimize building designs according to these two objectives (i.e. initial CO_2 and operation CO_2) and as will be explained below, this optimization problem involves hierarchical variables and design variables related to different design disciplines (i.e., architecture, HVAC systems, control, etc.).

4.1 Description of the building

The studied building is located in Holzkirchen, Germany and is named *Twin Houses*. It consists in a typical size single-family home. Measurement data have been compared with a TRNSYS energy model developed in the context of EBC Annex 58, a project leaded by the International Energy Agency, which makes the choice of this specific building very appealing for the purpose of this study. The results have shown a strong correlation between measured and predicted temperatures (Strachan, Scehla, et al. 2016; Strachan P. et al. 2015). In order to optimize the actual building, a set of variables has been selected and added to the TRNSYS model (see Table 1). First, a list of 19 different building envelopes has been selected, each of which is detailed in Table 2. Four types of structure are represented: light-frame wooden wall (LFW), double stud wall (DSW), concrete-block wall and cross-laminated timber wall (CLT). For every type of structure, several levels of insulation are selected with different types of insulation materials. The overall U-values of these building envelopes vary from 0.12 to 0.31 W/m^2-K accordingly. In a similar

manner, a list of 9 windows has been chosen: double glazed windows with low SHGC, double glazed windows with high SHGC and triple glazed windows with high SHGC. In each category, there is three possible frames composition: wood, aluminum or PVC. The overall U-values of these windows vary from 0.59 to 1.29 W/m^2-K and other details are shown in Table 3. The window and wall types are among the independent design variables.

The building dimensions are fixed to 10 m × 10 m with a variable windows-to-wall ratio for each façade. Figure 3 shows the layout of the ground floor, which is the focus of the multi-objective optimization. The design parameters of first floor and of the basement were not optimized in the present study and their zone temperatures were set to be constant (18°C and 21°C respectively). Only the heating need of the ground floor is included in the objective function calculation.

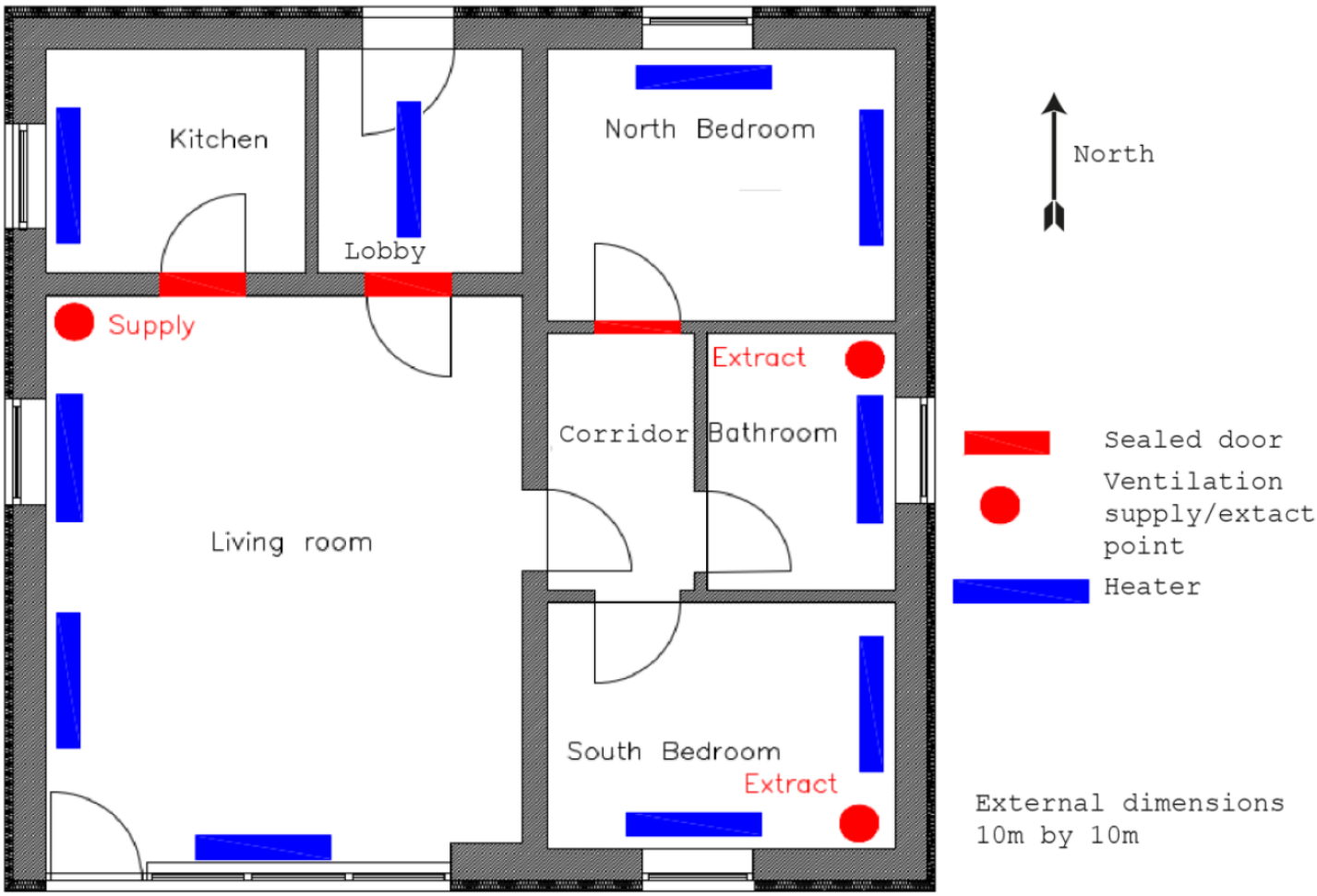

Figure 3: Experimental layout, adaptation from Ref. (Strachan, Scehla, et al. 2016).

4.2 Description of the heating systems

Contrarily to the previously defined variables (independent variables), the ones related to the heating systems are hierarchical variables and a schematic representation is shown in Figure 4. Three possible heating systems (high-level variable) have been chosen in order to meet the heating demand of the ground floor: gas boiler, electrical heat pump, electric heating. The heating capacity of each system is among the low-level variables (q_{bo}, q_{hp}, q_{er}). The heat distribution is provided by water radiators, water convectors or electrical radiators. The number of heaters per zone varies with the size of the room but this number is not optimized in this work and each heater is identical, i.e. same type and same heater rated capacity. Both last characteristics are also low-level design variables. The water radiator capacity is evaluated with a design surface temperature of 70°C and a surrounding air temperature of 20°C. The maximum water flow rate $\dot{m}_w$ is set to limit the water velocity to 1.75 m/s considering a pipe diameter of 12.7 mm (Vedavarz, Kumar, and Hussain 2007; McQuiston, Parker, and Spitler 2005). In the presence of a water radiator, the fluid is heated from a gas boiler or a heat pump[2]. On the other hand, the electrical heater does not require any other system than the baseboard electric heater itself.

The boiler and combustion efficiency are interpolated in a predefined table including several water inlet temperature and three part-load levels. The data were collected from a typical boiler (Thomas H. Durkin 2006). The characteristics remain the same irrespectively to the q_{bo} values. In a similar manner, heating performance data (heating capacity and power consumption) has been defined for the heat pump according to several water inlet temperatures and outdoor air temperatures. The data was collected from the characteristic curves of a LA 16MS heat pump (Dimplex 2008). They have been normalized in order to be scaled according to q_{hp}.

The operation of the heating systems (boiler and heat pump) is done by a proportional controller. This last one sets the water flow rate in the radiator based on the zone ambient temperature T_a. A deviation of 0.3°C from the general set point temperature T_z will bring the water flow rate to its maximum value $\dot{m}_w$ with $T_z = 21°\mathrm{C}$ during the occupied hour and $T_z = T_{sp}$ during the unoccupied hours, with T_{sp} the set point temperature for unoccupied hours. Moreover, the water temperature at the radiator inlet T_w enables or disables the heating system. With the heat pump heating system, a water tank is added and acts as a damper

[2] Type751 and type941 respectively in TRNSYS

between the heat source and the heat sink. It gives more stability to the TRNSYS model. $\dot{m}_w$ and T_w are low-level optimization variables.

It is a common practice to change the value T_w according to the outdoor temperature (outdoor reset temperature) (Arena and Faakye 2013). Therefore, another optimization variable named α represents the monthly water temperature increase rate for every degree less than 0°C (looking at the monthly average outdoor temperature). With $T_w = 40°C$ and $\alpha = 0.5$, an average monthly outdoor temperature of −10°C will increase T_w by 5 °C. An auxiliary electric power unit can be combined with the heat pump to cope with the efficiency drop during cold days. The temperature to enable or disable the auxiliary electric power unit (T_{aux}) is also a low-level optimization variable.

The operation of the electrical heating system is slightly different. Each zone is controlled by a three-stage room thermostat. The radiator is at full capacity with deviation of 0.3°C and at two third of its capacity with a deviation of 0.15°C.

4.3 Summary of the optimization variables

Table 1 presents all the optimization variables with their upper and lower bounds and Figure 4 shows a schematic representation of the hierarchy between variables.

Table 1: List of the optimization variables with their upper and lower bounds.

Variable Symbol	Description and unit []	Variable type	Range
T_{sp}	Temperature set point for the unoccupied hours [°C]	ID+CO	[16 - 21]
Env	Wall envelope type [-]	ID+DI	{0,2,..18}
WWR_N	Windows to wall ratio North façade [-]	ID+CO	[0.1-0.3]
$WWR_{E\ or\ W}$	Windows to wall ratio East façade [-]	ID+CO	[0.1-0.5]
WWR_S	Windows to wall ratio South façade [-]	ID+CO	[0.1-0.7]
Win	Windows type [-]	ID+DI	{0,2,…8}
HS	Heating system type (heat pump = 0, boiler = 1, electric rad. = 2)	HL+DI	{0,1,2}
q_{he}	Heater rated capacity [kW]	LL+CO	[0.25 - 1.5]
T_w	Heater inlet water temperature [°C]	LL+CO	[35 - 65]
q_{bo}, q_{hp}, q_{er}	Heating system rated capacity [kW]	LL+CO	[2.5 - 15]
$\dot{m}_w$	Maximum water flow rate in the heat emitter [kg/hr]	LL+CO	[100 - 800]
HE	Heat emitter type (Convector = 0, Radiator = 1) [-]	LL+DI	{0,1}
α^a	Water temperature increase rate @ heat emitter inlet [°C/°C]	LL+CO	[0 - 1]
q_{aux}	Heat pump auxiliary power [kW]	LL+CO	[0 - 7.5]
T_{aux}	Set point temperature to activate heat pump auxiliary power [°C]	LL+CO	[-10 - 5]

Legend: HL = High-level, LL = Low-level, ID = Independent, DI = Discrete, CO = Continuous

[a] Monthly water temperature increase rate for every degree less than 0°C (looking at the monthly average temperature).

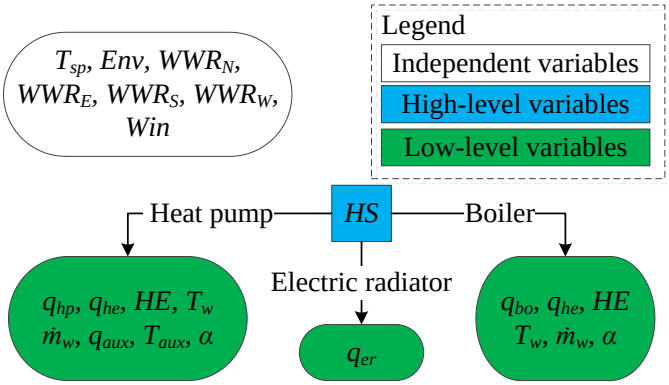

Figure 4: Schematic representation of the hierarchy between variables.

4.4 Objective functions and penalty function

In order to compare the optimization approaches outlined above with a multi-objective BPO problem, a set of two important objectives has been chosen to be minimized simultaneously. They are related to the environmental impact of the building: the initial equivalent emissions of carbon dioxide (CO_{2eq}) due to the building materials and heating systems materials ($CO2_M$) and the CO_{2eq} due to the building energy consumption ($CO2_E$). Although $CO2_M$ and $CO2_E$ have the same units and could have been summed to get a single objective, they were kept here as two independent objectives to minimize for the following reasons. It is more and more reported that in modern buildings, it becomes increasingly hard to make a tradeoff between the environmental impact of the input materials and that of the energy consumption (Sartori and Hestnes 2007; Winther and Hestnes 1999; Gagnon 2015). In high performance buildings, the environmental impact of both materials and energy can have the same order of magnitude and should be considered simultaneously (Chastas et al. 2017). Moreover, $CO2_M$ is the amount of CO_{2eq} emitted in the atmosphere during the construction phase whereas $CO2_E$ is the amount of CO_{2eq} that will be gradually released in the atmosphere over a period of time. The first one thus have immediate impacts on climate changes as opposed to the second one.

$CO2_M$ considers only the materials represented by the design variables. It is assumed that the other materials remain constant, i.e. they are not influenced by the values of the design variables, and thus do not influence the optimization. For the latter ($CO2_E$), it accounts for the CO_{2eq} related with the operation of the heating systems during a period of 15 years. The building location is taken into account for the calculation of both objectives functions. In order to find solutions capable of reaching T_z a penalty function

(*PF*) multiplies the CO_{2eq} when the heating system does not supply enough heat (Hamdy, Hasan, and Siren 2013; Magnier and Haghighat 2010; Nguyen, Reiter, and Rigo 2014):

$$CO2_E = PF \times E_{tot} \times I \times Yr \tag{1}$$

where E_{tot} is the total energy consumption of the heating system, I is the carbon intensity of the energy and Yr is the number of years considered. To determine PF, T_a is first compared to T_z at every time step and for each zone:

$$\Delta T_i = \begin{cases} 1 \text{ if } T_z - T_a \geq 1 \text{ °C} \\ 0 \text{ else} \end{cases} \tag{2}$$

When $\Delta T_i = 1$ the time is cumulated and *RSP* defined as:

$$RSP = \frac{\sum_{i=0}^{i=t} \Delta T_i \Delta t}{\delta} \tag{3}$$

with Δt being the time step and δ the number of hours of the simulation period. *RSP* represents the degree-weighted percentage of time during which the set point is respected during the simulation period (September to April). When RSP was below 3 °C–%, no penalty was applied (i.e., $PF = 1$). Otherwise, *PF* is defined by the following quadratic distance:

$$PF = \frac{1250}{9}(RSP - 0.03)^2 + 1 \tag{4}$$

Such a function has the advantage of penalizing gradually the solutions unable to reach T_z, with a very small penalty when *RSP* is close to 0.03. The coefficients of Eq. (4) have been selected in such a way that *PF* = 1.5 when *RSP* = 0.09. Other possible penalty strategies are detailed by Coello and Lamont (2004) and DEAP project (2017).

4.5 Carbon footprint evaluation method

The environmental performance evaluation is a complex topic and is guided by the norm ISO 14040, which describes the principles and framework for life cycle assessment (LCA). This standard concerns the evaluation of the environmental impacts associated with the entire lifetime of a product, from raw material extraction to disposal or recycling. Nowadays, several tools have been designed to help in performing an LCA for many types of products. A state-of-the-art (Zabalza Bribián, Aranda Usón, and

Scarpellini 2009) presents a list of tools developed specifically for buildings. Recently, Fraunhofer IBP has developed a user-friendly tool currently named *SBS Building Sustainability*[3]. It has been designed for a fast evaluation of the environmental building performance. It allows creating a construction by using data from many European countries from the ESUCO database.

For the purpose of this study, only the impacts from phases A1 to A3 are considered in order to determine $CO2_M$. They include the raw materials supply, transport and manufacturing. SBS also supplies the environmental impacts of a wide variety of heating systems and water radiator. It provides a generic value, not for a specific brand and it changes according to the size of the equipment. Table 2

.....

Table 3 present a summary of the carbon footprint of some of the most impactful optimization variables of the present study.

Table 2: Wall specifications.

No.	Composition	U-value [W/m^2-K]	kg of $CO_{2eq.}$/m^2
0	Concrete block, mineral wool	0.305	189.7
1	Concrete block, mineral wool	0.202	197.0
2	Concrete block, mineral wool	0.151	204.3
3	Concrete block, expanded polystyrene	0.271	183.4
4	Concrete block, expanded polystyrene	0.17	186.0
5	Concrete block, expanded polystyrene	0.131	188.7
6	LFW wall, mineral wool (between studs)	0.320	36.5
7	LFW wall, mineral wool (between studs)	0.231	40.1
8	LFW wall, mineral wool (between studs)	0.181	44.1
9	LFW wall, mineral wool (between studs), EPS (outside)	0.196	41.3
10	LFW wall, mineral wool (between studs), mineral wool (outside)	0.175	46.4
11	LFW wall, mineral wool (between studs), mineral wool (outside)	0.141	52.5
12	LFW wall, mineral wool (between studs), mineral wool (outside)	0.118	58.6
13	CLT, polyisocyanurate	0.167	-9.8
14	CLT, expanded polystyrene	0.187	-27.2
15	CLT, mineral wool	0.177	-15.5
16	DSW (double stud wall) filled with cellulose, EPS (outside)	0.204	-22.7
17	DSW (double stud wall) filled with cellulose, EPS (outside)	0.166	-24.7
18	DSW (double stud wall) filled with cellulose, EPS (outside)	0.140	-26.7

[3] https://www.gabi3.com/Signin.html?locale=en

Table 3: Windows specifications.

No.	Composition	SHGC [-]	U-values* [W/m^2-K]	kgCO_2/ m^2
0	Double glazed LowE Argon, Wooden frame	0.624	1.27	34.12
1	Double glazed LowE Argon, Aluminum frame	0.624	1.27	47.29
2	Double glazed LowE Argon, PVC frame	0.624	1.27	44.80
3	Triple glazed LowE Argon, Wooden frame	0.584	0.59	51.18
4	Triple glazed LowE Argon, Aluminum frame	0.584	0.59	87.81
5	Triple glazed LowE Argon, PVC frame	0.584	0.59	62.53
6	Double glazed LowE Argon & Low SHGC, Wooden frame	0.333	1.29	34.12
7	Double glazed LowE Argon & Low SHGC, Aluminum frame	0.333	1.29	47.29
8	Double glazed LowE Argon & Low SHGC, PVC frame	0.333	1.29	44.80

*Overall U-values including frame.

Environmental impact of energy sources are extracted from the GaBi database SP33. For electricity in Germany in 2017, the global warming potential (GWP 100 years) per kWh is 0.594 kg CO_2/kWh (dataset DE strom mix 1kV-60kV). From the same database, the carbon intensity of natural gas is set to 0.251 CO_2/kWh (GaBi 7 2013).

4.6 Performance metrics for assessment of optimization algorithm

Several performance metrics for the optimization procedures have been reviewed by (Riquelme, Von Lücken, and Baran 2015). In this paper, the three most cited ones have been selected to evaluate the proposed optimization algorithms.

Metric for convergence (Deb and Jain 2002): The first metric $C(P^{(t)})$ is a measure of how close is a set of non-dominated solutions to the Pareto-optimal points P^* (or a reference set). It is computed for each non-dominated set $F^{(t)}$ extracted from every population $P^{(t)}$, with *t* the generation number. For each point *i* in $F^{(t)}$, the normalized Euclidean distance d_i to P^* is computed as follows (for two objective functions):

$$d_i = \min_{j=1}^{|P^*|} \sqrt{\sum_{k=1}^{2} \left(\frac{f_k(i) - f_k(j)}{f_k^{max} - f_k^{min}} \right)^2} \tag{5}$$

$f_k^{\max}$ and $f_k^{\min}$ are the maximum and the minimum values of the kth objectives function (i.e. $CO2_M$ and $CO2_E$) in P^*. $C(P^{(t)})$ is the average of d_i for all points in $F^{(t)}$:

$$C\left(P^{(t)}\right)=\frac{\sum_{i=1}^{\left|F^{(t)}\right|} d_i}{\left|F^{(t)}\right|} \tag{6}$$

In order to get a metric within [0,1], $C(P^{(t)})$ values are normalized by dividing them by $C\left(P^{(t)}\right)_{\max}$.

Metric for diversity (Deb et al. 2002)**:** The second metric Δ measures how much the solutions from a non-dominated set are spread. The following equation is used to determine Δ:

$$\Delta=\frac{d_f+d_l+\sum_{i=1}^{N-1}\left|d_i-\bar{d}\right|}{d_f+d_l+(N-1)\bar{d}} \tag{7}$$

The Euclidian distance d_i is calculated for consecutive points in the selected non-dominated set of solutions with i = 1,2 …, (N-1). N is the number of solutions in the non-dominated set and $\bar{d}$ is the average value of these distances. d_f and d_l are the Euclidian distances between the extreme solutions as shown in Figure 5. If all solutions are equally spread along the front with the existence of the extreme solutions into it, Δ = 0, otherwise, the metric would be greater than 0.

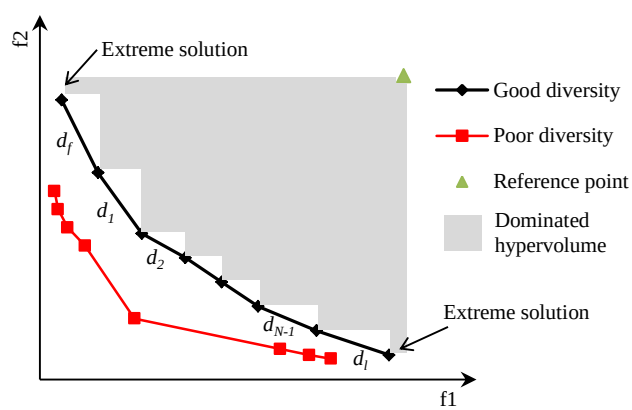

Figure 5: Visual representation of the metric diversity and hypervolume.

Metric for diversity and convergence: A third metric named hypervolume (*HV*), with a purpose similar to that of the convergence metric is described by Zitzler (1999). The output is a unary metric that evaluates how much the objective space is covered (Riquelme, Von Lücken, and Baran 2015). For any non-dominated set $F^{(t)}$ dominating completely the previous one ($F^{(t-1)}$), $HV_t > HV_{t-1}$. A major difference from

$C(P^{(t)})$ is that the evaluation of *HV* occurs between a non-dominated set of solutions and a reference point rather than with P^* (see Figure 5). It makes it sensitive to the choice of this point and the magnitude of each components of $F^{(t)}$ (Auger et al. 2009).

4.7 Description of the algorithm parameters and computer specifications

This section shows the main values of the optimization algorithm parameters, which are summarized in Table 4. Different rules of thumb have been derived over the years to properly set GAs, based on numerical experimentations (Deb et al. 2002; Carlucci et al. 2015; Palonen, Hasan, and Siren 2009; Delgarm et al. 2016). For NSGA-II, the number of variables is 22. Therefore, *Pop* is set to 220 individuals which represents a size of ~10 times the number of design variables. The number of active variables varies in the case of sNSGA. The population size is still set to 220 but P_m varies according to the heating system selection. For example, there are 17 active variables when the heat pump is selected, and thus $P_m = 1/17$.

Table 4: Algorithm parameters

	NSGA-II	sNSGA		
Gen	39	39		
Pop	220	220		
P_c	0.9	0.9		
P_m	0.043	*HS*=1	*HS*=2	*HS*=3
		0.059	0.067	0.111

The energy simulation and the optimization are launched on a Lenovo ThinkPad T450s. It comes with a processor Intel dual core i5-5200U and with 4 GB of RAM memory. The SCOOP module in Python combined with the computer performances allows the evaluation of 3 solutions simultaneously.

5 Comparison of performance between both algorithms

5.1 Graphical analysis of the evolutionary process This first overview shows that both algorithms converge in a similar manner toward P^*. Figure 6 will be used later in Section 6.

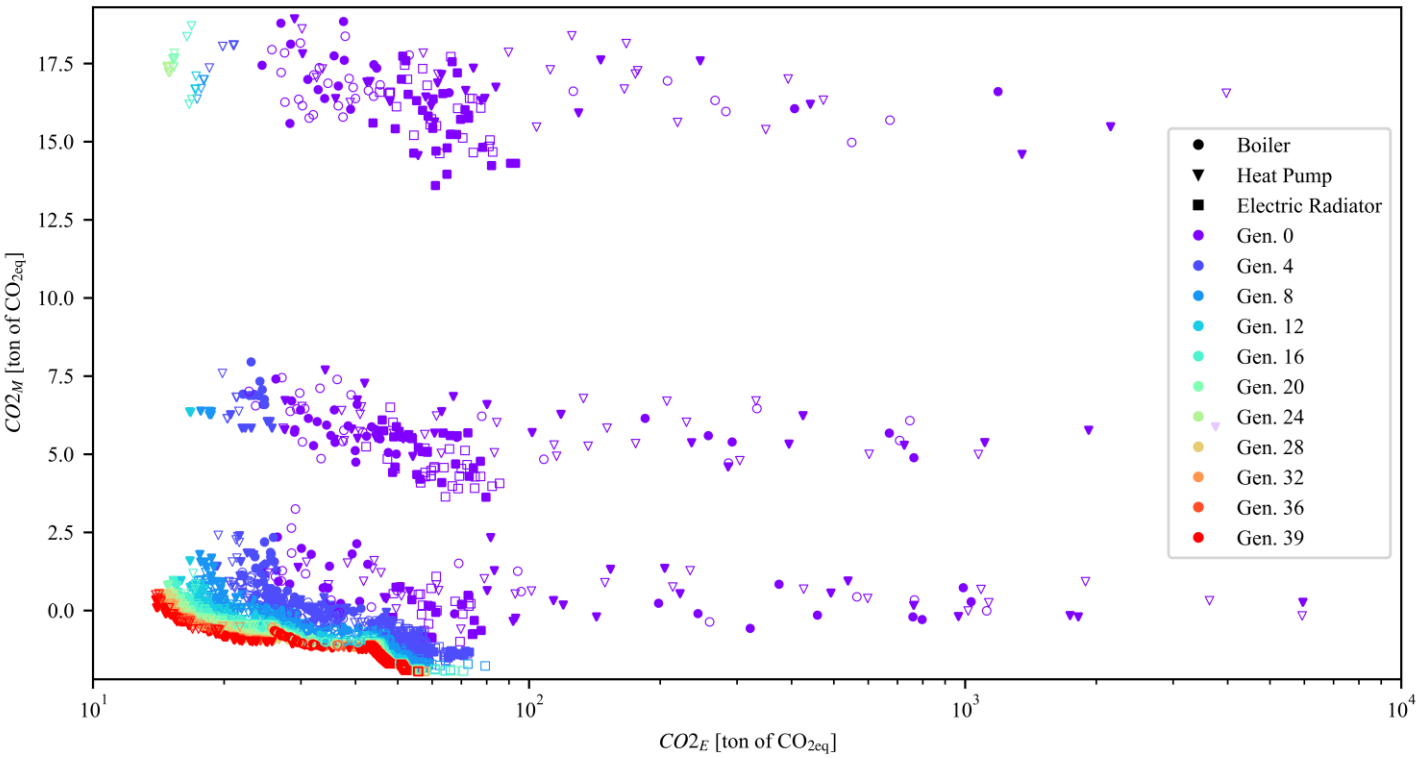

Figure 6: Solutions after each generation for both algorithms and for optimization #1 (filled markers = created solutions from NSGA-II and empty markers = created solutions from sNSGA).

First, the optimization problem described above was solved once with the two proposed algorithms and the results are shown here (optimization #1). Figure 6 shows the population evolution occurring during the optimization process. Each point represents an individual which is a combination of variables resulting from selection, crossover and mutation operators and their position in the figure is related to the values of the fitness functions. This figure gives a quick appreciation of the evolution process. The resulting individuals from NSGA-II are shown by the filled markers and the empty markers show the ones from sNSGA. The color scale is related to the generation number and the marker shape indicates the type of heating system. This first overview shows that both algorithms converge in a similar manner toward P^*. Figure 6 will be used later in Section 6.

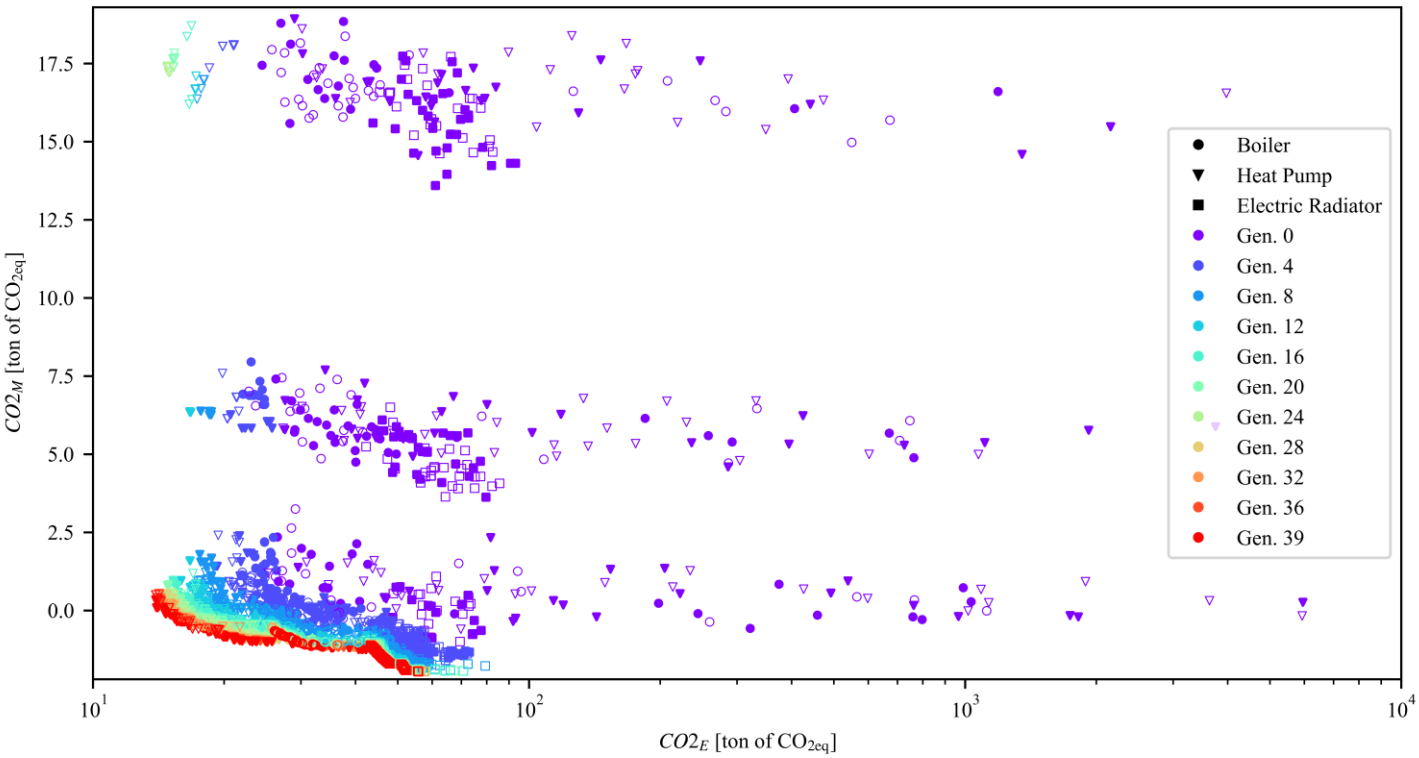

Figure 6: Solutions after each generation for both algorithms and for optimization #1 (filled markers = created solutions from NSGA-II and empty markers = created solutions from sNSGA).

Another interesting aspect of the optimization process is the evolution of the value of the average penalty function (PF_{avg}) over generations. At the beginning of the optimization process, some combinations of variables leading to a high *RSP* value (i.e. designs not respecting the temperature set point) are created and penalized by *PF*. During the optimization process, PF_{avg} tends towards 1 as shown in Figure 7 (i.e., the constraint is respected) and the range of $CO2_E$ values is getting smaller. PF_{avg} for sNSGA is smaller than PF_{avg} for NSGA-II during the optimization. This can be explained by the major difference between the two algorithms: the consideration of the heating system type during the crossover operation process. In the NSGA-II, two individuals with different heating systems can mate and mix all their variables. These situations enhance the chance to create solutions that do not perform well, i.e. solutions that do not success to reach the temperature set point in more than 97% of the time steps.

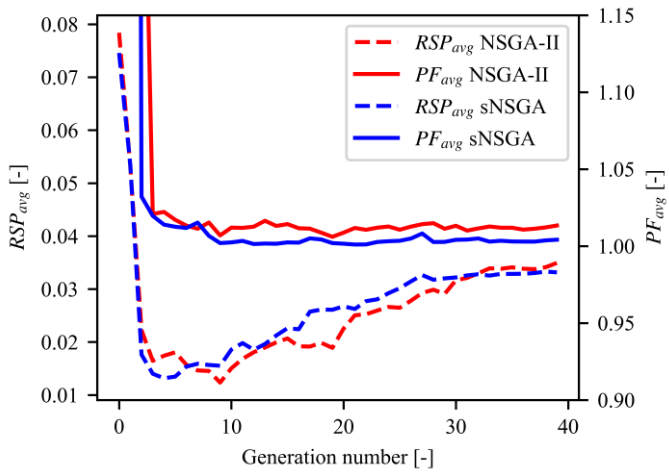

Figure 7: Average of *RSP* and *PF* after each generation for the optimization 1.

The large number of markers in Figure 6 makes it hard to compare the population diversity between both algorithms. An alternative is to plot only the non-dominated solutions after each generation as in Figure 8. One can observe that with NSGA-II, designs with a boiler are eliminated from the non-dominated sets starting at the 20th generation since they are outperformed by designs with a heat pump. With the other algorithm, boilers remain present in the non-dominated set of solutions for a longer number of generations. In fact, it appears that sNSGA was not able to explore some parts of the Pareto front such as one marked by an ellipse in Figure 8. It can be explained by the constraint applied to the crossover with sNSGA. As explained above, when *HS* is different for two individuals selected for a crossover, only the independent variables mate. The NSGA-II mates every individual without any consideration of the high-level variable values as described in Section 3.2. The consequence is that the solutions involving a boiler are rejected faster than with sNSGA, giving more opportunity for individuals with a heat pump heating system to reproduce.

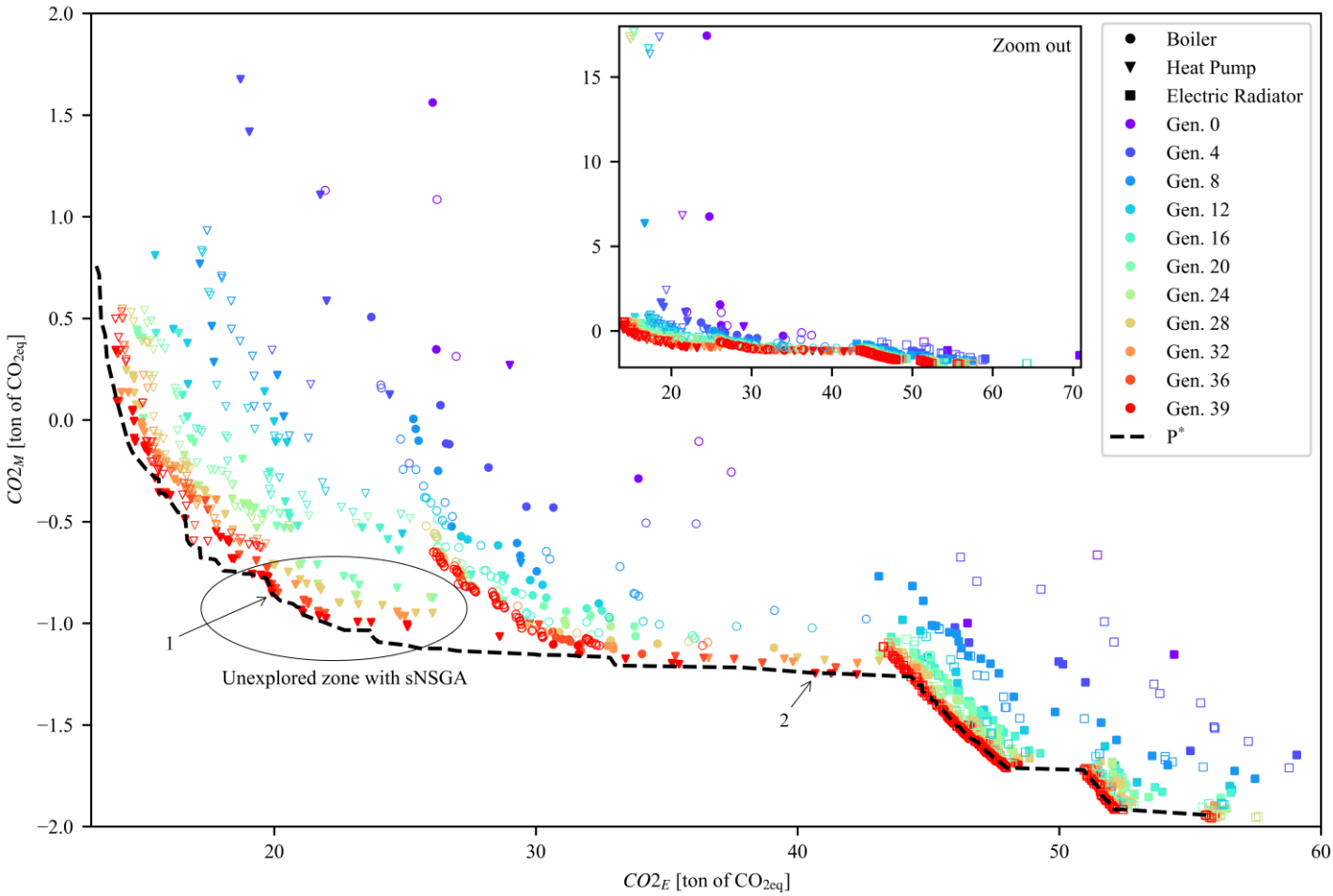

Figure 8: Non-dominated set of solutions after each generation for both algorithms and the optimization #1 (filled markers = solutions from NSGA-II and empty markers = solutions from sNSGA).

As mentioned before, other research (Chen et al. 2008; Manso and Correia 2013; Dasgupta and McGregor 1993; Molfetas 2006; Dasgupta 1994) has suggested that standard GAs can fail in solving deceptive problems because of premature convergence in one area of the design space. However, the opposite was observed with the present test case: the standard NSGA-II appeared to provide a better convergence than the sNSGA specifically adapted to hierarchical variables. On the other hand, an analysis of *RSP* (i.e. the constraint related to the set point temperature) shows that the solutions generated by the sNSGA respect more frequently the condition of $RSP \leq 0.03$. Figure 9 shows a distribution of *RSP* among the solutions for the last 5 generations. It appears that NSGA-II has produced many solutions that did not respect this condition. Moreover, those solutions violating the constraint occur in general with a boiler or a heat pump.

All the solutions defined by an electric radiator, for NSGA-II as well as for sNSGA, respect the *RSP* constraint.

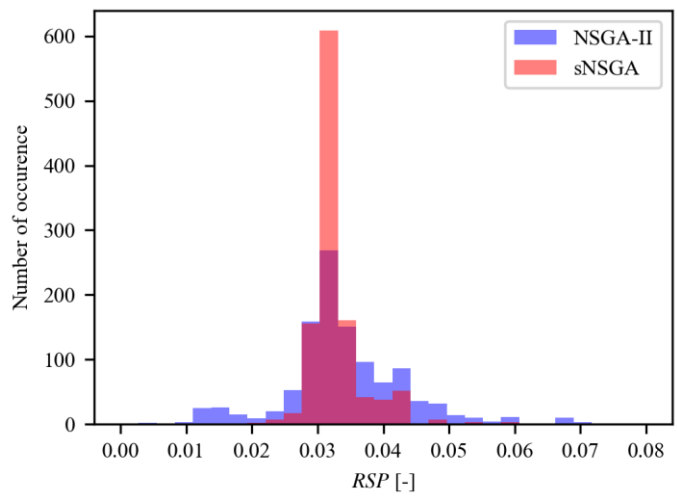

Figure 9: Solutions distribution with respect to *RSP* for both algorithms and the optimization #1 (1100 solutions per algorithm are considered, which were taken from the last 5 generations).

5.2 Algorithm evaluation based on metrics

With stochastic algorithms such as NSGA-II and sNSGA, it is a good practice to repeat several times the optimization procedure. This allows to verify whether an algorithm is consistent and can converge to the same set of solutions, even given its stochastic features. In this case, each algorithm has been repeated four times. All the sets of solutions from each run were combined as follows:

$$P_{tot} = \bigcup_{t=0}^{Gen} P_{NSGA-II}^{(t)} \cup \bigcup_{t=0}^{Gen} P_{sNSGA}^{(t)} \,. \tag{8}$$

to compute the metric values presented in this section.

The metric $C\left(P^{(t)}\right)$ requires Pareto-optimal points P^* for its evaluation, which is not precisely known in this case given the nature of the problem. From P_{tot}, 67 fronts are determined, i.e. the designs are separated into 67 levels of optimality based on non-domination. It can be observed that there is almost no improvement between the successive first ten fronts (the front 0 being the best of all 67 fronts), as shown in Figure 10. Therefore, it was decided to set P^* as being equivalent to front 0. Thereafter, a non-

dominated set of solutions is extracted from $P^{(t)}$ after each generation and $C\left(P^{(t)}\right)$ is computed. Figure 11 shows the evolution of the convergence of both algorithms. The results for each optimization are shown including the average of $C\left(P^{(t)}\right)$ when using NSGA-II or sNSGA.

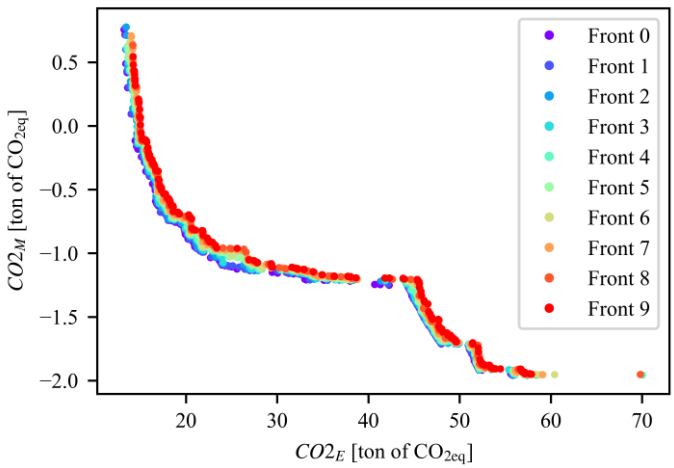

Figure 10: Solutions included in the first 10 levels of optimality fronts in P_{tot}.

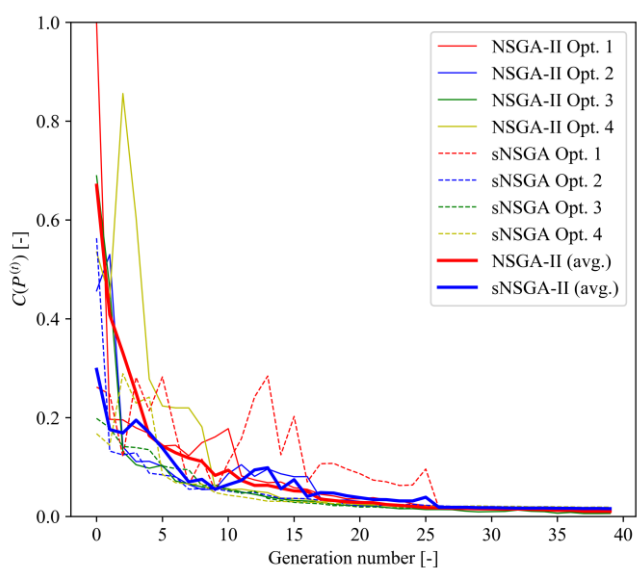

Figure 11: Convergence metric of both algorithms and for each optimization run.

After ~26 generations, both algorithms have converged closely to P^*. Before that point, however, the $C\left(P^{(t)}\right)$ value with sNSGA exhibits several peaks. They are caused by the occurrence of solutions on the

non-dominated set with a specific envelope type. The solutions with concrete blocks or LFW walls have a much higher $CO2_M$ than the ones with a CLT or DSW wall, as mentioned in Section 5.1. P^* is composed of optimal solutions with $CO2_M$ values in the range of [-2.0 – 2.5] ton. Therefore, these few designs with a high $CO2_M$ value affect significantly $C\left(P^{(t)}\right)$. During this test case, they have reappeared more often in sNSGA partly because P_m is higher than in NSGA-II.

The metric Δ shows a similar tendency for both algorithms (Figure 12): it tends to increase over generations and stabilize around generation 26. It means that the diversity degrades during the evolution process, even though the selection based on crowding distance helps to preserve a certain level of diversity in the set of optimal solutions. The analysis presented in Table 5 explains partly why. It shows the evolution of the solutions with respect to the number of heating systems of each type per generation for the optimization #1 (NSGA-II and sNSGA). The proportion of solutions with a heat pump (hp), a boiler (bo) or an electric radiator (er) in the population changes over generation. As observed in Figure 8, the heating system selection, a discrete variable, locates the solutions in specific areas of the fronts and therefore, affects significantly Δ. In fact, large gaps in term of $CO2_E$ and $CO2_M$ exist between those groups of solutions.

Table 5: Percentage of heating system types in the population as a function of the generation for the optimization 1.

	t =	0	9	19	29	39
NSGA-II	bo [%]	33	26	10	9	4
	hp [%]	33	15	30	36	46
	er [%]	34	59	60	55	50
sNSGA	bo [%]	33	29	28	24	26
	hp [%]	34	32	25	18	12
	er [%]	33	39	47	58	62

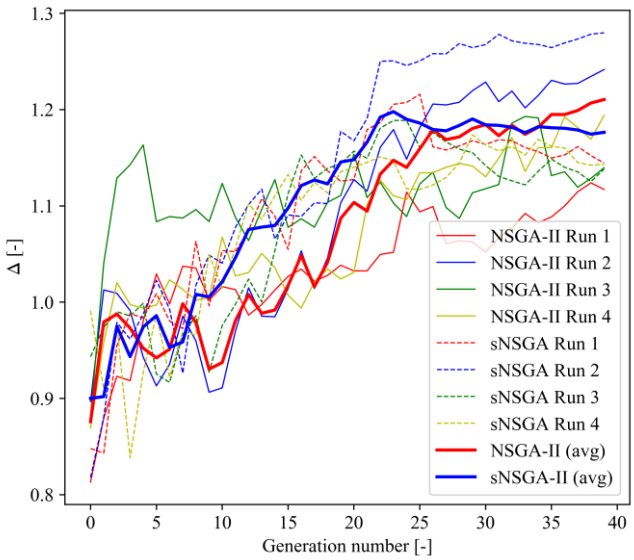

Figure 12: Diversity metric of both algorithms and for each optimization run.

The criterion *HV* introduced above characterizes the quality of the diversity of the non-dominated set of solutions and the convergence at the same time. When *HV* value stabilizes, it means that the non-dominated set of the new generation is close to the last one. For the present case, the reference point is located at $CO2_E = 246.86$ ton and $CO2_M = 17.58$ ton. It corresponds to the only existing point on the front 67, determined from P_{tot}. Figure 13 shows that during the first generations (approximately 10 generations), the optimal solutions progress quickly toward the Pareto front and then move more slowly towards P^*.

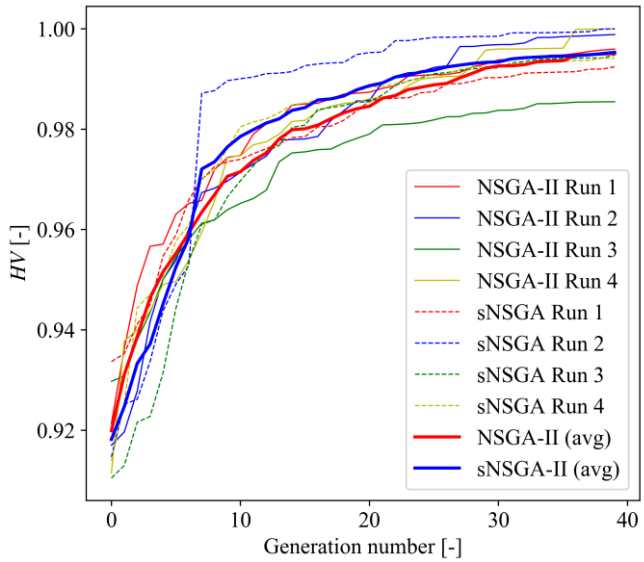

Figure 13: Hypervolume metric of both algorithms and for each optimization run.

The analysis of Sections 5.1 and 5.2 allows concluding that the NSGA-II could properly handle the hierarchical variables in the context of this test case. NSGA-II has converged closer to P^* than sNSGA after 40 generations. In fact, both algorithms perform similarly with minor variations according to the optimization number. Nevertheless, it is important to remember that the system of hierarchical variables was relatively simple, with only one layer of low-level variables. The benefits of sNSGA could be more tangible as the number of hierarchical levels increases.

6 Analysis of the optimal solutions

This section presents and analyzes the main features of the optimal set of non-dominated solutions that were found during our study. This characterizes further the optimization problem addressed here and reveals best design trends and environmental tradeoffs.

A first observation concerns the distribution of the fitness value along the y-axis ($CO2_M$) (see Figure 6). The visible distinct strata or layers of solutions can be explained by the carbon footprint of the envelope type (*Env)*, which accounts for a high fraction of $CO2_M$. *Env* is a discrete variable and there are major gaps between the different families of walls in terms of carbon footprint. Moreover, it can be observed that the designs with concrete blocks or LFW walls were not performing enough to be selected for reproduction after generation 4. Another observation is the effect of the penalty function (*PF)* on $CO2_E$ in Figure 6. It explains why some values of $CO2_E$ are very high in the first generation.

In Figure 8, the solutions tend to be clustered on the Pareto front according to the type of heating systems. This reveals the strong influence of this hierarchical variable on both objectives. The optimal solutions with a heat pump have the highest $CO2_M$ with the lowest $CO2_E$ and the opposite is observed for individuals with an electric radiator. In fact, the electric radiator has a low embodied CO_{2eq} in comparison with a heat pump. However, the heat pump has a higher COP, and therefore, requires less energy to fulfill the heating load than electric radiators. The shape of P^* also illustrates the relevance of certain solutions. In fact, zones where P^* is either vertical or horizontal are zones where a large degradation of one objective is followed by a very small improvement of the other and vice versa. For example, the selection of solution #2 over #1 in Figure 8 means that the $CO2_E$ value is doubled for an insignificant reduction of $CO2_M$.

Analyzing individually each solution on P^* can be time consuming. In this case, there are 987 solutions on P^*, 71.3% of which were found by the sNSGA. From the 987 solutions, many are almost identical as revealed by Figures 14-16. Therefore, the number could be reduced considerably if similar solutions were clustered together. There are 698 solutions that use an electric radiator (group “er”), 274 that use a heat pump (group “hp”) and only 15 that use a boiler (“bo”). Most of the electric radiator optimal solutions were found by the sNSGA although the solutions found by the NSGA-II were not far behind. The selection of a heating system dictates a part of the other design variables (low-level variables), therefore, an analysis of each group was performed.

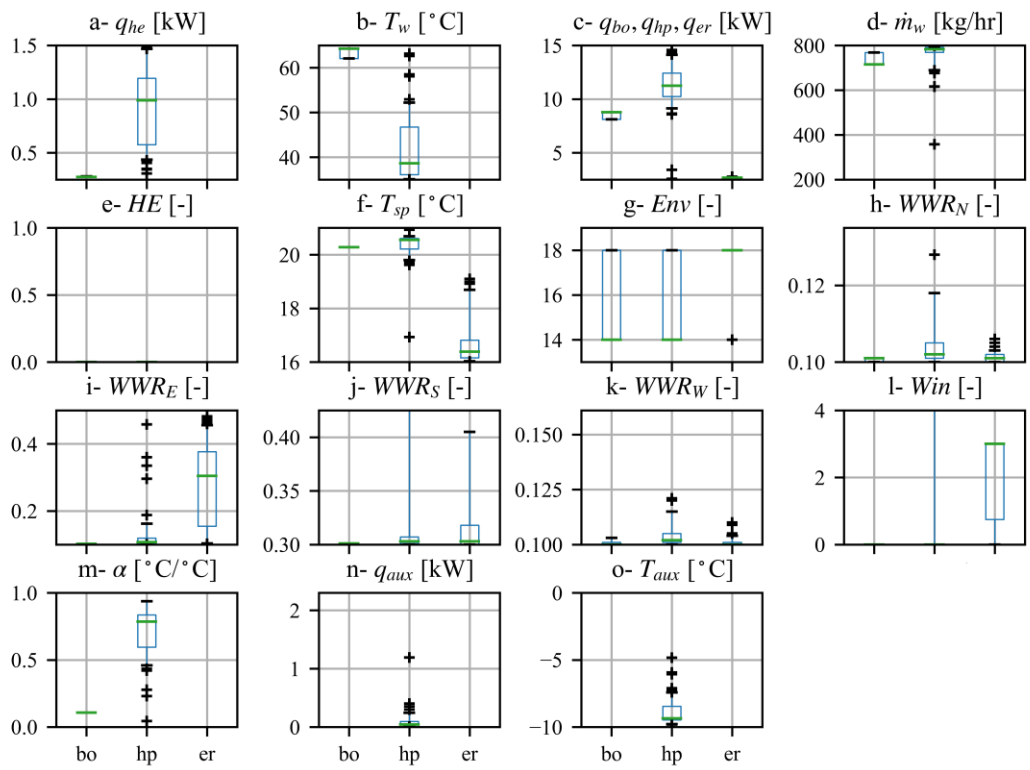

Figure 14: Boxplot for each design variable in P^*. The green line represents the median, the side of the box the first and third quartile and the whiskers the 5 and 95 percentiles. The sign "+" indicates data outside the whiskers.

Figure 14 shows the boxplots for each design variable. First, one finds that some variables are essentially identical for all the solutions: Env = 18 (rarely Env = 14) and HE = 0 in more than 95% of the optimal solutions. In other words, the optimal solutions are all defined by a double stud wall filled with cellulose, and in the presence of a heat pump or a boiler, the heat emitter is always a convector. The analysis of the continuous variables WWR_N and WWR_W also demonstrates that they should take most of the time the lowest possible value based on the two objectives. A small WWR means that the opaque portion of the envelope is larger. As mentioned above, in the optimal set of solutions, the opaque envelope 18 was almost always selected. It is mainly composed of wooden materials and thus acts as a carbon sink. In other words, smaller WWR values help to reduce $CO2_M$. Similarly, the windows are responsible for significant heat losses, which increases the energy consumption. Therefore, it is preferred to have a WWR as small as possible for reducing $CO2_E$.

Large differences between the median values of the rated heating capacity of each heating system are noted in Fig. 14. The efficiency drop of the heat pump at low temperature leads to an oversizing in

comparison with the boiler. On the other hand, the stable efficiency of the electric radiator keeps their rated heating capacity to low values.

Usage of parallel coordinates is a good alternative to visualize multidimensional data sets and highlight the relations between optimal variables (see Figures 15-16). In this representation, each design is represented by a series of continuous line segments indicating the values of objective functions and design variables. In Figures 15-16, a color scale related to the objective function $CO2_E$ was applied to facilitate the interpretation of the figure.

Figure 15 shows all the optimal solutions on P^* with an electric radiator. It can be seen that solutions with a low $CO2_E$ values (44.34 to 48 ton of CO_{2eq}) use the window type 3 (triple glazed LowE argon, wooden frame) as defined in Table 3 whereas the optimal solutions with a $CO2_E$ value between 51 and 52 ton of CO_{2eq} use a window type 0 (double glazed LowE argon, wooden frame). In the first case, the selected window reduces the heat losses because of its low U-value but increases $CO2_M$ because of its high carbon footprint. There is also an observable relation between T_{sp} and Win: solutions with a triple glazed window have a temperature set point for the unoccupied hours lower than solutions with a double glazed window.

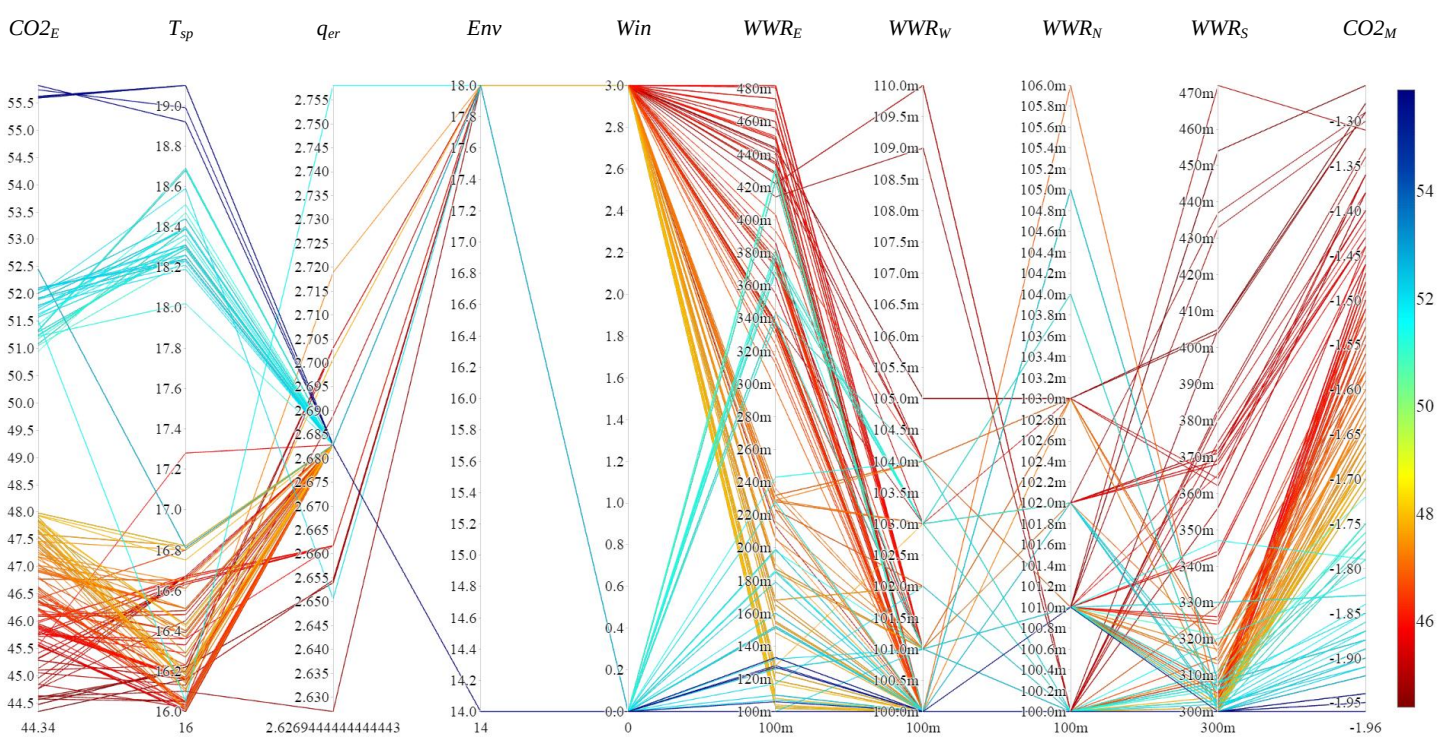

Figure 15: Parallel coordinates for all the solutions with electric radiator in P^*.

The analysis of the heat pump solutions also highlights relations between design variables and the objectives functions (Figure 16). Because the variability of Env, q_{aux}, T_{aux}, and $WWRs$ was small, they are not reported in Figure 16. Increasing the heat emitter capacity q_{he} leads to a reduction of the energy consumption (and thus $CO2_E$) while increasing $CO2_M$. At the same time, the reduction of q_{he} in optimal solutions is compensated by the increase of T_w. Similarly, when the heat pump capacity is low (~ 3kW), the supply water temperature T_w is high (~60°C). These solutions lead to higher energy consumption. However, the usage of smaller mechanical equipment leads to a lower embodied energy (and thus a lower $CO2_M$). Finally, among almost all solutions, the parameters α, $\dot{m}_w$ and T_{sp} exhibit a small variability.

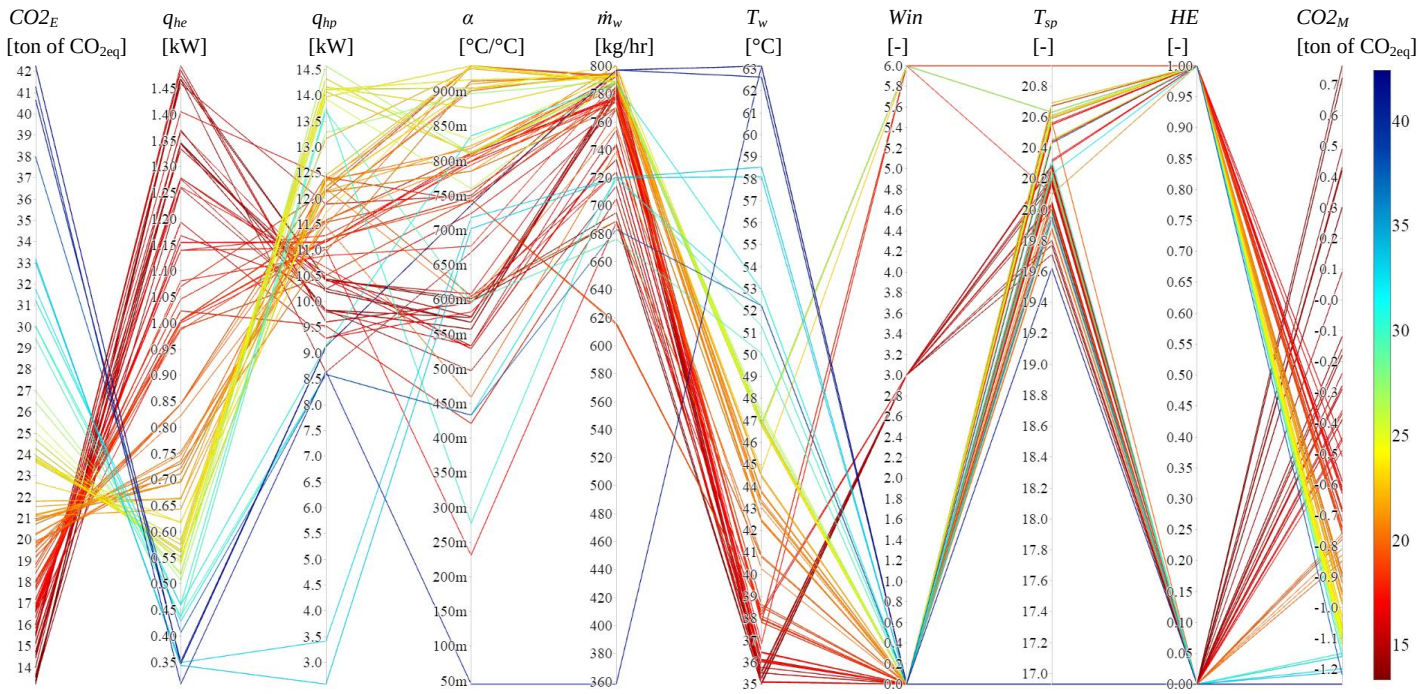

Figure 16: Parallel coordinates for all the solutions with heat pump in P^*.

Solutions with gas boilers are not analyzed here since they were very few in P^*.

7 Conclusions

This paper had two main objectives: (i) to select an algorithm dealing with hierarchical variables in a real BPO multi-objective problem, and (ii) to provide the set of optimal solutions minimizing simultaneously the CO_{2eq} emissions related to the materials and to the energy consumption. A house located in Germany

was chosen as the test case building since it was already well documented. More specifically, this paper compared the well-known NSGA-II with a new algorithm dealing with hierarchical variables (sNSGA). Different performance criteria were defined and used to compare the two algorithms. Although some differences were noted in terms of convergence rate and level of diversity in the evolving population, both algorithms performed similarly for the selected design variables and test case. It was found that NSGA-II could properly handle hierarchical variables although the core of the algorithm does not pay any attention to this particularity. This test case did not prove that a sGA was needed for such a problem.

Different modifications could be implemented to sNSGA to improve its performance such as a dual mutation rate, decreasing during the evolution process. Moreover, the high-level variables could be allowed to mate. In this case, this could have accelerated the elimination of solutions using a boiler as a heating system. In further studies, one could include other objective functions, such as the cost of construction and operation or thermal comfort. Some variables in the present study did not exhibit a contradictory effect in terms of $CO2_E$ and $CO2_M$, but this could prove different with more objectives. Furthermore, it would be interesting to test the algorithms in a problem involving more levels of hierarchy in the variables. sNSGA could potentially become more attractive in such cases.

The test case has shown an existing trade-off between solutions in order to minimize simultaneously $CO2_E$ and $CO2_M$. The solutions with an electric radiator have the lowest $CO2_M$ but the highest $CO2_E$. The contrary occurs for solutions with a heat pump: the COP of the heat pump reduces the energy consumption and thus $CO2_E$, but heat pump solutions tend to have a higher value of $CO2_M$. Most optimal solutions relied on an electrical heater (80 %) or a heat pump (18.6 %). The selected opaque envelope was almost always the double stud wall filled with cellulose. In rare situations, the CLT wall was selected. As mentioned before, other objective functions, such as cost and thermal comfort could be considered. In order to help decision makers to select the best solution among a set of several hundreds, a multi-criteria analysis should follow the multi-objective optimization. Furthermore, it would be interesting to determine to what extent the optimal solutions are affected by the location of the building.

BPO has a great potential to improve the building design process and reduce the carbon footprint of buildings. Nevertheless, it should be noted that the average computing time for one generation was 6 hours in the present work. As a result, a total of 10 days of computation was required to complete the optimization with one algorithm. Therefore, finding ways to reduce computational and implementation time would be helpful for such techniques to be fully deployed in the industry.

Acknowledgements

Natural Sciences and Engineering Research Council of Canada (NSERC) and Fonds de recherche du Québec - Nature et technologies (FRQNT). Details have been removed for anonymity.

References

Agrawal, Ram Bhushan, K. Deb, and R. B. Agrawal. 1995. "Simulated Binary Crossover for Continuous Search Space." *Complex Systems* 9 (2): 115–148.

Alajmi, Ali, and Jonathan Wright. 2014. "Selecting the Most Efficient Genetic Algorithm Sets in Solving Unconstrained Building Optimization Problem." *International Journal of Sustainable Built Environment* 3 (1): 18–26. doi:10.1016/j.ijsbe.2014.07.003.

American Society of Heating Refrigerating and Air-Conditioning Engineers. 2004. *ANSI/ ASHRAE Standard 55 (2004): Thermal Comfort Conditions for Human Occupancy*. Atlanta, GA: ASHRAE.

Arena, L, and O Faakye. 2013. "Optimizing Hydronic System Performance in Residential Applications." Denver, USA: The National Renewable Energy Laboratory.

Attia, Shady, Mohamed Hamdy, William O'Brien, and Salvatore Carlucci. 2013. "Assessing Gaps and Needs for Integrating Building Performance Optimization Tools in Net Zero Energy Buildings Design." *Energy and Buildings* 60 (May): 110–24. doi:10.1016/j.enbuild.2013.01.016.

Auger, Anne, Johannes Bader, Dimo Brockhoff, and Eckart Zitzler. 2009. "Theory of the Hypervolume Indicator: Optimal μ-Distributions and the Choice of the Reference Point." In *Proceedings of the Tenth ACM SIGEVO Workshop on Foundations of Genetic Algorithms*, 87–102. FOGA '09. New York, NY, USA: ACM. doi:10.1145/1527125.1527138.

Bichiou, Youssef, and Moncef Krarti. 2011. "Optimization of Envelope and HVAC Systems Selection for Residential Buildings." *Energy and Buildings* 43 (12): 3373–82. doi:10.1016/j.enbuild.2011.08.031.

Brownlee, Alexander E.I., Jonathan A. Wright, and Monjur M. Mourshed. 2011. "A Multi-Objective Window Optimisation Problem." In *Proceedings of the 13th Annual Conference Companion on Genetic and Evolutionary Computation*, 89–90. New York, NY, USA: ACM. doi:10.1145/2001858.2001910.

Caldas, L. G., and L. K. Norford. 2003. "Genetic Algorithms for Optimization of Building Envelopes and the Design and Control of HVAC Systems." *Journal of Solar Energy Engineering* 125 (3): 343. doi:10.1115/1.1591803.

Carlucci, Salvatore, Giulio Cattarin, Francesco Causone, and Lorenzo Pagliano. 2015. "Multi-Objective Optimization of a Nearly Zero-Energy Building Based on Thermal and Visual Discomfort Minimization Using a Non-Dominated Sorting Genetic Algorithm (NSGA-II)." *Energy and Buildings* 104 (October): 378–94. doi:10.1016/j.enbuild.2015.06.064.

Chastas, Panagiotis, Theodoros Theodosiou, Karolos J. Kontoleon, and Dimitrios Bikas. 2017. "Normalising and Assessing Carbon Emissions in the Building Sector: A Review on the Embodied CO 2 Emissions of Residential Buildings." *Building and Environment*.

Chen, Yang, Jinglu Hu, Kotaro Hirasawa, and Songnian Yu. 2008. "Solving Deceptive Problems Using a Genetic Algorithm with Reserve Selection." In *Evolutionary Computation, 2008. CEC 2008.(IEEE World Congress on Computational Intelligence). IEEE Congress On*, 884–889. IEEE. http://ieeexplore.ieee.org/abstract/document/4630900/.

Coello, Carlos A. Coello, and Gary B. Lamont. 2004. *Applications of Multi-Objective Evolutionary Algorithms*. World Scientific.

Dasgupta, Dipankar. 1994. “Handling Deceptive Problems Using a Different Genetic Search.” In *Evolutionary Computation, 1994. IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference On*, 807–811. IEEE. http://ieeexplore.ieee.org/abstract/document/349952/.

Dasgupta, Dipankar, and Douglas R. McGregor. 1993. “SGA: A Structured Genetic Algorithm.” Glasgow: Department of Computer Science, University of Stratchclyde.

DEAP project. 2017. “DEAP User’s Guide.” http://deap.readthedocs.io/en/master/.

Deb, Kalyanmoy, and Sachin Jain. 2002. “Running Performance Metrics for Evolutionary Multi-Objective Optimization.” Technical report 2002004. KanGAL Report. Kanpur, India: Indian Institute of Technology Kanpur. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.9.159.

Deb, Kalyanmoy, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. 2002. “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II.” *IEEE Transactions on Evolutionary Computation* 6 (2): 182–197.

Delgarm, Navid, Behrang Sajadi, Saeed Delgarm, and Farshad Kowsary. 2016. “A Novel Approach for the Simulation-Based Optimization of the Buildings Energy Consumption Using NSGA-II: Case Study in Iran.” *Energy and Buildings* 127 (September): 552–60. doi:10.1016/j.enbuild.2016.05.052.

Dimplex. 2008. “What Are the Benefits of a Heat Pump?” Technical report. Germany: Dimplex. http://www.dimplex.de/fileadmin/dimplex/downloads/planungshandbuecher/en/dimplex_php_heating_gb_102008.pdf.

Evins, Ralph. 2010. “Configuration of a Genetic Algorithm for Multi-Objective Optimisation of Solar Gain to Buildings.” In *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation*, 2003–2006. ACM. http://dl.acm.org/citation.cfm?id=1830845.

Evins, Ralph. 2013. “A Review of Computational Optimisation Methods Applied to Sustainable Building Design.” *Renewable and Sustainable Energy Reviews* 22 (June): 230–45. doi:10.1016/j.rser.2013.02.004.

Evins, Ralph. 2015. “Multi-Level Optimization of Building Design, Energy System Sizing and Operation.” *Energy* 90 (October): 1775–89. doi:10.1016/j.energy.2015.07.007.

Evins, Ralph. 2016. “A Bi-Level Design and Operation Optimization Process Applied to an Energy Centre.” *Journal of Building Performance Simulation* 9 (3): 255–71. doi:10.1080/19401493.2015.1045034.

GaBi 7. 2013. *GaBi 7* (version Construction Database (SP33)). Software Und Datenbanken Zur Ganzheitlichen Bilanzierung. Echterdingen, Germany: Thinkstep GaBi.

Gagnon, Richard. 2015. “Le Québec Est-Il En Mesure d’atteindre Ses Objectifs de Réduction de GES d’ici 2020? État Des Lieux Du Secteur Du bâTiment et Des Besoins Énergétiques En Milieu Urbain.” Montréal, Canada.

Goldberg, David E., Kalyanmoy Deb, and James H. Clark. 1991. “Genetic Algorithms, Noise, and the Sizing of Populations.” *Urbana* 51: 61801.

Gossard, D., B. Lartigue, and F. Thellier. 2013. “Multi-Objective Optimization of a Building Envelope for Thermal Performance Using Genetic Algorithms and Artificial Neural Network.” *Energy and Buildings* 67 (December): 253–60. doi:10.1016/j.enbuild.2013.08.026.

Hamdy, Mohamed, Ala Hasan, and Kai Siren. 2011a. “Applying a Multi-Objective Optimization Approach for Design of Low-Emission Cost-Effective Dwellings.” *Building and Environment* 46 (1): 109–23. doi:10.1016/j.buildenv.2010.07.006.

Hamdy, Mohamed, Ala Hasan, and Kai Siren. 2011b. "Impact of Adaptive Thermal Comfort Criteria on Building Energy Use and Cooling Equipment Size Using a Multi-Objective Optimization Scheme." *Energy and Buildings* 43 (9): 2055–67. doi:10.1016/j.enbuild.2011.04.006.

Hamdy, Mohamed, Ala Hasan, and Kai Siren. 2013. "A Multi-Stage Optimization Method for Cost-Optimal and Nearly-Zero-Energy Building Solutions in Line with the EPBD-Recast 2010." *Energy and Buildings* 56 (January): 189–203. doi:10.1016/j.enbuild.2012.08.023.

Hutter, Frank, Holger H. Hoos, and Kevin Leyton-Brown. 2011. "Sequential Model-Based Optimization for General Algorithm Configuration." In *Learning and Intelligent Optimization*, edited by Carlos A. Coello Coello, 6683:507–23. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-25566-3_40.

Lessard, Yannick, Chirjiv Anand, Pierre Blanchet, Caroline Frenette, and Ben Amor. 2017. "LEED v4: Where Are We Now? Critical Assessment through the LCA of an Office Building Using a Low Impact Energy Consumption Mix: Critical Life Cycle Assessment of LEED V4." *Journal of Industrial Ecology*, September. doi:10.1111/jiec.12647.

López-Ibáñez, Manuel, Jérémie Dubois-Lacoste, Leslie Pérez Cáceres, Mauro Birattari, and Thomas Stützle. 2016. "The Irace Package: Iterated Racing for Automatic Algorithm Configuration." *Operations Research Perspectives* 3: 43–58. doi:10.1016/j.orp.2016.09.002.

Magnier, Laurent, and Fariborz Haghighat. 2010. "Multiobjective Optimization of Building Design Using TRNSYS Simulations, Genetic Algorithm, and Artificial Neural Network." *Building and Environment* 45 (3): 739–46. doi:10.1016/j.buildenv.2009.08.016.

Manso, António Manuel Rodrigues, and Luís Miguel Parreira Correia. 2013. "A Multiset Genetic Algorithm for the Optimization of Deceptive Problems." In *Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation*, 813–820. ACM. http://dl.acm.org/citation.cfm?id=2463471.

McQuiston, Faye C., Jerlad D. Parker, and Jeffrey D. Spitler. 2005. *Heating, Ventilating, and Air Conditioning, Analysis and Design*. Sixth Edition. Wiley.

Molfetas, Angelos. 2006. "Multiple Control Levels in Structured Genetic Algorithms." Sydney, Australia: University of Western Sydney. http://researchdirect.westernsydney.edu.au/islandora/object/uws%3A2478/.

Nguyen, Anh-Tuan, Sigrid Reiter, and Philippe Rigo. 2014. "A Review on Simulation-Based Optimization Methods Applied to Building Performance Analysis." *Applied Energy* 113 (January): 1043–58. doi:10.1016/j.apenergy.2013.08.061.

Palonen, Matti, Ala Hasan, and Kai Siren. 2009. "A Genetic Algorithm for Optimization of Building Envelope and HVAC System Parameters." In *Proc. Of the 11th IBPSA Conference, Glasgow, Scotland*. https://www.researchgate.net/profile/Matti_Palonen/publication/252429941_A_genetic_algorithm_for_optimization_of_building_envelope_andhvac_system_parameters/links/565ecb3c08ae1ef92983ebb0.pdf.

Parmee, IC. 1998. "Evolutionary and Adaptive Strategies for Efficient Search across Whole System Engineering Design Hierarchies." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 12 (5): 431–45.

Preechakul, Chirdpong, and Soorathep Kheawhom. 2009. "Modified Genetic Algorithm with Sampling Techniques for Chemical Engineering Optimization." *Journal of Industrial and Engineering Chemistry* 15 (1): 110–18. doi:10.1016/j.jiec.2008.09.003.

Rafiq, MY, JD Mathews, and GN Bullock. 2003. “Conceptual Building Design - Evolutionary Approach.” *JOURNAL OF COMPUTING IN CIVIL ENGINEERING* 17 (3): 150–58. doi:10.1061/(ASCE)0887-3801(2003)17:3(150).

Raich, Anne M., and Jamshid Ghaboussi. 2000. “Evolving Design Solutions Using an Unstructured, Multi-Objective Problem Domain.” *Urbana* 51: 61801–2397.

Rainville, De, Félix-Antoine Fortin, Marc-André Gardner, Marc Parizeau, Christian Gagné, and others. 2012. “Deap: A Python Framework for Evolutionary Algorithms.” In *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation*, 85–92. ACM. http://dl.acm.org/citation.cfm?id=2330799.

Riquelme, Nery, Christian Von Lücken, and Benjamin Baran. 2015. “Performance Metrics in Multi-Objective Optimization.” In *Computing Conference (CLEI), 2015 Latin American*, 1–11. IEEE. http://ieeexplore.ieee.org/abstract/document/7360024/.

Sartori, I., and A.G. Hestnes. 2007. “Energy Use in the Life Cycle of Conventional and Low-Energy Buildings: A Review Article.” *Energy and Buildings* 39 (3): 249–57. doi:10.1016/j.enbuild.2006.07.001.

Seshadri, Aravind. 2006. “A Fast Elitist Multiobjective Genetic Algorithm: NSGA-II.” *MATLAB Central* 182. https://web.njit.edu/~horacio/Math451H/download/Seshadri_NSGA-II.pdf.

Strachan P., Monari F., Kersken M., and Heusler I. 2015. “IEA Annex 58: Full-Scale Empirical Validation of Detailed Thermal Simulation Programs.” *Energy Procedia* 78 (November): 3288–93. doi:10.1016/j.egypro.2015.11.729.

Strachan, Paul, Katalin Scehla, Matthias Kersken, Hugo Heusler, and KU Leuven. 2016. “Empirical Validation of Common Building Energy Simulation Models Based on in Situ Dynamic Data.” *IEA*, EBC Annex 58, Report of Subtask 4a, . http://www.iea-ebc.org/projects/completed-projects/ebc-annex-58/.

Strachan, Paul, Katalin Svehla, Ingo Heusler, and Matthias Kersken. 2016. “Whole Model Empirical Validation on a Full-Scale Building.” *Journal of Building Performance Simulation* 9 (4): 331–50. doi:10.1080/19401493.2015.1064480.

Thomas H. Durkin. 2006. “Boiler System Efficiency.” *ASHRAE Journal* Vol. 48 (July): 51–57.

Tiwari, Ashutosh, and Rajkumar Roy. 2002. “Variable Dependence Interaction and Multi-Objective Optimisation.” In *Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation*, 602–609. GECCO’02. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. http://dl.acm.org/citation.cfm?id=2955491.2955582.

Tuhus-Dubrow, Daniel, and Moncef Krarti. 2010. “Genetic-Algorithm Based Approach to Optimize Building Envelope Design for Residential Buildings.” *Building and Environment* 45 (7): 1574–81. doi:10.1016/j.buildenv.2010.01.005.

Vedavarz, Ali, Sunil Kumar, and Muhammed Iqbal Hussain. 2007. *HVAC - The Handbook of Heating, Ventilation and Air Conditioning for Design and Implementation*. http://app.knovel.com/hotlink/toc/id:kpHVACTHH2/hvac-handbook-heating/hvac-handbook-heating.

Verbeeck, G., and H. Hens. 2007. “Life Cycle Optimization of Extremely Low Energy Dwellings.” *Journal of Building Physics* 31 (2): 143–77. doi:10.1177/1744259107079880.

Wang, Weimin. 2005. “A Simulation-Based Optimization System for Green Building Design.” Montréal, Québec, Canada: Concordia University. http://spectrum.library.concordia.ca/8272/.

Wang, Weimin, Hugues Rivard, and Radu Zmeureanu. 2005. “An Object-Oriented Framework for Simulation-Based Green Building Design Optimization with Genetic Algorithms.” *Advanced Engineering Informatics* 19 (1): 5–23. doi:https://doi.org/10.1016/j.aei.2005.03.002.

Wang, Weimin, Radu Zmeureanu, and Hugues Rivard. 2005. “Applying Multi-Objective Genetic Algorithms in Green Building Design Optimization.” *Building and Environment* 40 (11): 1512–25. doi:10.1016/j.buildenv.2004.11.017.

Winther, Beate Nemeth, and Anne Grete Hestnes. 1999. “Solar versus Green: The Analysis of a Norwegian Row House.” *Solar Energy* 66 (6): 387–393.

Wright, Jonathan, and Ali Alajmi. 2016. “Efficient Genetic Algorithm Sets for Optimizing Constrained Building Design Problem.” *International Journal of Sustainable Built Environment* 5 (1): 123–31. doi:10.1016/j.ijsbe.2016.04.001.

Wright, Jonathan, and Raziyeh Farmani. 2001. “The Simultaneous Optimization of Building Fabric Construction, HVAC System Size, and the Plant Control Strategy.” In , 8. Rio de Janeiro, Brazil: Building simulation.

Zabalza Bribián, Ignacio, Alfonso Aranda Usón, and Sabina Scarpellini. 2009. “Life Cycle Assessment in Buildings: State-of-the-Art and Simplified LCA Methodology as a Complement for Building Certification.” *Building and Environment* 44 (12): 2510–20. doi:10.1016/j.buildenv.2009.05.001.

Zitzler, Eckart. 1999. *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*.

Zitzler, Eckart, Kalyanmoy Deb, and Lothar Thiele. 2000. “Comparison of Multiobjective Evolutionary Algorithms: Empirical Results.” *Evolutionary Computation* 8 (2): 173–195.