Holistic optimization of buildings based on the evaluation of annual performances from short simulation sequences

Hasan Sayegh

UNIVERSITÉ SAVOIE MONT BLANC

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ SAVOIE MONT BLANC

Spécialité : **Génie civil et sciences de l'habitat**

Arrêté ministériel : 25 Mai 2016

Présentée par

Hasan SAYEGH

Thèse dirigée par **Gilles FRAISSE** et **Etienne WURTZ**
encadrée par **Antoine LECONTE** et **Simon ROUCHIER**

préparée au sein du **laboratoires LISE du CEA** et **LOCIE de l'Université Savoie Mont Blanc**
dans **l'École Doctorale SISEO**

Holistic optimization of buildings based on the evaluation of annual performances from short simulation sequences

Thèse soutenue publiquement le **« 3 Décembre 2020 »**, devant le jury composé de :

Mrs, Evelyne, LUTTON
Directrice de recherche, INRAE, Président
Mr, Bruno, PEUPORTIER
Directeur de recherche, Ecole de Mines de Paris, Rapporteur,
Mr, Jean-Jacques, ROUX
Professeur, INSA Lyon, Rapporteur
Mr, Jean-Michel, RENEAUME
Professeur, Université de Pau et des pays de l'Adour, Examinateur
Mr, Gilles, FRAISSE
Professeur, Université Savoie Mont Blanc, Directeur de thèse
Mr, Etienne, WURTZ
Directeur de recherche, CEA, Codirecteur de thèse
Mr, Antoine, LECONTE
Docteur, CEA, Encadrant
Mr, Simon, ROUCHIER
Maitre de conférence, Université Savoie Mont Blanc, Co encadrant

Abstract

Performing global approach studies on buildings, which take into consideration both the envelope and the connected systems, lead to the complexity of models under study. Simulation of such models may lead to high computational time expenses. Usually, simplified or surrogate models instead of detailed ones are used to avoid this issue. A global approach based on the reduction of input data profiles rather than the model itself is a current case of interest. The approach evaluates annual performances of a model starting from a short simulation sequence of typical selected days instead of complete data profiles.

After presenting and analyzing the methods used in the literature for typical day selection, the thesis presents a new iterative approach with an embedded grouping algorithm. The new algorithm, called TypSS (Typical Short Sequence) Algorithm, creates and enhances iteratively a short simulation sequence of typical days based on target criteria reflecting the annual performances of a model. The algorithm was applied on a detailed building model and led to much faster simulations while obtaining results of high correlation with the reference ones. Results were also compared to an iterative and a clustering approach used for day selection and its potential was noticed. The approach also showed its efficiency when generalized, and a sensitivity analysis on its input parameters was performed to evaluate its sensitivity to initial inputs imposed by operators.

Finally, the reduced sequence was used in a heavy multi-objective optimization study by NSGA-II. An adaptive strategy for optimization employing reduced sequences named OptiTypSS was introduced comparing the obtained results to an adaptive metamodel based approach. The method succeeded in obtaining optimal results very close to the ones from a reference full year simulation requiring less heavy simulations (30 for the metamodel approach while 9 for OptiTypSS). On the other hand computational time taken by the proposed strategy was higher than the one of metamodel due to the time consumed in the day selection process which could be enhanced in future work.

Keywords: Buildings, energy systems, short sequence, computation time reduction, multi-objective optimization.

Résumé

Les approches holistiques en modélisation des bâtiments sont des démarches globales considérant les fortes interactions entre l'enveloppe, les systèmes, l'environnement et les usagers. Par contre, ils sont très pénalisants en temps de calcul du fait de l'utilisation de modèles détaillés en régime dynamique et de périodes simulées longues. Dans ce contexte, la réduction du temps de calcul est un véritable défi pour les études holistiques.

La démarche classique utilise les méta-modèles ou des modèles réduits. La thèse explore une autre voie basée sur la réduction de la période simulée au lieu du modèle lui-même. L'objectif est de définir une séquence de jours suffisamment courte pour déterminer avec le modèle dynamique complet les performances qui sont ensuite extrapolées à l'année complète. Cela permettrait ainsi de développer une approche méthodologique plus rapide et plus accessible pour la conception des bâtiments.

Après avoir présenté et analysé les méthodes utilisées dans la littérature, la thèse présente une nouvelle approche itérative intégrant un algorithme de regroupement. Le nouvel algorithme, appelé TypSS (Typique Short Sequence) Algorithme, crée et améliore de manière itérative une séquence courte de jours typiques basée sur des critères de sélection reflétant les performances annuelles d'un cas d'étude. L'algorithme a été appliqué sur un modèle de bâtiment détaillé et a conduit à des simulations beaucoup plus rapides tout en obtenant des résultats très proches des résultats annuels. Les résultats ont également été comparés à une approche itérative et de regroupement utilisées pour la sélection de jours et son potentiel a été remarqué. L'algorithme a également montré son efficacité lorsqu'elle est généralisée. Une analyse de sensibilité sur les paramètres d'entrée a été réalisée pour évaluer la sensibilité aux paramètres devant être fixés par un utilisateur.

Enfin, la séquence réduite a été utilisée dans une étude d'optimisation multicritères par NSGA-II. Une approche adaptative d'optimisation utilisant des séquences réduites nommée OptiTypSS est introduite en comparant les résultats obtenus à une approche adaptative basée sur le métamodèle. La méthode a permis d'obtenir des résultats très proches des individus optimaux obtenus à partir de simulations sur une année complète. D'autre part, le temps de calcul pris par la stratégie proposée était plus élevé

que celui du métamodèle en raison du temps consommé dans le processus de sélection du jour. En conséquence, elle pourrait être amélioré dans les travaux futurs.

Mots clés: Bâtiments, systèmes énergétiques, séquence courte, réduction du temps de calcul, optimisation multi-objectifs.

Acknowledgments

Three years of continuous work to be written in a single report was hard, but not as hard as writing this section. This path was full of great people that supported, encouraged and influenced me throughout this journey and I cannot mention them all in couple pages. To them all I dedicate this work.

I would like to acknowledge the financial support provided by the Region Auvergne-Rhône-Alpes, France for this PhD thesis through the project “Optimization of energy networks and energy producing buildings” OREBE. Thanks also to all partners contributing in the project for their interesting exchanges and valuable advices during the periodic meetings.

I would like to thank my directors and supervisors who were always around and assuring that I have all the necessary conditions to succeed. Gilles Fraisse for his excellent direction and all the time he dedicated for the success of this thesis throughout its course. Etienne Wurtz for the support, for believing in me, giving me the opportunity to participate in international projects and the efforts he did to assure the success of my international exchange program to Berkeley Lab. Antoine Leconte for his big heart and daily presence for all my questions making the long days of work go very smooth. And finally Simon Rouchier for his remarks regarding the coding and writing process.

I would like to thank every single colleague from CEA who became family after sharing a lot of good memories both in and outside the working hours. Special thanks go to Franck who was always a great support and helped me blend in since day one through his kind gestures and great humor. You were a huge part of this. Special thanks go also to Blaise and Clémence for always being kind and helping me fit better.

I would like to thank colleagues in LOCIE. From its director Monika Woloszyn to all permanent and non-permanent members and especially PhD students, we have shared a lot together. Special thanks go to Ainagul, Madina, Amin, Parul, Marie, Hugo and Gaetan, I am glad that we ended this journey as we started it, together. Thanks go also to Mathi, Manu and Taini for all the good memories we shared in and outside the lab.

I would also like to thank colleagues from Berkeley lab for their kind welcome in my short yet very rich stay with them. Thanks to Michael, Jianjun, Kun, Nari, Tea, Haris, Lisa, Hagar, Antoine, Jose and Javi for the great time we spent. Thanks also to my roommates Arjit, Surej and Mohit. My American experience was unforgettable thanks to you all.

Thanks to all my Lebanese friends in France who were a great part of this journey. From Paris to Chambery you always made me feel home. I wish I could mention you all but you are too special to miss.

Finally I would like to thank my family. To my mom and dad who are the reason I made it so far. You have given me the light that I am following to pursue my dreams and for that I am forever thankful. Thanks to Ali, Hussein, Lamia and Youssef, my siblings and backbones whom I am blessed to have.

To you all thank you. This work is for you.

Table of contents

List of figures

List of tables

Abbreviations

ANN	Artificial Neural Network
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
BES	Building Energy Simulation
BIM	Building Information Modeling
BOP	Building Optimization Problem
BPS	Building Performance Simulation
CEA	French Alternative Energies and Atomic Energy Commission
CHP	Combined Heat and Power
CVRMSE	Coefficient of Variation of Root Mean Square Error
CO_2	Carbon dioxide
DHW	Domestic Hot Water
GHG	Greenhouse Gas
HDD	Heating Degree Day
IIASA	International Institute for Applied Systems Analysis
IPCC	Intergovernmental Panel on Climate Change
LHS	Latin Hypercube Sampling
MILP	Mixed Integer Linear Programing
NMBE	Normalized Mean Bias Error
NSGA	Non Dominated Sorting Genetic Algorithm

NSGA-II	Improved Non Dominated Sorting Genetic Algorithm
OECD	Organization for Economic Co-operation and Development
OptiTypSS	Strategy for multi-objective optimization with TypSS
PSO	Particle Swarm Optimization
RMSE	Root-mean square error
SCSPT	Short Cycle System Performance Test
TypSS	Typical Short Sequence Algorithm

Nomenclature

Variables

Used in model

INS	Insulation thickness	*m*
VST	Volume of storage tank	m^3
SCOLL	Surface of solar collector	m^2

Used in TypSS

$crit$	target criterion	-
$crit_{year,indv_{n,j}}$	Criterion value obtained with the reference annual simulation for day j of the period n by the individual indv	*criterion unit*
$crit_{year,indv_n}$	Criterion value obtained with the reference annual simulation for period n by the individual indv	*criterion unit*
$crit_{year,indv_{max}}$	Maximum criterion value obtained with the reference annual simulation for all periods by the individual indv	*criterion unit*
$crit_{year,indv_{min}}$	Minimum criterion value obtained with the reference annual simulation for all periods by the individual indv	*criterion unit*
$\widetilde{crit}_{year,indv_n}$	Normalized criterion value obtained with the reference annual simulation for period n by the individual indv	-
$\overline{crit}_{year,indv}$	Mean criterion value obtained with the reference annual sequence by the individual indv	*criterion unit*
$crit_{red,indv_n}$	Criterion value obtained with the short sequence for period n by the individual indv	*criterion unit*
$\widetilde{crit}_{red,indv_n}$	Normalized criterion value obtained with the short sequence for period n by the individual indv	-
$crit_{year,indv_{sum}}$	Criterion annual sum value obtained with the reference annual sequence by the individual indv	*criterion unit*

$crit_{red,indv_{sum}}$	Criterion annual sum obtained with the short sequence by the individual indv	*criterion unit*
$\overline{crit}_{year_{period}}$	Mean criterion value obtained with the reference annual sequence by an individual	*criterion unit*
dat_{ref}	Reference data (year, 3 years, 10 years…)	-
dat_{year}	Yearly data	-
$dat_{red_{START}}$	Initial reduced sequence data	-
$dat_{red_{STOP}}$	Intermediate reduced sequence data	-
$dat_{red_{final}}$	Final reduced sequence data	-
d^n	Characteristic day of a period n	-
$day_{n_{start}}$	First day of a period n	-
$day_{n_{stop}}$	Last day of a period n	-
E_{crit}	Relative annual sum error of a criterion by an individual	%
E_{Global}	Relative annual sum error of all criteria by an individual	%
E_{max}	Maximum global annual sum error of all criteria between all the individuals	%
I_{indv}	Tested individuals	-
mod	Tested model	-
$n_{criteria}$	Number of target criteria	-
n_{indiv}	Number of individuals	-
n_{period}	Number of periods	-
n_{START}	Number of days in the initial sequence	*days*
n_{STOP}	Number of days in the generated sequence	*days*
n_{size}	Number of days in a period n	*days*

N_{days}	Number of days in the test sequence	*days*
par_{indv}	Parametric configuration of an individual	-
R^2_{crit}	Coefficient of determination of a target criterion by an individual	-
$R^2_{crit_{all}}$	Coefficient of determination of a target criterion by data points from all individuals	-
R^2_{Global}	Global coefficient of determination of all criteria by data points from all individuals	-
$Score_{crit_n}$	Period n score for a criterion	-
$Score_{p_n}$	Period n score for all criteria	-
$targ_{period}$	Target period	-
$(\Delta_{crit_n})_{indv}$	Difference between normalized reference and predicted criterion value for a period n and an individual indv	-
Δ_{p_n}	Difference between normalized reference and predicted criterion value for a period n and all individuals	-
$\Delta T^{day_{n_{start}}@day_{n_{last}}}$	Target period specifying its limits	
	Used in SCSPT	
G'coll	Target irradiation sum	*kWh*
G'hor	Target horizontal irradiation sum	*kWh*
T'amb	Target ambient temperature	*°C*
μ_{SCSPT}	Breaking threshold	*%*
$\Delta \mathrm{E}_{SCSPT}$	Global error	*%*
	Used in NSGA-II	
$c(i)$	Child individual at iteration i	-

$p_1(i)$	Parent 1 individual at iteration i	-
$p_1(i)$	Parent 2 individual at iteration i	-
α	Percentage of crossing	*%*
	Used in OptiTypSS	
$\mathrm{cost}_{\mathrm{Total}}$	Total investment cost	*€*
$\mathrm{cost}_{\mathrm{Coll}}$	Investment cost of solar collector	*€*
$\mathrm{cost}_{\mathrm{Vol}}$	Investment cost of storage tank	*€*
$\mathrm{cost}_{\mathrm{Ins}}$	Investment cost of insulation material	*€*
$\mathrm{cost}_{\mathrm{Ins}_{\mathrm{ext\ wall}}}$	Investment cost of external wall insulation	*€*
$\mathrm{cost}_{\mathrm{Ins}_{\mathrm{roof}}}$	Investment cost of roof insulation	*€*
$\mathrm{Q}_{\mathrm{backup}}$	Annual backup energy	*kWh*
$n_{indv_{ini}}$	Number of initial tested individuals	*indvs*
$n_{indv_{selec}}$	Number of individuals selected from the predicted Pareto	*indivs*
µ	Breaking threshold	*%*

General Introduction

The building sector in its two forms, residential and commercial, accounts for about one-third of the global energy demand. However, the sector offers significant potential for improved energy efficiency with high-performance envelops and energy-efficient systems. Building energy simulations (BES) and optimization are increasingly demanded in the field because of its emphasis on sustainability. Yet, performing global approache studies on buildings, which takes into consideration both the envelope and the systems, leads to the complexity of models under study, leading therefore to unfeasible computational time expenses. Usually, simplified or surrogate models instead of detailed building models are used to avoid this issue.

However, this model replacement may affect severely the representation of the tested case study and therefore rises doubts concerning the credibility of applied studies. In addition to that, surrogate models may be inapplicable in case of large complex models due to the need of numerous learning data to construct. A holistic approach that might solve those doubts is a current case of interest. It is based on the reduction of input data profiles rather than the model itself.

The thesis presents and evaluates a developed day selection approach called TypSS (Typical Short Sequence) Algorithm to generate reduced sequences that can be applied on detailed models, despite their level of complexity, in dynamic simulations. A multi-objective optimization approach, named OptiTypSS, is then presented and evaluated in this research work. It employs reduced sequences generated by TypSS to accelerate heavy multi-objective optimization studies.

The manuscript is divided into four chapters starting from a thorough literature review. In this first chapter, the problem of energy sources depletion and global warming is discussed showing the role of the building sector in this worldwide crisis. A literature review is then conducted around the studies and approaches applied in the domain as attempts to reduce the energy impact of the sector. The main issues facing researchers and engineers are discussed which are directing toward global approaches for building performance evaluation such as the use of reduced sequences in building simulations, the subject of this thesis. Approaches for typical day’s selection used in the literature

are described afterwards showing the interest in developing a new generalized approach.

Chapter two presents this generalized approach, named TypSS, explaining the process it takes in each step starting from reference data to a final reduced sequence of typical days. The process is explained on a building connected to a combined solar thermal and heat pump system to simplify its presentation.

The third chapter is divided in two parts. The first part presents the obtained results upon a dynamic simulation of the case study and compares them to those obtained by two other methods of sequence reduction used in the literature. Using the sequence obtained by TypSS was 25 times faster than the annual one and best performant with respect to the others. Moreover, the chapter evaluates the generalization potentials of the algorithm through simulating the case study but this time with several parametric configurations simultaneously and not considered in the day selection process. The sequence estimated the performances with relative errors inferior to 10% and Coefficient of Variation of Root Mean Square Error (CVRMSE) inferior to 25% the limit specified by ASHREA. The second part presents a sensitivity analysis on the input parameters of the algorithm and implements recommendations for a better performing reduced sequence.

Finally, chapter four presents the results obtained upon using the obtained reduced sequence in a sequential and an adaptive multi-objective optimization study applying the conclusions acquired in the previous chapter. The adaptive OptiTypSS approach is introduced showing the accuracy of its obtained Pareto front by comparing it to the results obtained by a surrogate model of the tested case study. Obtained Pareto fronts by the two approaches were very close to the reference one but the global computational time was much higher with the new proposed strategy. Therefore, improvements are required and several measures are proposed in the perspectives that open the door to new more profound work.

Chapter I Concept of building performance evaluation and model study by reduced sequences

I.1. Introduction

During the industrial revolution, humanity witnessed a transition phase from an agricultural dominated society to a commercial industrial one [1]. World population has increased from around 700 million to more than 7 billion people nowadays. A growth pattern expected to exceed 9.7 billion by the year 2050 according to the International Energy Agency [2]. This is accompanied by a 70% expected increase in worldwide household (home unit) with respect to the year 2010 [3]. Due to this trend, man's daily habits and living conditions have transformed radically, leading to a change in daily life style and the urge for resources to power the new growing communities.

Steam engines were soon powering transportation, factories, homes and farm implements. Coal was also used for heating buildings. At the end of the 19th century, oil, processed into gasoline, began trending as the main energy resource for internal combustion engines. Energy use was increasing rapidly, doubling every year while the cost of energy production was declining steadily. However, this was accompanied by the depletion of those abundant yet limited resources and a drastic increase in air, water and soil pollution.

Temperature measurements made in different places of the globe during the 20th century show an increase in average temperature compared to the previous century. This increase has taken place in two stages, the first from 1910 till 1945, the second from 1976 till today [4]. Moreover, the work of several researchers of the Intergovernmental Panel on Climate Change (IPCC) shows the existence of a correlation between the CO_2 concentration and the temperature at the surface of the earth [5]. Following these observations, the massive exploitation of fossil fuels has been singled out as certainly responsible. The assertions of the IPCC, expressed in the various reports it has produced [6], have ruled on the responsibility of fossil fuels in the increase of gases to greenhouse effect (GHG) in recent decades. Currently the majority of decision makers recognize that global warming is anthropogenic in origin.

The overall energy and environmental situation is even more complex. The world population continues to increase almost linearly during the three past decades. Population growth naturally generates more activities and creates more needs. Between 1980 and 2015, carbon dioxide emission increased by around 60%. Recent forecasts show that demand for energy will continue to increase reaching up to 39 gigatonne of CO_2 emissions by the year 2030 (Figure I-1). According to [7] the energy

consumption in developing countries will increase with an annual average of 3.2% exceeding that of developed countries.

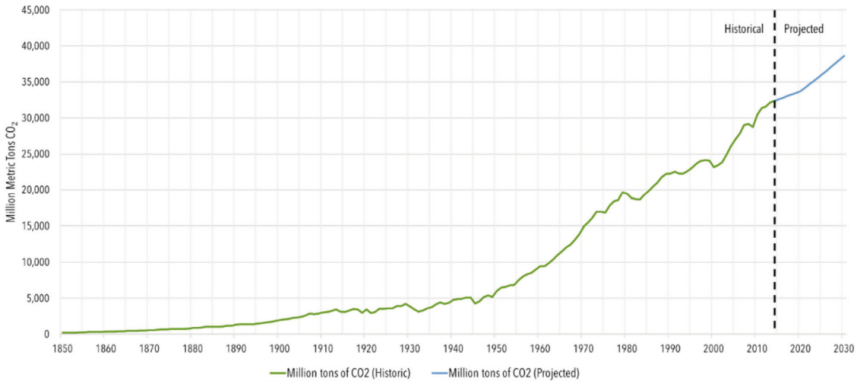


Figure I-1 . Evolution in global carbon dioxide emissions from 1850 to 2030. Source: IEA. [8]

Examining the origin of this massive need, the final energy consumption is often attributed to four main economic sectors: industry, transport, residential and commercial. Figure I-2 reflects the share of energy consumption divided between the four sectors for both Organization for Economic Co-operation and Development (OECD) and non-OECD countries. Curves show that although the industrial sector dominates the global energy consumption, the building sector in its two forms, residential and commercial, accounts for about one-third of the global demand. A demand projected to increase progressively as the global demand increases. The constant search for comfort, growth of the population as well as the increase in time spent in buildings are the main reasons behind this growing need [9].

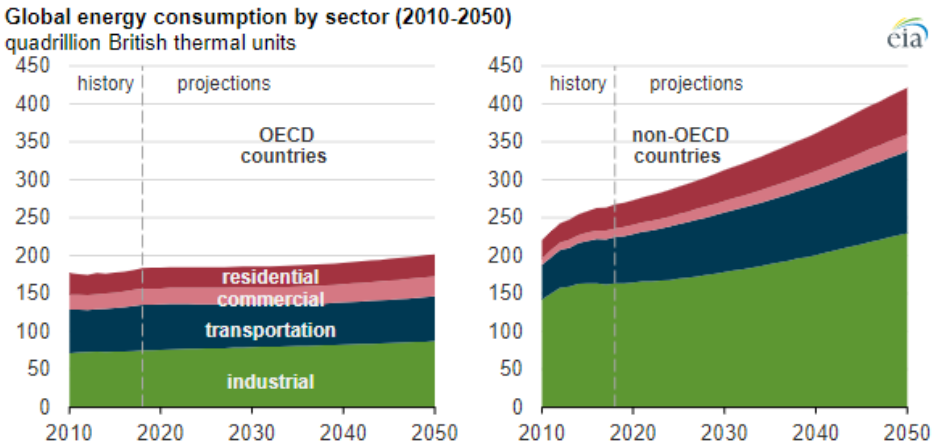


Figure I-2. Global final energy consumption by sector, history and projection. Source: IEA

I.2. Buildings sector on worldwide scale

Requirements for heating, production of domestic hot water (DHW) and for air conditioning (in hot climates) are the fields responsible for much of energy consumption in buildings (all types of buildings). As shown in Figure I-3, the use of energy allocated to these services, in the developed countries, represents approximately between 60% and 85% of energy consumed in the building.

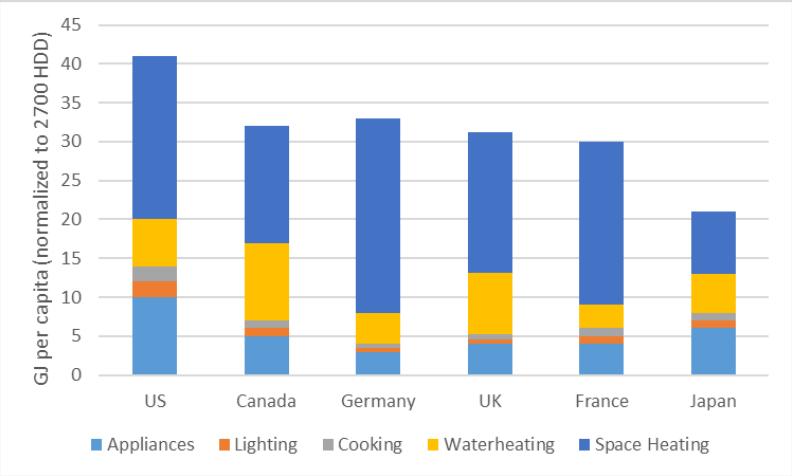


Figure I-3. Final energy use per service in developed countries in 2007. Source: report of IIASA [10] (HDD = heating degree day).

Much of this energy is from a fossil source (Figure I-4). The housing and tertiary-type buildings are therefore responsible for around 33% of emissions of CO_2, 66% of chlorofluorocarbons and between 25% and 33% of carbon black [10]. In hot climatic zones the air conditioning needs take the place of heating needs. With a population extremely large and energy resources solely based on energies emitting greenhouse gases, the energy and environmental situation in these countries could be more worrying in the future.

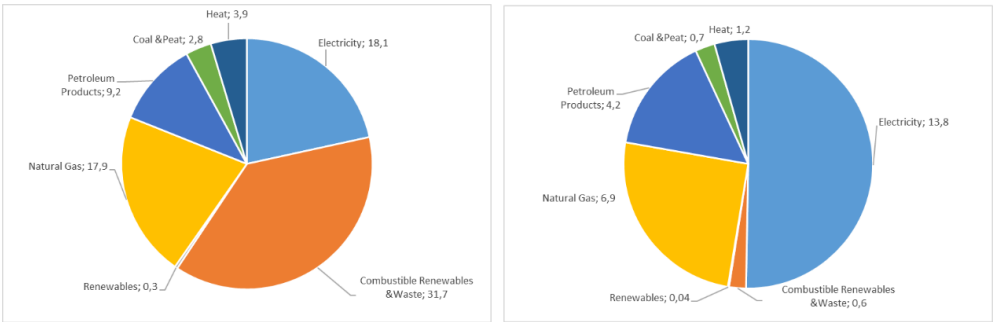


Figure I-4. World final energy consumption by source in residential sectors (left), commercial and public (right) in 2007 [10]

I.3. Buildings sector on French scale

Regarding France, it is true that the unit needs of buildings is decreasing over time (thanks to thermal insulation, considering energy aspects during construction, housing rehabilitation etc.). However, energy demand in this sector remains very high especially due to the large stock of strongly consuming existing buildings. Overall, the consumption in the building sector (residential and tertiary) has been practically stable since 2003. It represents 44% of the total consumption, far ahead of transport (32%), industry (21%) and other sectors mainly agriculture (3%) (Figure I-5).

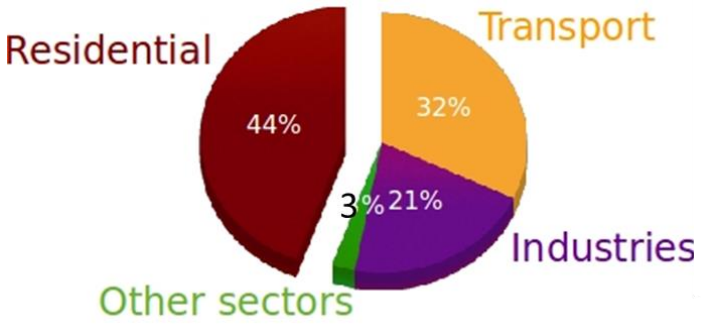


Figure I-5. Final energy use for each sector adapted from French Environmental Energy Agency ADEME (Source: D. Mauree, 2014)

Buildings sector consumption breaks down into two thirds for residential buildings (main and secondary residences) and a third for the commertial sector [11]. In 2013, space heating and cooling accounted for 68% of this share thus being the major contributor in the energy consumption in buildings. The remaining share was divided between domestic hot water needs, cooking and specific electricity such as lighting and appliance functioning requirements as shown in Figure I-6.

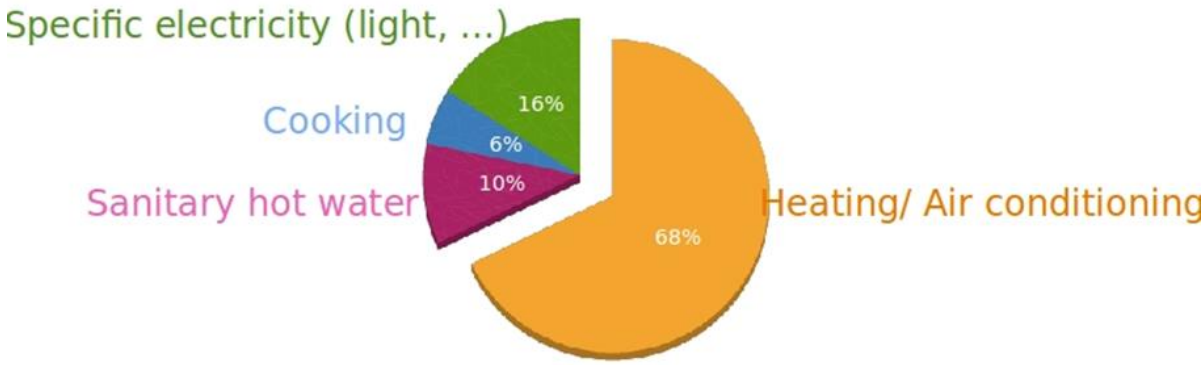


Figure I-6. Final energy use inside buildings adapted from French Environmental Energy Agency ADEME (Image credit: D. Mauree, 2014)

I.4. Concept of building performance simulation (BPS) and optimization

From this point, governments all over the world started adopting new regulations and laws that take into consideration environmental impacts for new projects. In addition, research to improve the different sector performances has become more supported by governments through more funding and new policies.

For instance, the European Union had put a policy that requires to commit a 9% reduction in energy use by 2016 based on the 2006/32/EC directives [12], in addition to decreasing the greenhouse gas emissions as well as primary energy consumption by 20% as indicated by the climate change package legislation [13]. Paris climate change accord, COP21 [14], mandates involved countries to limit the total CO_2 emission to 40 billion tons emitted per year in order to limit the global warming to 1.5°C.

The building sector offers significant potential for improved energy efficiency with high-performance envelops and energy-efficient systems. From this point, the interest in building design studies has risen.

Building performance simulation (BPS), also denoted Building Energy Simulation (BES), is increasingly used to design buildings because of its emphasis on sustainability [15]. The requirement of building design are comprised of qualitative elements (social impact, esthetics, special planning, etc.) and quantitative elements (cost, yearly-consumed energy, amount of daylight, etc.). The design aims on satisfying multiple criteria in addition to measurable performances. Several papers examine how a geometric model can dynamically be operated in relation to BPS [16]–[18].

The potential of using different BPS tools can be categorized in two possible stages, simplified and detailed design stages [19]. Simplified tools have shown to be useful at certain point of early design stage but might be limited to apply on later design evaluation. Many researchers have published various methods containing high precision calculations, focusing on manual variations ([20], [21]) while others used Monte Carlo algorithms ([19], [22]). When it comes to optimization, most studies focus on optimizing singular or very few objectives such as the electrical consumption ([16], [23], [24]). In general, such methods seek high precision of performance functions, which in turn penalizes the speed of calculation time.

Optimization algorithms run numerical models iteratively, constructing sequences of progressively better solutions up to a point that satisfies pre-defined optimal conditions. This point is not necessarily the globally optimal solution since it might be unfeasible due to the nature of the case study [25] or even the program itself [26]. Because of code features, the search space may be non-linear and have discontinuities, requiring the use of special optimization methods that do not require the computation of the derivatives of the function [27]. In building optimization studies, the building simulation model is usually coupled with an optimization engine, which runs algorithms, and strategies to find what is described to be an optimal solution [28]. Two of the optimization examples used in building optimization and usually applied on simplified models are described hereunder:

- *Pattern search,* which is an iterative search for the optimum that does not require a gradient and therefore can be used in non-differentiable or continuous functions. The step size is halved in case of no more improvement is possible [29].
- *Linear programing* that simplifies the problem into a linear problem (matrix) to compute directly the optimum. The optimum falls in an external point if all objective functions and constraints are linear [30].

In their review on simulation based optimization studies for building performance analysis, Nguyen et al. [31] divided the process in three phases:

- *Preprocessing phase* where the formulation of the optimization problem takes place including the building model, the objective functions and constraints, selecting the appropriate optimization algorithm and coupling it with the model. It is important in this phase for the model to be simplified to avoid severely delaying the optimization process, but not too simplified to avoid inaccurate modeling of building phenomena [32].
- *Optimization phase* where monitoring, controlling and detecting errors of the study takes place. It is worth mentioning that in this phase, it is almost impossible to estimate the time of convergence of the optimization algorithm. Researchers do not usually mention the time taken by the algorithm to converge to an optimal solution since the behavior of the optimization algorithms is not trivial. However, several attempts have been applied to speed up the time of simulation while still reaching good final results such as in [33].

- *Post processing phase* where interpretation, verification, presenting of results and decision making take place.

In addition to simplified models based optimization methods, evolutionary algorithms are very common in the building optimization field. They are usually applied in dynamic detailed model optimization due to their learning process that helps in converging faster to optimal solutions based on results from previous iterations. Such algorithms apply the Darwinian principle of survival of the best by keeping a population of solutions of which the poorest are eliminated. Types of such algorithms include Genetic Algorithms (GA) [34], Evolutionary Programing (EP) [35], [36], Covariance Matrix Adaptation, Evolutionary Strategy (CMA-ES) [37] and Differential Evolution (DE) [38]. Other algorithms that mimic natural processes include Harmony Search (HS) [39], Particle Swarm Optimization (PSO) [40], Ant Colony Optimization (ACO) [41] and Simulated Annealing (SA) [42].

As mentioned previously, many building optimization studies use the single objective approach where one objective function can be optimized in an optimization run [43]. However in real world, designers have to deal with several contradictory design criteria simultaneously such as minimizing energy demand while minimizing cost or maximizing internal comfort [44], [45]. Therefore, multi-objective optimization is more relevant than the single objective approach and there exist numerous research papers that consider this approach for optimization as will be shown in the following paragraphs.

In their review done on the optimization methods applied to renewable and sustainable energy, Banos et al. [25] have shed the light on the concept of single and multi-objective optimization. They introduced the fact that in many applications, multi objective optimization is inevitable because of the interaction of several decision parameters.

There exist two main approaches to solve multi-objective problems:

- *Scalarization approach* that assigns different weight factors to each criterion and therefore back to a single objective problem, the weighted sum of the criteria [46].
- *Pareto optimality approaches* where a trade-off optimal solution is examined and appropriate solutions are then determined. The approach is referred to as

"Pareto optimization". The basic principle, established by Pareto in 1896, is as follows: "In a multi-objective problem, there is such a balance that one cannot improve one criterion without deteriorating at least one of the other criteria ". This equilibrium is called the Pareto optimum. A solution is said to be Pareto optimal if it is not dominated by any other solution where there is no other solution that can better improve one criterion without deteriorating another. The Pareto front is the set of optimal Pareto solutions. Due to the complexity of BOPs, researchers often use up to two objective functions with very few studying three or more functions such as in [47] who optimized energy consumption, CO_2 emission and initial investment cost or in [48] who optimized energy consumption, thermal comfort and initial investment cost. The process of selecting the optimal solution from the front is not trivial and is known as multi-criteria decision-making. Many decision making techniques have been developed [49] such as "pros and cons", "simple prioritization" and "bureaucratic".

Stadler et al. [50] created a multi-objective process to minimize CO_2 emissions by optimizing the energy systems linked to the building. Similarly, Merkel et al. [51], Milan et al. [52], Lauinger et al. [53] and others have studied building and energy supply system optimization by multi-objective approaches.

In multi-objective problems, splitting building design problems into sub-problems (envelope, systems, renewables…) may lead to missing out on synergies between different areas. As a result, many researchers through optimizing variables from different areas considered the building globally such as in [54]. Yet, performing holistic approaches on buildings, which takes into consideration both the envelope and the systems, leads to the complexity of models under study, especially when analyzing heat networks in the case of multiple buildings i.e. districts or blocks, leading therefore to unfeasible computational time expenses. Simulation of detailed building models may take several minutes in building energy simulation [32]. On the other hand, simulation-based optimization techniques require up to thousands of simulations to evaluate the case study. The optimization schemes may therefore become infeasible due to such computationally expensive models. Usually, very simplified models instead of detailed building models are used to avoid this issue, as in [54]–[56]. Particularly, in [57] Lee used a two-step optimization scheme to deal with an expensive CFD model. In the first

step, Lee performed the optimization on the simple CFD model. Then he performed a few detailed CFD simulations on the optimal candidate solutions found in step 1 to refine the results. Other methods employ reducing the population size and/or the number of generations. Mancarella et al. [58] used spatial aggregation to reduce the number of nodes in an energy system network study and Milan et al. [59] reduced nonlinearities and discontinuities to avoid non-convexity of the program. Other work using simplified analytical models can also be found in [60]–[64].

However, these reductions significantly lower the performance of optimization algorithms, and may result in sub-optimal solutions [65]. Surrogate models are among promising solutions to this problem. A surrogate model (meta-model or emulator) is an approximation model of the original. It typically mimics the behavior of the original model to be able to produce the model responses at reduced computational cost.

In the context of optimization, surrogate models can speed convergence by reducing function evaluation cost and/or smoothing noisy response functions [66]. After running the surrogate-based optimization, other refined optimization around the optimal points using the original model can be performed to obtain exact solutions. Klemm et al. [67] employed surrogate based optimization in their study by applying a polynomial regression method on CFD simulation results to derive explicit analytic objective functions, then optimizing them using a simple deterministic optimization method. Magnier and Haghighat [32] used TRNSYS simulations to train an artificial neural network (ANN), then used the trained – validated ANN to couple with the genetic algorithm (GA) to optimize thermal comfort and energy consumption. The database for training the ANN consists of output of 450 simulations. Time for generating the database was 3 weeks, but optimization time was very small. If direct coupling between TRNSYS and GA was used, it would need 10 year to finish the task [32]. Chen et al. [68] used a feed forwards neural network for the identification of temperature in intelligent buildings and then optimize by the particle swarm optimization (PSO). Eisenhower et al. [69] used the Support Vector Machines method to generate several meta-models of a 30-zone EnergyPlus building model and then performed sensitivity analysis to select the most influential variables for optimization. These authors stated that the optimization using the meta-model offers nearly equivalent results to those obtained by EnergyPlus model.

They also recommended that the use of Gaussian Process regression, sometimes-denoted Kriging models, for optimization of complex buildings require further investigations. Gengembre et al. [70] minimize 20-year life cycle cost of a single-zone building model using a surrogate model and the PSO. They concluded that the accuracy of their surrogate model is acceptable and such a surrogate model can further help designers in design space exploration with cheap simulation cost.

However, the accuracy and sensitivity of surrogate based optimization is currently not a well-developed area, especially when the number of input variables is large [71], the cost function is highly discontinuous or in cases many discrete input variables exist.

The strength and weakness of various surrogate methods is a great research field of computational and statistical science and well beyond the scope of the building simulation community. There is currently no consensus on how to obtain the most reliable estimate of accuracy of a surrogate model, thus the coefficient of correlation R^2 is often applied, as in [32], [72]. R^2 is the proportion of the variance of a dependent variable that is predictable from independent variable(s). Furthermore, the random sampling method of inputs, the number of building model evaluations used to construct and validate a surrogate model is still problematic and is often chosen empirically by analysts. It also needs more studies to see whether significant difference between optimization results given by a surrogate model and an ‘actual’ building model exists. In addition to that, the processing time of optimization studies can be severely affected by the balance between the number of variables and their options. Usually, computer clusters are used for complicated optimization problems with large number of variables [73]. These questions are explicit challenges of the building research community.

On the other hand, the use of detailed models is very useful for accurate and credible studies. A holistic approach that might solve those doubts, and working on detailed models, is a current case of interest. It is based on the reduction of input data profiles rather than the model itself. The approach evaluates annual performances of a model starting from a short simulation sequence of typical selected days instead of complete 365 days input data profiles. Therefore, instead of simplifying the models, running short sequences is used to reduce the computational time expenses of a fully dynamic simulation. Figure I-7 illustrates the different approaches of simulation adopted in the BPS domain and their relation to the complexity of the model.

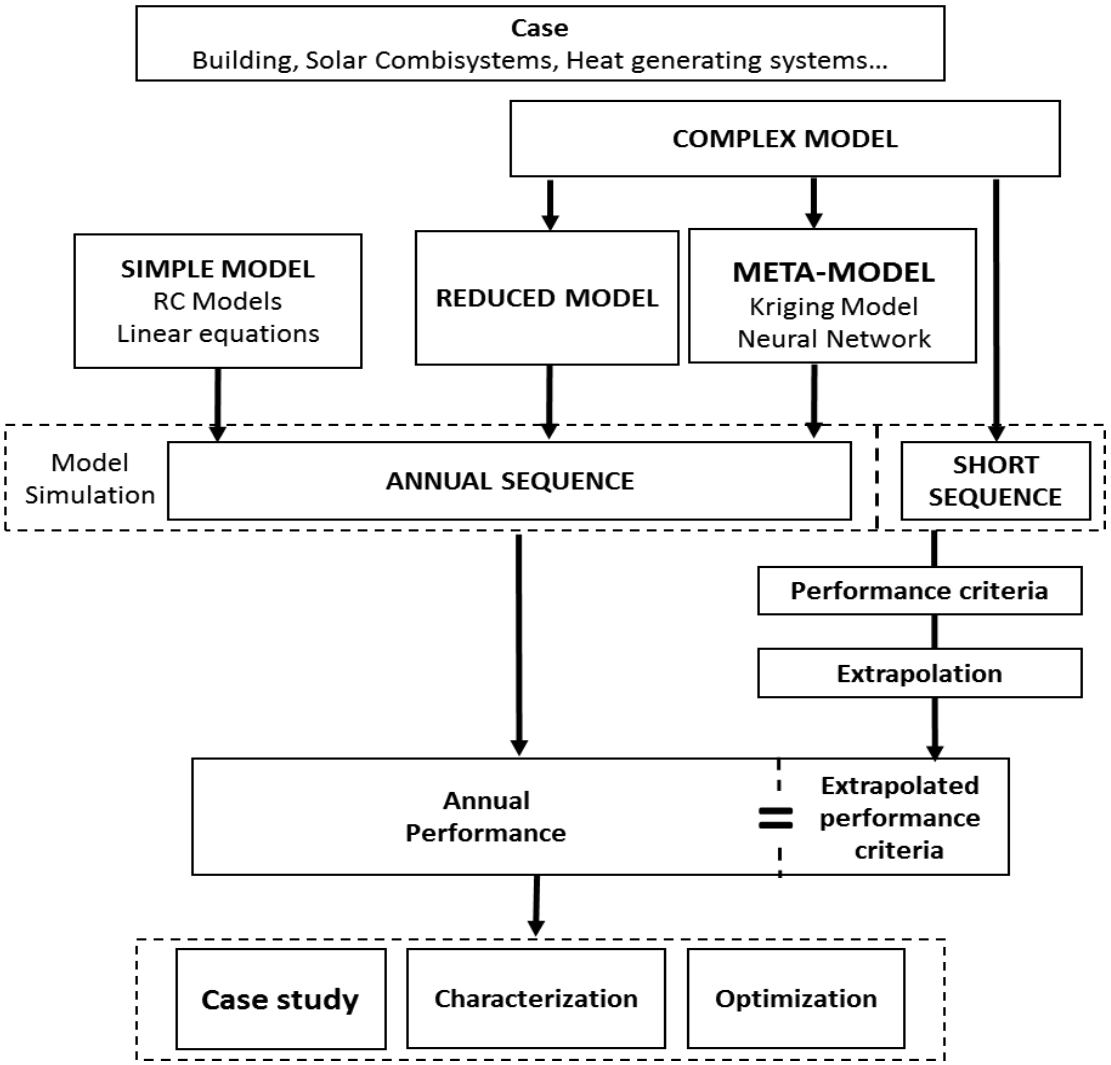


Figure I-7. The relation between the complexity of the case study and the type of time sequence used for simulations.

Table I-1 shows the main advantages and disadvantages of these approaches. As mentioned above, reduced order models such as modal analysis, RC models and metamodels are derived from the case study numerical models and are simulated on complete annual profiles. On the other hand, the reduced simulation sequences are derived from input data profiles and introduced directly to complex models for dynamic simulations.

Approach	Case study state	Annual profiles state	Advantages	Disadvantages
Reduced order models	Reduced	Complete (365 days data)	Software widely available Easy to implement	Requires specialist experience Time consuming to develop
Simplifying model through analogical RC models	Reduced	Complete (365 days data)	Easy to implement on envelop studies Rapid thermal dynamics become negligible	Not useful in non-linear problems Requires specialist experience Inaccurate when considering short dynamics
Metamodel	Reduced	Complete (365 days data)	Useful for deterministic applications Flexible Requires fewer parameters to fit than other methods	Complex method Requires specialist experience Time consuming to develop for each case study
Reduced simulation profiles	Complex	Reduced (Typical selected days)	Applicable despite model changes Applicable in nonlinear problems Flexible to apply	Requires previously calculated inputs to generate the reduced sequence Uses profile data for day selection, requires dynamic simulation

Table I-1. The approaches used to reduce computational time expenses and their pros and cons.

I.5. **Model study by short sequence**

The literature contains various approaches to select a representative set of historical periods. As shown in Figure I-8, the process starts by the original annual data and ends in a short sequence that will be later used in model testing or optimization. In between, the reduction approach implements day selection algorithms or works through continues testing to generate a sequence to reproduce the annual performance criteria after extrapolating the results found by the reduced simulation. These approaches can be grouped in three main categories: Heuristic Approaches, Iterative Approaches and Grouping Algorithms.

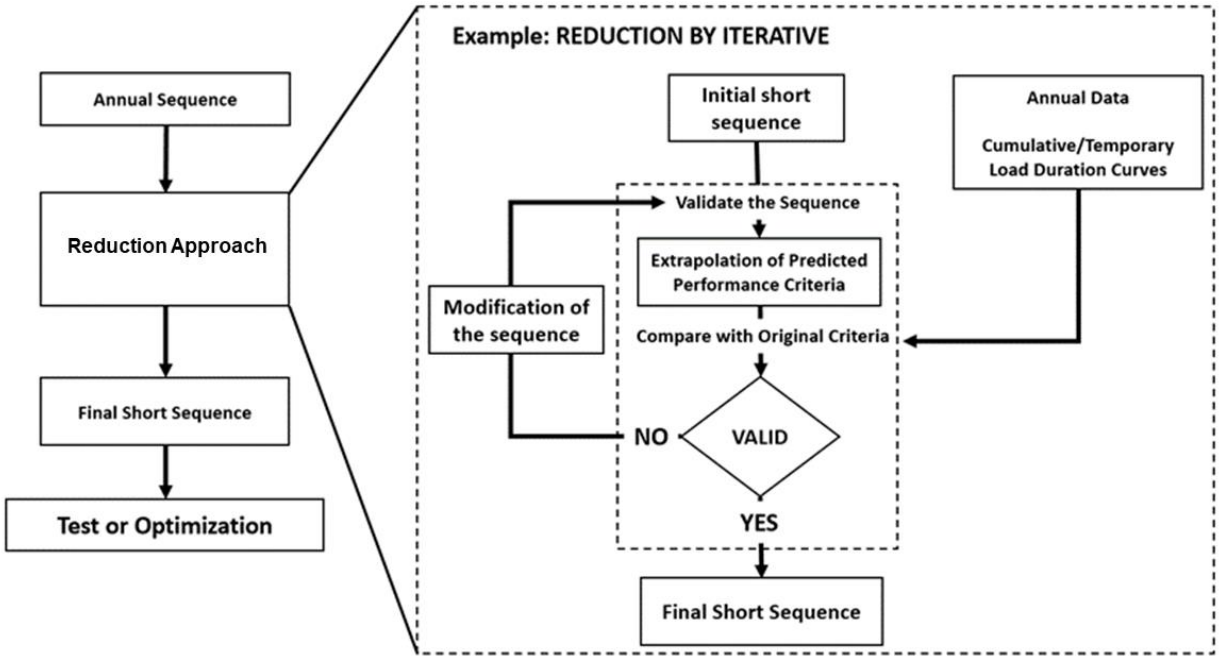


Figure I-8. General schema of the short sequence selection process followed by the iterative reduction approach.

I.5.1. Heuristic Approaches

Heuristic approaches are practical methods that select directly a set of typical days highly influenced by the personal expertise or experience of the developer. The selection is quick but not guaranteed to be optimal, Figure I-9. In their study, Belderbos et al. [74] selected the day that contains the minimum demand level of the year, the day that contains the maximum demand level and the day that contains the largest demand spread in 24 hours. Haller et al. [75] defined short-term fluctuation patterns represented by 13 days from the four seasons, each with three characteristic days that cover low, medium and high renewable energy supply regimes. They added an additional peak time day representing high demand and low renewable energy supply. Fripp et al. [76] discussed within investment periods optimized based on 12 days of sampled data: two for each even-numbered month. One day in each month corresponds to conditions that occurred on the peak-load day of the same. The second day of data for each month corresponds to a randomly selected day from the same month. Hart et al. [77] reduced the data size of energy generation by variable renewables by selecting eight specific days that contain hours with extreme meteorological and load events and 20 random days to characterize typical system

behavior. Weights for each day were assigned using least squares to best match the annual load, wind speed, and irradiance distributions.

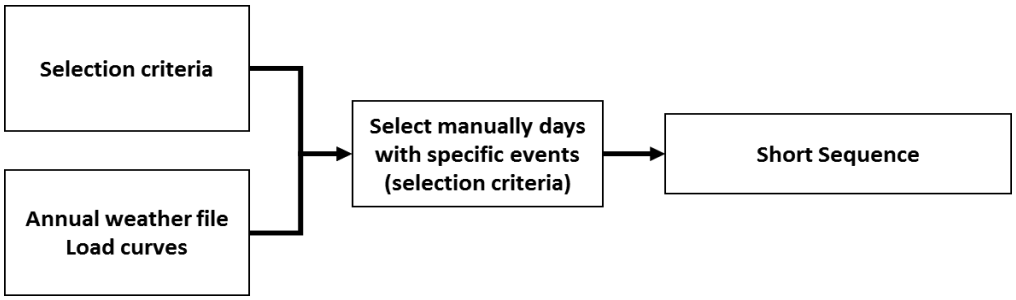


Figure I-9. Heuristic method in typical day selection.

I.5.2. Iterative Approaches

Iterative approaches search for the best solution after repeating the same action several times and comparing the quality of results in each iteration Figure I-10. There are many examples in the literature that use this approach for day selection, either directly by implementing iterations or indirectly through performing graphical methods or performing Mixed Integer Linear Programing (MILP) based on repetitive iterations. Ortiga et al. [78] who reproduced two cumulative energy demand curves, one for heating and the other for cooling, used a graphical method of iteration while studying the optimization of cogeneration and tri-generation models for building.

The French Alternative Energies and Atomic Energy Commission (CEA) has developed an iterative approach that reduces a whole year into twelve days and was used for testing solar combisystems [79]. The test was called Short Cycle System Performance Test (SCSPT), which selects the short sequence based on weather data, the energy demand, comfort and energy stored by the system. Results were very promising and the sequence was able to reproduce the annual performance with a good degree of accuracy and worked for different models.

Poncelet et al. [80] developed a MILP iterative approach to predict the electricity demand, the onshore wind generation and the PV solar generation data supplied by the Belgian transmission system operator. The basic model divides each cumulative load duration curve into a number of bins. Each bin corresponds to values within a specific range. MILP is then employed in an iterative way to identify a representative day of each bin as well as the weight assigned to each day based on the weight of the

bin thus minimizing the difference between the original and predicted curve until finally obtaining a duration curve as close as that of the original.

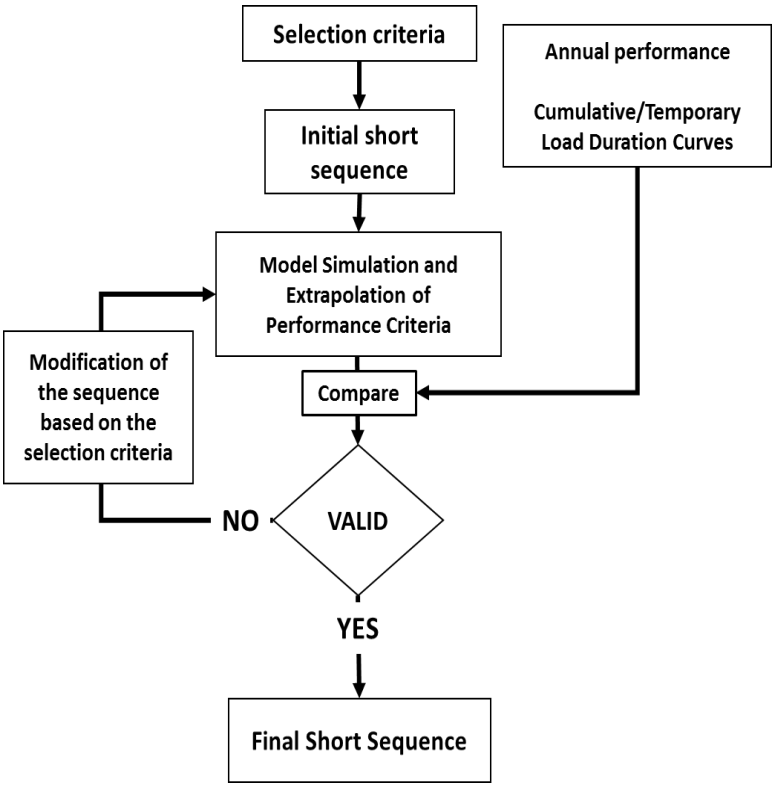


Figure I-10. Iterative method in typical day selection.

I.5.3. Grouping Algorithms

Grouping algorithms are more advanced approaches to select a representative set of historical periods. Days with similar attributes are grouped into clusters followed by day selection of each group, Figure I-11.

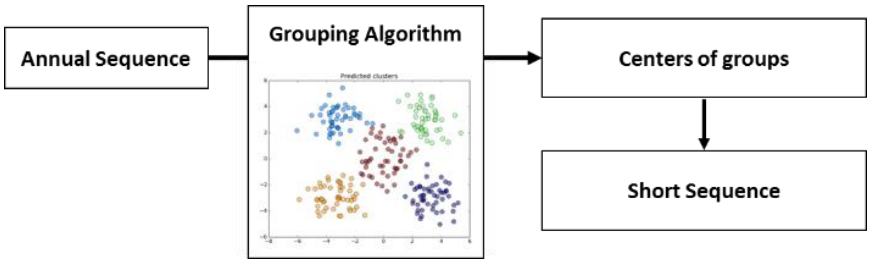


Figure I-11. Grouping method in typical day selection.

While clustering algorithms were the most preferred in studies for their simplicity and precision, some studies employed discriminant analysis to achieve grouping. Clustering algorithms are classified into exclusive and non-exclusive algorithms as

shown inFigure I-12. Exclusive clustering algorithms are those in which each data segment belongs to only one cluster, whereas for non-exclusive clustering (also known as fuzzy c-means clustering) each data segment may belong to more than one cluster with different degrees of membership. Exclusive clustering can be further classified into hierarchical and partitional clustering. Partitional clustering directly divides data segments into a pre-determined number of clusters without building a hierarchical structure, whereas hierarchical clustering seeks to build a hierarchy of clusters with a sequence of nested partitions, either from singleton clusters to a cluster including all data segments or vice versa, Figure I-13. The former is known as agglomerative hierarchical clustering, and the latter is called divisive hierarchical clustering.

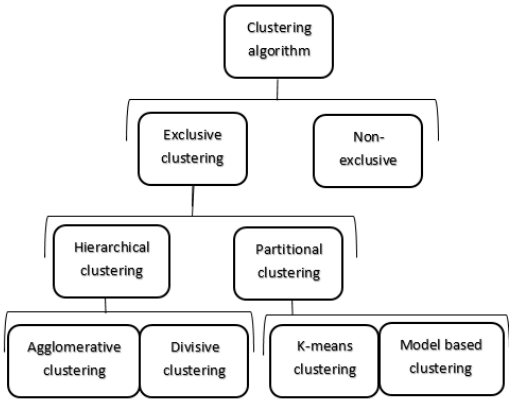


Figure I-12. Classification of clustering techniques [81]

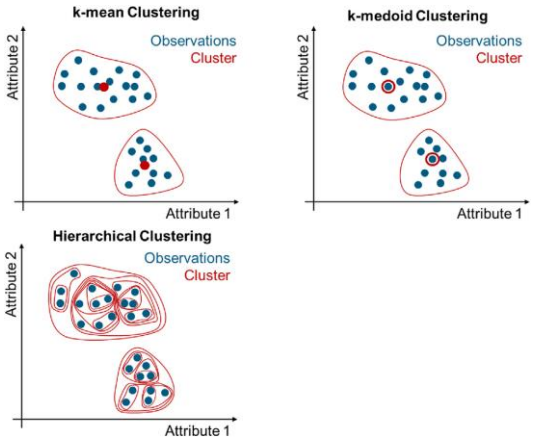


Figure I-13. Principle of partitional and hierarchical clustering [82]

Divisive is very computationally intensive. Therefore, agglomerative methods are usually preferred. Partitional clustering can be classified into k-means clustering algorithm and the model-based clustering (also known as probabilistic clustering or a mixture of Gaussians clustering). In model-based clustering, each cluster can be mathematically represented by a parametric distribution, like Gaussian (continuous) or Poisson (discrete) distribution. A mixture of these distributions therefore models the entire data segments. The probabilistic clustering algorithm seeks to optimize the parameters of the mixture model to “cover” the data segments as much as possible, which is considered a very computationally intensive process. Most of the studies relied on clustering approach use k-mean clustering as their favored approach. Fazlollahi et al. [83] used k-mean clustering to perform a multi-objective optimization

of district energy systems. Using performance indicators and statistical measures to assess reproduced profile error deviations, they succeeded in reaching a sequence of 8 days capable of reproducing several attributes duration curves with a good degree of accuracy and 40 times faster operation. Dominguez et al. [84] used k-means approach while optimizing a CHP system. Following a concept close to that of Fazlollahi, they succeeded in creating a calendar showing the day distribution between clusters, which showed the seasonal effects, and the work periods that influence the building performance. Kotzur et al. [82] tested the efficiency of hierarchical clustering in estimating and optimizing the performance of a residential energy supply system and compared it to other partitional clustering approaches. Menegon et al. [85] developed a new dynamic test procedure for laboratory characterization of energy systems using k-means clustering algorithm. Some other studies used clustering algorithm as a part of a bigger approach for time reduction. In his thesis study on multi-criteria optimization method for urban densification project, Ribault [86] divided the year into 14 clusters by k-means after which he started reducing each cluster by a deleting algorithm that deletes repetitive days by an aggregate weight function based approach.

On the other hand, discriminant analysis was used by Blachandra et al. [87] to reclassify days of 12 months of the year comparing the monthly average load curves of electricity demand to the daily ones. Discriminant analysis works in a way where first days of the month found to be misclassified were grouped with the previous group (month) while those of the last days were grouped to the next and the middle stray days were ignored. This regrouping approach ended with nine representative curves of nine groups with a percentage error of estimation in the range of 5% for 60% of the hourly loads and 10% for about 86% of the hourly loads.

Table I-2 presents a summary of the studies found in the literature and the approaches used in each.

The table also synthesizes the approaches presenting the case study or the model that was either tested or optimized, it was found that in the majority, the methods were used as a way to ease and fasten the testing simulations while a little actually used the obtained sequence to perform direct optimization studies. In addition to that, in most studies, the used models were reduced ones and therefore not working on detailed models. The profiles used for the study were either annual profiles from weather data

or load duration curves of the performance criteria, noting that the duration curves used were either temporal load curves or cumulative while mostly temporal. Evaluated criteria or attributes are also shown in the table. They depend directly on the case study, energy demand (heating and cooling) was tested in energy systems while power generation in CHP models and so on. Methods of extrapolation of the obtained results were evaluated and found to be related to the category of the method of reduction.

Table I-2. The different approaches found in the literature and their field of application.

References	Approaches	Case study	Type of the model	Used profile	Evaluated criteria	Extrapolation of sequence	Objective of the reduction
Balachandra et al.	Discriminant Analysis	Electrical plants	Reduced Model	Load profile (temporal)	Generation of electricity	-	Test method for system modeling
Ribault et al.	Clustering+ iterative	Urban densification	Complex model	Climate data (temporal), energy demand (temporal)	Energetic need	Weight of cluster	Optimization
Menegon et al.	Clustering (k-means)	Energy systems	Complex model	Climate data (temporal), load profile (temporal)	Climate data, thermal load	Weight of cluster	Test method for laboratory system characterization
Domeniguez et al.	Clustering (k-means)	CHP	Reduced Model	Load profile (cumulative)	Power	Weight of cluster	Test method for system characterization
Fazlollahi et al.	Clustering (k-means)	Energetic system of a district	Reduced Model	Climate data (temporal), load profile (cumulative)	Climate data, electric load, thermal load	Weight of cluster	Optimization
Kotzur et al.	Clustering (hierarchic, k-means, k-mediods)	CHP, energy supply systems	Reduced Model	Climate data (temporal), Electricity demand (temporal)	Climate data, electric load	Weight of cluster	Test method for system design
Poncelet et al.	Iterative (MILP), clustering (hierarchic)	PV panels, wind turbines	Reduced Model	Load profile (Cumulative)	Generation of electricity	Weight of cluster	Test method for system characterization
Albaric et al.	Iterative (SCSPT)	Solar combisystems	Complex model	Climate data (temporal)	Climate data, heating, thermal storage	Proportion	Test method for laboratory system characterization
Ortiga et al.	Iterative (graphical)	Regeneration systems	Complex model	Load profile (Cumulative)	Cooling, heating	Repetition factor for each time period	Optimization
Hart et al.	Heuristic	System of variable renewables	Reduced Model	Climate data (temporal), Electricity demand (temporal)	Climate data, electric load	-	Test method for system planning and assessment
Fripp et al.	Heuristic	Power systems	Reduced Model	load profile (temporal), Electricity demand (temporal)	Power generation, electricity demand	-	Test method for system planning
Haller et al.	Heuristic	Long term decarburization strategies	Reduced Model	Load profile (temporal)	Fluctuation of renewable supply	-	Test method for scenario evaluation
Belderbos et al.	Heuristic	Power plant	Reduced Model	Electricity demand (temporal)	Electricity demand	-	Test method for system planning

I.6. Extrapolation of results

The process of time reduction is also directly related to the method of extrapolation of the results found by the selected days. The adequacy of the extrapolated results and their proximity to the real values are the indicators for the success or failure of the method. Extrapolation is usually performed by two main methods: multiplying the obtained results by a proportion or multiplying the obtained results by the weight of the group represented by a day. The former is usually used in heuristic and iterative approaches. Multiplying by a proportion is the multiplying of the results of a short sequence simulation by a single scalar depending on the number of elements (days) in the selected data set (short sequence), so it gives the same effect for all the days of the sequence regardless of their real representation. On the other hand, extrapolation by actual group weight multiplies each value by a unique scalar that depends only on the weight of the group it came from. This way of extrapolation is inevitable in clustering algorithms where the scalar depends on the size of each cluster. Most studies in the literature used the extrapolation by a proportion when not using clustering or grouping algorithms.

I.7. Analysis and discussion

The predilection by the researchers into using clustering algorithms rather than other approaches, with a special interest in the K-means approach was noticeable. This distribution is shown in Figure I-14. This interest in clustering approach is due to its good performance in achieving the objective. Poncelet et al. [80] showed that in their study. After validating the accuracy of their MILP based optimization approach for different cases as shown in Figure I-15, they compared the performance with the other approaches. Figure I-16 compares the accuracy of their approach (denoted OPT) by that of a heuristic approach (denoted H), hierarchical clustering approach (denoted CA) and random selection approach (denoted RS) for two representative days. Since each approach takes different computational time to converge, the figures are plotted with generalized axes for better visualization. Time is recorded in (%) representing the advancing course of the process. Evaluated criteria are normalized to keep all plots in the same range (between 0 and 1).

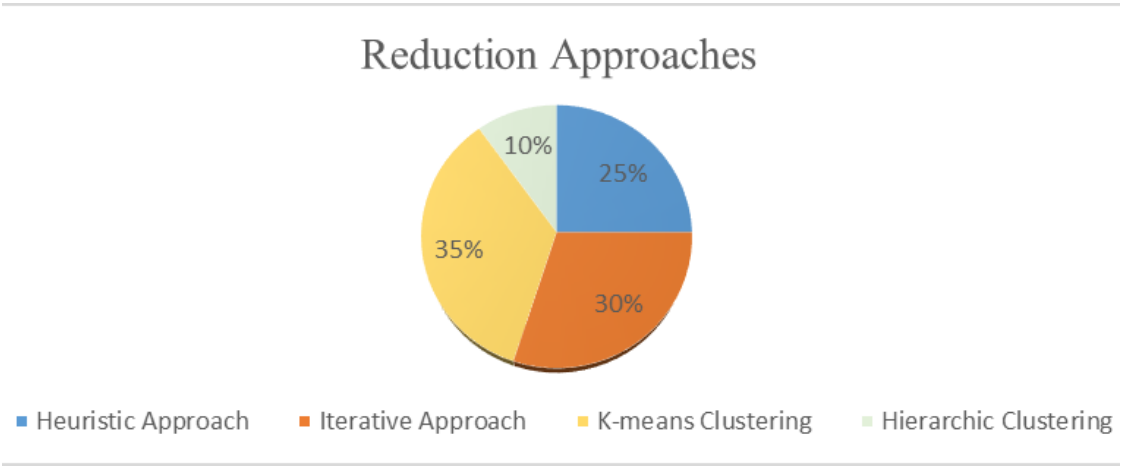


Figure I-14. Distribution of the reduction approaches as found in the literature.

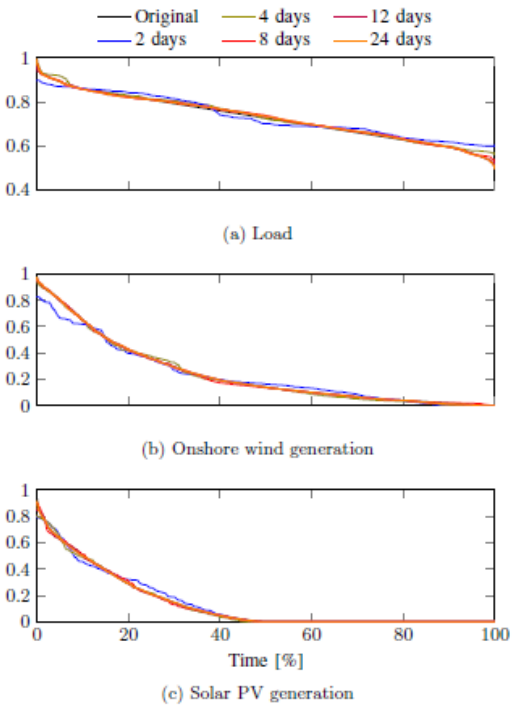


Figure I-15. Approximation of the duration curves using the OPT approach to select a varying number of representative days [80]

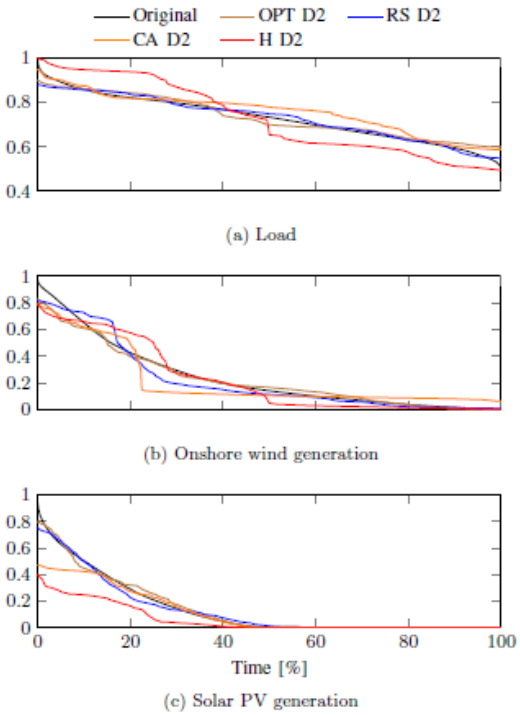


Figure I-16. Approximation of the duration curves for two representative days selected by the different approaches [80]

While the heuristic approach showed the worst accuracy, the accuracy of the other approaches was good with almost the same performance. However, they noted that the time execution cost was least in heuristics and clustering. In addition to that, while the implementation cost was best in random selection and heuristics, clustering

approach showed a better implementation cost than iterative by MILP with a relatively good flexibility [80].

On the other hand, Kotzur et al. [82] validated the capability of each of k-means, k-mediods and hierarchical approaches to regenerate the annual performance for different systems and for different number of selected days with smoothed typical periods. The first aspect found was that while energy systems based on centralized supply resources (CHP systems and residential systems) can be well represented with a few typical days, energy systems heavily relying on storage technologies (island system) could not be properly represented by independent typical days at all. In addition to that, hierarchical approach showed a better performance than the k-mean and k-mediod approach in both computational load and reproducibility. The following figures shows the relative errors for the approaches used on the studied CHP system (Figure I-17) and residential system based on heat pumps (Figure I-18).

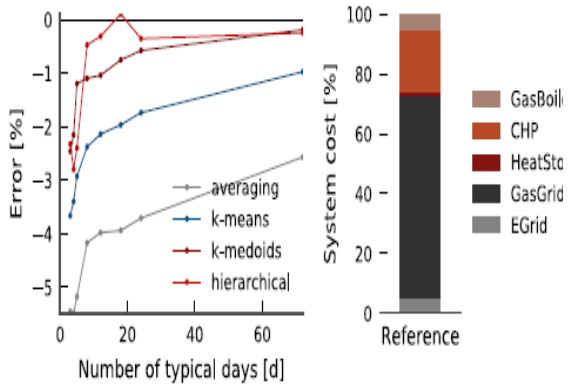


Figure I-17. Relative error for the case of CHP system [82]

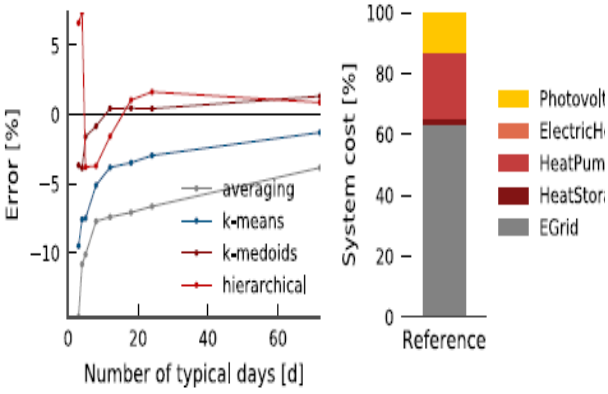


Figure I-18. Relative error for the case of residential system based on heat pumps and photovoltaics [82]

The work done in the literature concluded that in function of the flexibility of the approach and the capability of studying special predefined days, heuristic approach comes in the lead while clustering algorithms are not very favorable. The reason behind that is the way the approach itself works, while clustering algorithms search for the best day selection by a machine learning process, simply heuristic approach works by preselection of days and performing simulations on this sequence. In addition to that, heuristics is the simplest way of reducing algorithms where no difficult coding with high

math equations is required through the procedure. However, when it comes to the precision of the obtained results after simulating the short sequence and comparing its compatibility with the simplified sequence, the performance of the heuristic approach is the worst. Expecting precise results from a heuristic method requires high experience by the operator to get an efficient sequence and not wasting time by trial and error. This is not found in clustering and iterative cases where it is left for the computer to do all the trials and give finally the selected sequence, with a higher precision, and a shorter period. This approach comparison is summarized in Table I-3.

Criteria	Iterative Approach	Heuristic approach	Clustering
Precision	Good	Average	Very Good
Flexibility	Good	Very Good	Good
Simplicity	Good	Very Good	Average

Table I-3. Comparison of the approaches

On the other hand, the diversity of the case studies found in the literature does not make us able to favor a method over another. The efficiency of the method is directly related to the system studied or optimized. However, it was shown that the same method could be used to study several case studies, which is a major benefit in the sake of defining a new approach to be used in optimization studies. Indeed, continuous modification of model parameters takes place throughout the optimization procedure of such studies. Therefore, the use of an approach that shows stability despite the modifications is necessary. In addition to that, studied models should be valid on a great domain of parameter modifications to be able to catch good optimal solutions. Having an adaptive experimental design combining the definition of the short sequence and the optimization process could not only limit the number of heavy simulations required for model validation but also accelerate them through running the reduced simulations.

As explained previously, optimization studies include a lot of simulation while searching for the optimal model. Moreover, since in reality, buildings do not exist solely while rather in groups, which interact between each other such as districts or blocks, this

complicates models and makes performing simulations time consuming. Even a single building in reality is a complex model of several parameters that constitute its energy systems and envelop. Therefore, instead of simplifying the models, which is inevitable yet not preferable, using short simulation sequences is very interesting in such studies. Ortiga et al who used a reduction approach of iterative aspect, continued their study in optimizing different cases of cogeneration and trigeneration models [78]**.** The results showed very well coherence with the optimization made based on the annual sequence with less than 10% errors except for Micro-CHP total efficiency that showed a great difference. In addition to that, they concluded that for optimization, longer time sequence does not mean better results, where a sequence of 5 days gave better results than that of 10 days. In another study of multi-period optimization of district energy systems, Fazlollahi et al used the short sequence they developed in [83] to perform their district optimization study [88]. The results showed that by selecting the adequate resources, centralized and decentralized conversion technologies and distribution networks, the environmental impacts could be reduced down to 65% and the total annual costs down to 27%. The time reduction approach used was k-means clustering algorithm where they generated a sequence of eight days and performed the simulations based on it.

I.8. Conclusion

The world population continues to increase almost linearly during the three past decades. Population growth naturally generates more activities and creates more energy need to power them. The final energy consumption is often attributed to four main economic sectors: industry, transport, residential and commercial. The industrial sector dominates the global energy consumption while the building sector in its two forms, residential and commercial, accounts for about one-third of the global demand. A demand projected to increase progressively as the global demand increases.

From this point, research to improve the different sector performances has become more supported through more funding and new policies. The building sector offers significant potential for improved energy efficiency with high-performance envelops and energy-efficient systems. Building performance simulations (BPSs) are increasingly used to design buildings because of its emphasis on sustainability.

When it comes to optimization, multi-objective optimization is more relevant than the single objective approach. Yet, performing holistic approaches on buildings, which takes into consideration both the envelope and the systems, leads to the complexity of models under study, especially when analyzing heat networks in the case of multiple buildings i.e. districts or blocks, leading therefore to unfeasible computational time expenses. Simulation of detailed building models may take several minutes in building energy simulation. On the other hand, simulation-based optimization techniques require up to thousands of simulations to evaluate the case study. Usually, very simplified models instead of detailed building models are used to avoid this issue.

However, these reductions significantly lower the performance of optimization algorithms, and may result in sub-optimal solutions. Surrogate models are among promising solutions to this problem. However, the accuracy and sensitivity of surrogate based optimization is currently not a well-developed area, especially when the number of input variables is large.

A holistic approach that might solve those doubts is a current case of interest. It is based on the reduction of input data profiles rather than the model itself. The approach evaluates annual performance, of a complex model, including both the envelope and the connected systems, starting from a short simulation sequence of typical selected days instead of complete 365 days input data profiles. The literature contains various approaches to select a representative set of historical periods. These approaches can be grouped in three main categories: Heuristic Approaches, Iterative Approaches and Grouping Algorithms. The process of time reduction is also directly related to the method of extrapolation of the results found by the selected days. The adequacy of the extrapolated results and their proximity to the real values are the indicators for the success or failure of the method.

The predilection by the researchers into using grouping algorithms rather than other approaches, with a special interest in the K-means clustering approach was noticeable. On the other hand, the diversity of the case studies found in the literature does not make us able to favor a method over another. The efficiency of the method is directly related to the case study studied or optimized.

Based on this thorough literature review, a new approach called TypSS (Typical Short Sequence) Algorithm, was developed and tested in the following chapters. The

approach is of an iterative aspect with an embedded grouping algorithm. It uses defined target criteria chosen by the user for the typical day selection. The choice of merging those two approaches together was based on the previous examinations that reflected the advantage of the grouping algorithm regarding precision and that of the iterative approach regarding flexibility profiting from advanced computer technics that automate and enhance the process without the need for direct interferer by the operator.

Chapter II Description of the Typical Short Sequence algorithm (TypSS)

II.1. Introduction

II.1.1. Objectives

As explained in Chapter 1, physical models are used in building performance simulations (BPSs). Each model is a function of a plurality of parameters related to physical systems to be modeled. In order to make an accurate prediction, the physical models are often detailed, which leads to long simulation times. It is therefore common to seek a compromise between model accuracy and reasonable simulation time. Such a compromise leads to a decrease in quality in the simulation and therefore in the prediction of the physical behavior of the system. There is therefore a need for a method making it possible to simulate a physical system using a complex model over a long time horizon while being compatible with limited computing power or computing time.

In that respect, this PhD thesis puts forward and studies a new algorithm called Typical Short Sequence algorithm (TypSS). TypSS is a method for determining a series of typical days allowing to obtain a short simulation sequence to predict the thermal behavior of a detailed modeled building, composed of the envelope and connected systems. The approach is of an iterative aspect with an embedded grouping algorithm. It uses defined target criteria chosen by the user for the typical day selection. The approach employs averaged and cumulative values of target criteria to evaluate both temporal performances per period and annual performances as a complete simulation. The aim is to replicate the annual performance profile of the chosen criteria of the system, including annual global values, which could be later used in system characterization or optimization. These criteria are part of the performance functions of the model, i.e. the output of a model's simulation. On the other hand, it is much easier to understand the algorithm through an example; therefore, a case study is presented in the next section and the process of the algorithm will be explained while it is applied on it

II.1.2. Case study

Being a simulation based algorithm, TypSS requires a model mod to be applied on. The model is the set of mathematical, physical and logical equations calculated under typical boundary conditions dat_T for technical decision-making. In the case of building models, the boundary conditions are model-independent data such as weather data,

occupant profiles and operation schedules that are needed for the simulation process. Figure II-1 shows a general scheme of a typical building model. The model is built from the combination of envelope (walls, windows, roof…) and the connected systems (heating, cooling, electric…) which parametrize the model equations. Model parameters sets are denoted par_{indv} with $indv$ being the tag of an individual. As explained previously, a model with a unique parametric combination is considered an individual with unique performances. Simulating an individual under the given boundary conditions leads to the calculation of the internal equations giving output results, denoted model performances. These performances are used by TypSS algorithm as target criteria $crit_{T,indv}$ to select the typical days of its typical short sequence.

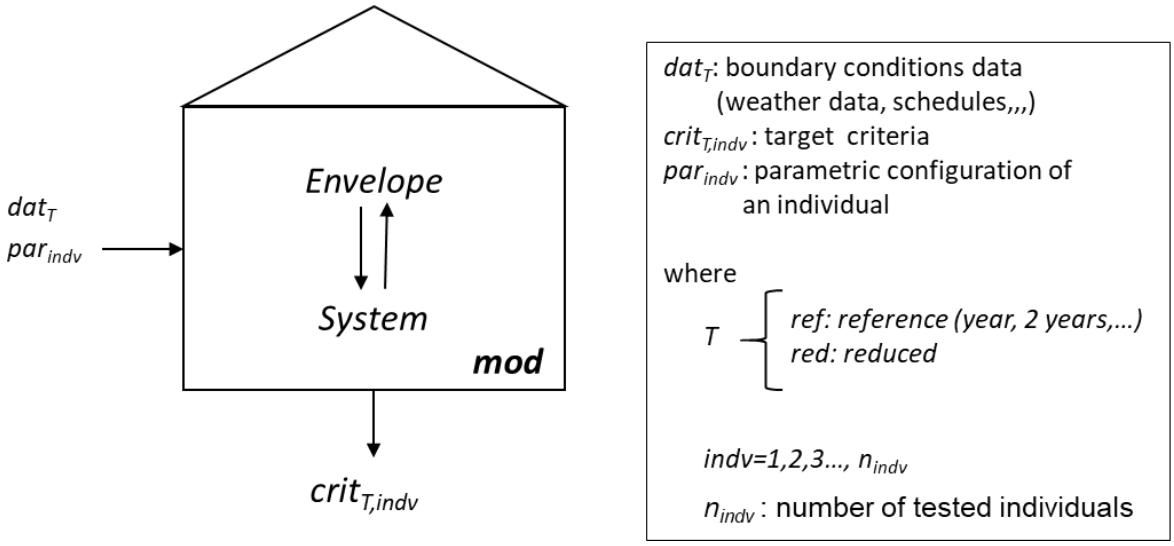


Figure II-1. A general scheme of a building model.

In the given example, the tested model mod is a building connected to a combined solar thermal and heat pump system presented in Figure II-2 and described in [89] [90]. The performance of the system was estimated based on component testing and system simulations. Most of the components used in the simulations have been validated against data from prototypes or commercial products as indicated in [90]. This model was part of the European project MacSheep [91] and is designed in Trnsys17. Within the project, the energetic and economic performances were evaluated against eight different reference systems that originate from two heat sources, i.e. air and ground, two climates Zürich and Carcassonne, and two different building standards, i.e. well insulated and medium to low insulated. The developed

components were combined into three different systems that were tested by whole system test methods. A system of them was selected to evaluate the TypSS algorithm.

The system is made of solar thermal collectors of a surface of 9.28m², a storage volume of 0.763m3 and a heat pump to assure the space heating and domestic hot water DHW supply in case of poor solar supply for the building. The envelope is a two-storey building with an effective floor area of 70m². The building is simulated as one common thermal zone. Internal capacities caused by such building structures are simplified in the simulation as one 200 m² large inner wall, thus representing 400 m² of wall surface. The inner (air) volume of the buildings is 389.45 m^3; the net floor area (first plus second floor) is 140 m² as explained in [90].

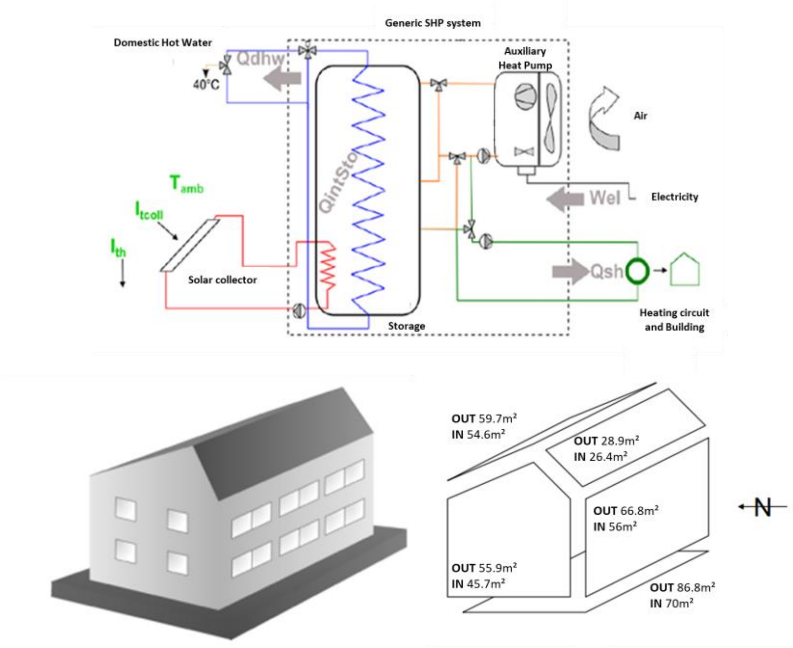


Figure II-2. Case study: (Up) solar combisystem connected to a building, (bottom) envelope parts with inside and outside facade areas.

Three model performances, or target criteria $crit_{year,indv}$, are chosen for the study and plotted in Figure II-3,

- the daily-integrated backup energy (in kWh),
- the daily-integrated energy stored in the tank (in kWh)
- and the daily-averaged internal room temperature of the building (in °C).

The backup energy is the electrical energy consumed by the heat pump to supply space heating and domestic hot water. The energy stored in the tank gives an image

of the energy content of the store based on its mean temperature. The choice of those criteria was based on the nature of the model and the aim for a later optimization of its energy consumption taking into consideration the comfort of the occupants.

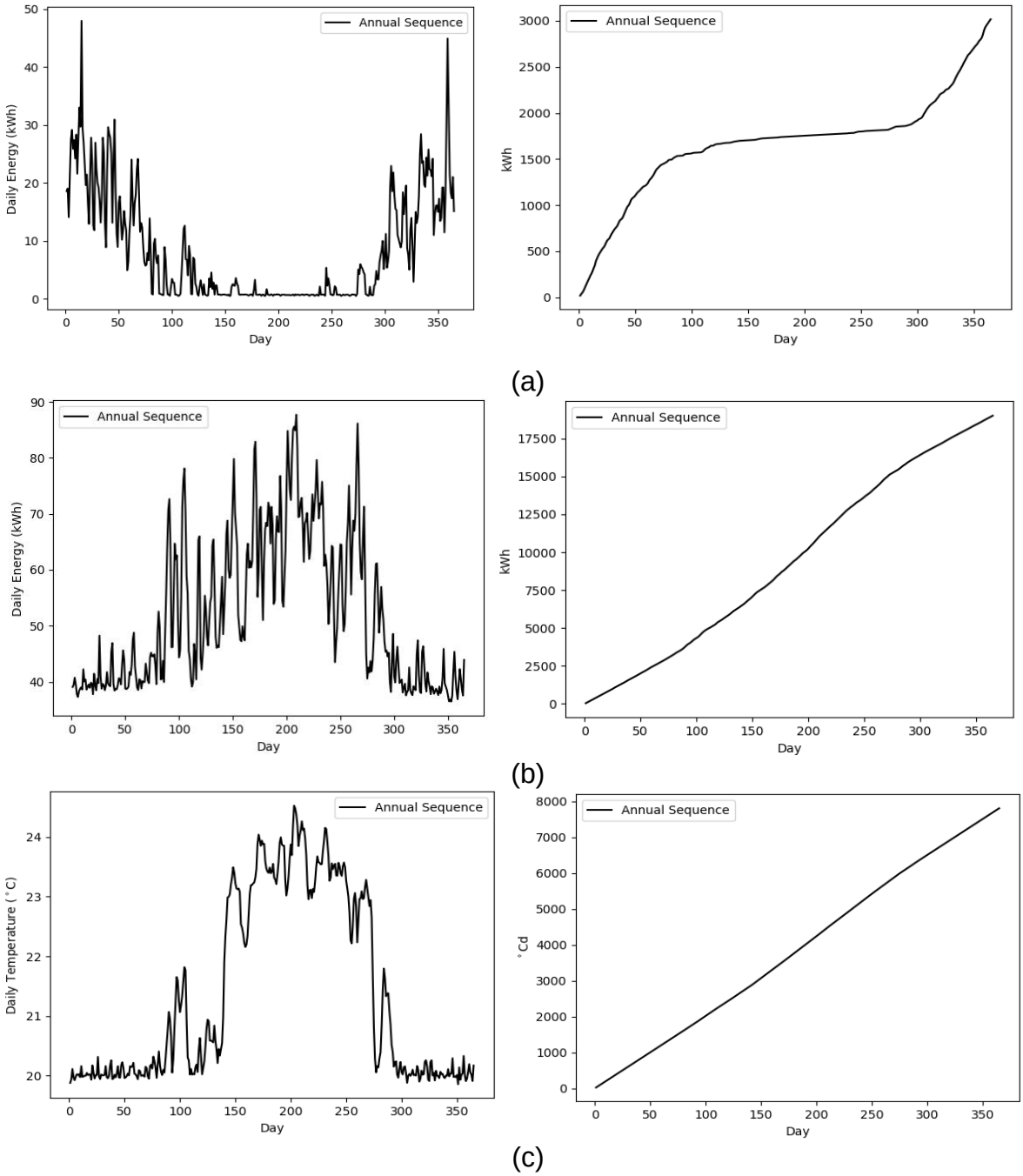


Figure II-3. Daily (Left) and cumulative (Right) profiles of the target criteria: (a) backup energy, (b) energy stored in the tank, (c) internal room temperature.

TypSS algorithm will now try to reproduce the profiles of those target criteria for all individuals at the same time.

II.2. TypSS: The process of the algorithm

II.2.1. Global Methodology

The proposed approach generates a reduced simulation sequence by dividing the year into distinct periods and selecting a representative day for each taking into consideration the performance of the sequence as a whole. This is achieved through three main parts which are launched after parametrizing the algorithm and introducing a physical model to simulate.

The main parts include:

- *Initialization* phase where initial variables are calculated and provided to the other parts of the algorithm
- *Period setting* phase that divides the time horizon (typically a year) starting from the initial data supplied from the *Initialization* phase into periods of different sizes enabling more focus on periods with higher performance changes. This is accompanied by locating period centers and proposing them as representatives of their periods forming therefore an intermediate reduced sequence.
- *Typical days' enhancement* phase that enhances the sequence generated from the *Period setting* phase by searching better representative days for each period based on global performance values.

Figure II-4 shows the global scheme of the algorithm and a detailed scheme is presented in APPENDIX A.

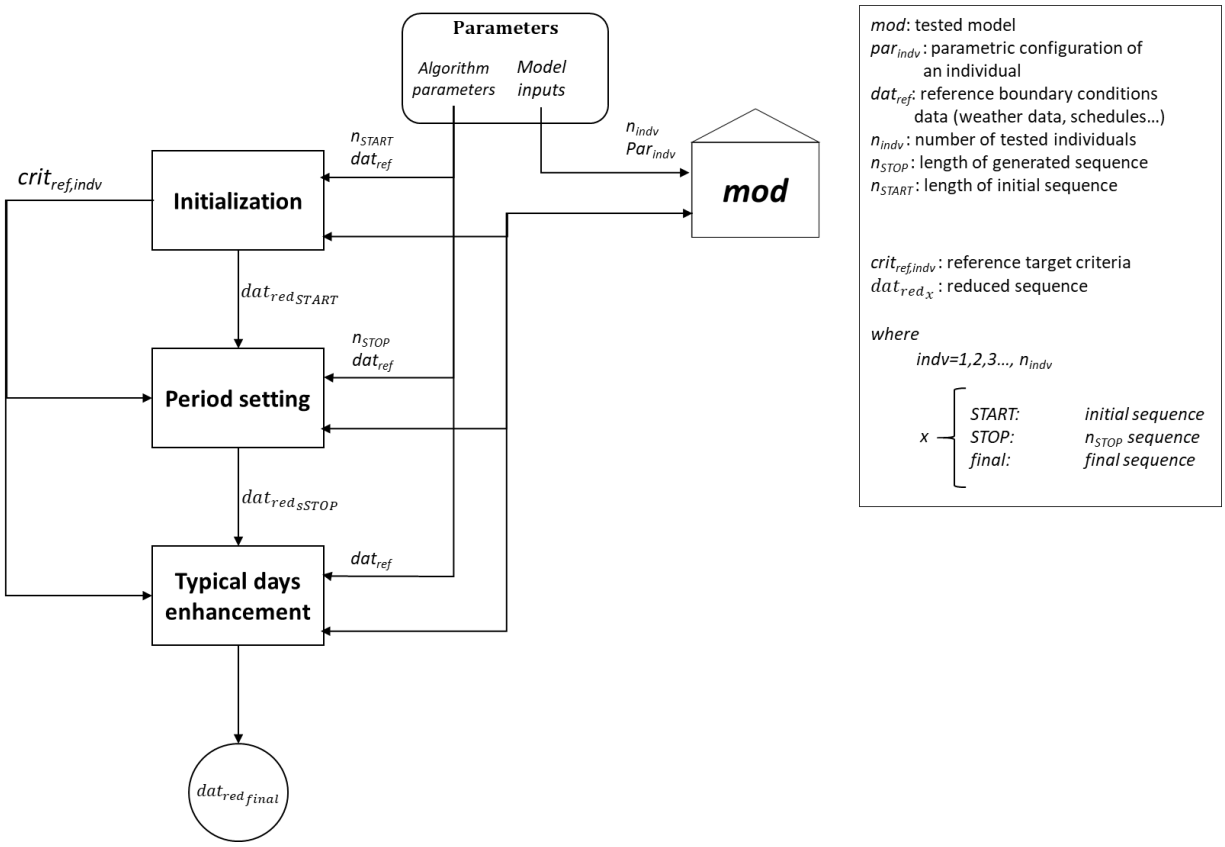


Figure II-4. The global scheme of the algorithm TypSS.

The parameters include all the data specified by the operator to the model and algorithm needed to operate. This includes:

- par_{indv}: the parametric configuration of an individual. An individual being a model with unique parametric characteristics
- n_{indv}: the number of tested individuals
- dat_{ref}: the reference data such as the weather data, occupants profile and operation schedules. Reference data could be data of a year or more depending on the performed study
- n_{STOP}: the length of the generated sequence or number of typical days produced
- n_{START}: the length of the initial sequence or number of initially considered typical days

Initialization and *Period setting* phase take those inputs and produce intermediate short sequences, $dat_{red_{START}}$and $dat_{red_{STOP}}$ respectively while the *Typical days' enhancement* phase produces the final sequence $dat_{red_{final}}$. In addition to that, the

reference target criteria $crit_{ref,indv}$, which the algorithm bases its selected days on is generated, for each individual, in the *Initialization* phase. It is supplied to the two following parts to produce their sequences. The process in each block is explained in details in the following sections.

II.2.2. Parameters

As shown in Figure II-4, several data are essential for the algorithm to function. They can be divided into two groups:

- Parameters related to the algorithm itself:
  Those parameters control the process of the algorithm and serve as triggers or breaks of its different parts. Those inputs include:
  - the length of the initial sequence n_{START} taken as four days in this example. This way, the initial sequence roughly represents the four seasons as the starting point of the algorithm.
  - the length of the generated sequence or number of typical days produced n_{STOP}. This parameter will be the breaker that allows the transition from dividing the year into partitions, *Period setting* phase, to global performance assessment and improvement, *Typical days' enhancement* phase.

  In addition to triggers and breakers, there exist data base related to the day selection process inside the algorithm:

  - the reference data dat_{ref} which represent the boundary conditions of the model. Yearly occupants' profile and operation schedules were provided to the model. In addition to that, yearly weather file of Chambery France, which has a moderate climate, cold in winter and relatively warm with occasional showers in summer was selected as the climate data file. Therefore, the reference case is that of one year and the reference data are now denoted dat_{year}.
- Parameters related to the case study, typically known as "inputs" of the algorithm, which define the nature of the tested model through assigning its parameters. Those inputs include:
  - number of tested individuals n_{indv}which has been fixed to five in this study.

- the different parametric configurations par_{indv} which define an individual.

The algorithm evaluates at least one individual I_1. On the other hand, for life cycle assessment, statistical studies or optimization studies, the parameters of the models are modified which affects the output of the simulations. Those outputs are later used to define the best performing model. However, when it comes to sequence reduction methods, those outputs are used as target criteria $crit_{T,indv}$ as stated previously. Thus, a sequence which was generated based on a certain output data of a certain parametric combination par_{indv}, might not replicate other outputs of the same model with different parameters since the initial conditions are now different. Therefore, and in order for the generated sequence remaining applicable in such parameter-modifying studies, it is essential that a single short simulation sequence would be able to replicate the performance functions despite a great number of parametric modifications so that the obtained results are reliable throughout the whole study and with all cases. For this sake, the TypSS algorithm was adapted to work simultaneously on several individuals I_{indv}. For instance, modifying the surface of the solar collector (SCOLL), the volume of the storage tank (VST) of the solar combisystem and the thickness of the insulating material (INS) of the building in the case study leads to different individuals I_{indv}. Using Latin Hypercube Sampling (LHS), 50 individuals I_{50} of the model in Figure II-2 were generated with SCOLL ranging between 6.5-25m², VST ranging between 0.3-1m^3 and INS ranging between 0.04-0.3m. Those ranges were put randomly yet still respecting sizing limits in such systems. The 50 individuals are presented in Figure II-5 and listed in Appendix B.

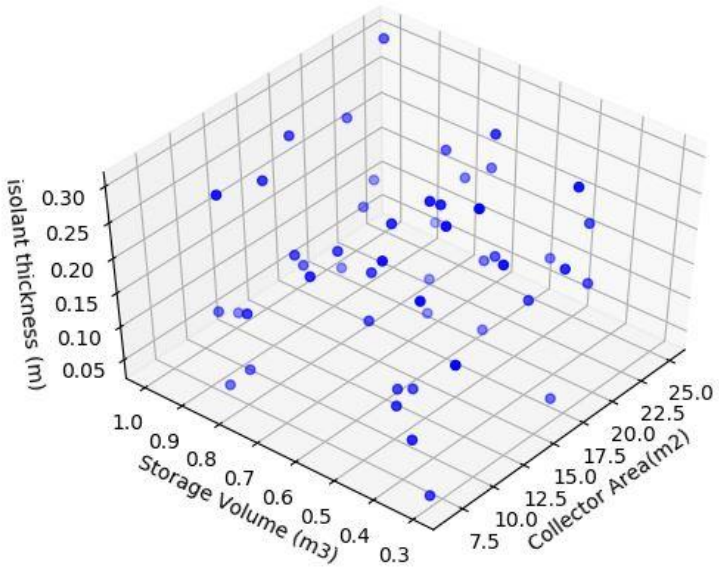


Figure II-5. Fifty individual samples I_{50} found by Latin Hypercube Sampling.

The aim now is to generate a single short sequence that can predict the annual performances of all 50 individuals. However, running TypSS on the 50 individuals is unpractical and very time consuming, so sampling should be done to select a smaller yet representing number of individual out of the 50. By clustering based on their performance, the individuals are divided into five clusters, of whom a single individual is selected from each. Therefore, five distant individuals I_5 are then selected out of the 50 to be run by the TypSS algorithm.

The selected individuals are now called “tested individuals” and are presented in Table II-1.

$indv$	par_{indv}		
	Collector surface (m²)	Storage volume (m³)	Isolation thickness (m)
1	6,5	0,3	0,04
2	7,85	0,61	0,05
3	7,08	0,77	0,25
4	17,49	0,88	0,07
5	25	1	0,3

Table II-1. The parametric characteristics of the five initial individuals.

II.2.3. Initialization

The *initialization* phase takes in addition to the model, two main parameters to initialize the process. The dynamic simulation of the n_{indv} individuals of the model under the imposed yearly boundary conditions dat_{year} will result in the individuals performances which will be used as target criteria $crit_{year,indv}$ by TypSS. This output will be used inside the algorithm itself as a reference. Running a complete simulation of the detailed model is a very important step and gives the algorithm the ability, in every iteration, to decide whether the test sequence is a good one or needs to be modified in the following iteration.

In addition to that, the block generates an initial reduced sequence $dat_{red_{START}}$ based on the previous parameters, i.e. target criteria $crit_{year,indv}$ and annual data dat_{year}. The method comprises a step of dividing the time horizon T, year in the example, into a plurality of periods $\Delta T^{day_{n_{start}}@day_{n_{last}}}$ where $day_{n_{start}}$ is the index of the first day of the considered period and $day_{n_{last}}$ is the index of the last day of the period. The schematic example of this step is given in Figure II-6 in which the year has been divided into four equal periods with n_{START}= 4 days. For instance, the second period after dividing the year into four quarters is $\Delta T^{92@182}$ beginning at the 92nd day of the year (included) and ends at the 182nd day (included).

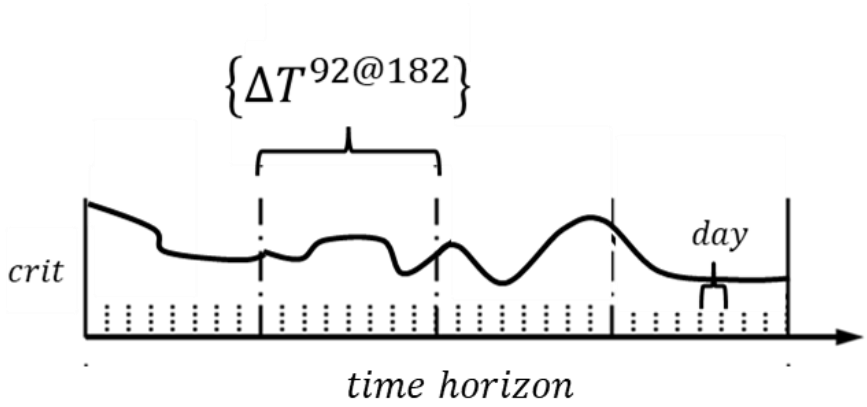


Figure II-6. Schema of a profile and dividing the year T_{sim} into four initial periods.

The method then comprises, for all the periods a step of determining the characteristic day d^n for each period with $n \in [i, j]$. This step is illustrated in Figure II-7. The curve shows the simulation performed for a given individual over the entire considered

period. The marked section shows the part of the simulation corresponding to the typical day d^n, for which the reduced criteria of the individuals are considered.

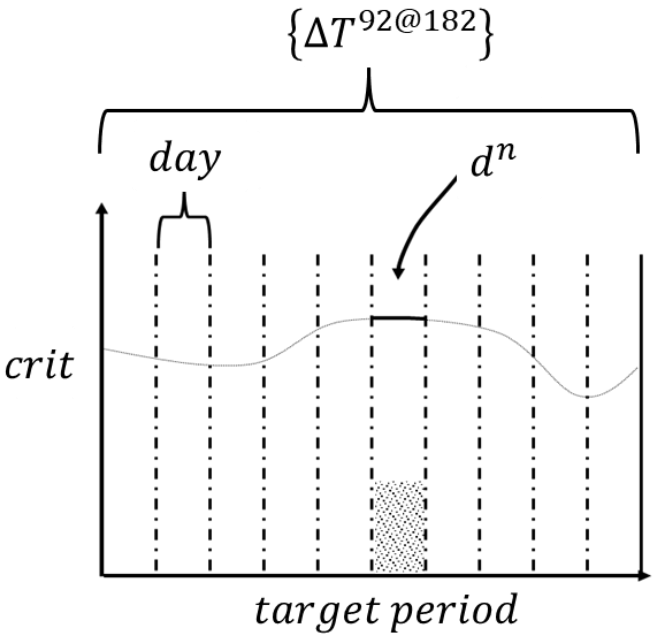


Figure II-7. Profile of a criterion in period $\Delta T^{92@182}$ showing the day distribution in the period and a characteristic day d^n.

To identify the typical day, the method calculates the Euclidean distance, the straight line distance between two points, from the criteria of each day to those of all the other days in the period. The day showing the smaller sum of distances to the other days is selected as the representative day of the period. For example, a point whose coordinates are given by the values of the three target criteria can represent each day of the considered period. Then, the center of dispersion is calculated for the considered period. The day associated with the closest point of this center of dispersion is chosen as the characteristic day d^n. This approach is inspired from the k-medoids clustering algorithm.

As already mentioned, this step takes all individuals into account: it is therefore a matter of taking into account, for a given period, a number of points equal to the number of days of the period multiplied by the number of individuals considered. For example, if the number of individuals is equal to five (5) and the number of days for the period considered is equal to ninety-one (91) then the selection will be made among the four hundred fifty five (91 × 5 = 455) points thus constituted to select only one point and therefore only one representing day. This is presented in Figure II-8 showing in 3D,

regarding the three target criteria, the distribution of days of each period (in blue) and the selected characteristic day d^n (in red) starting from four initial periods.

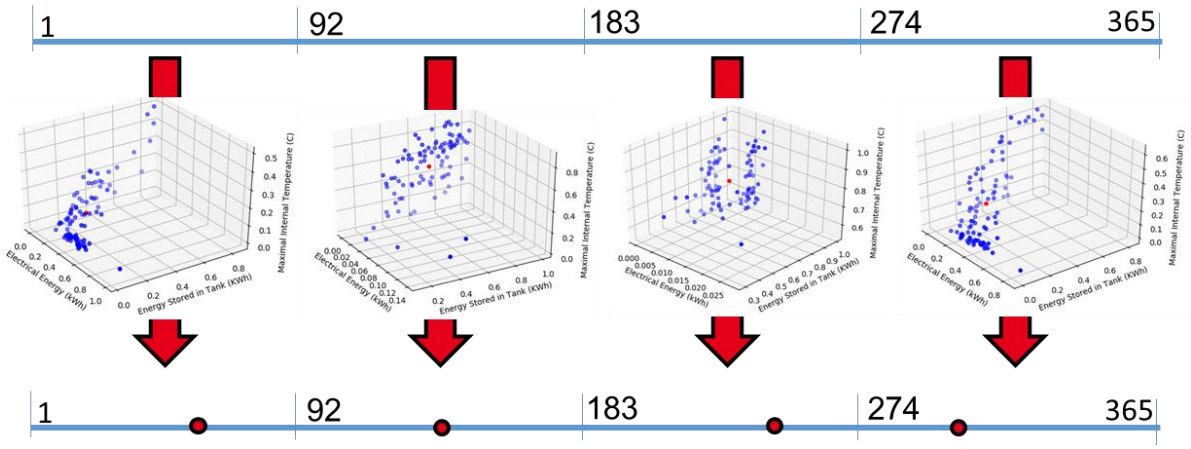


Figure II-8. The process of generating the initial sequence of four days.

When the characteristic days of each period are found, they are isolated and stacked after each other respecting their true order in the year. Thus, a first profile of the reduced sequence dat_{red_4} of four typical days (in this example) is generated and will be directly used by the following part of the algorithm. The process of day selection, isolation and stacking after each other is presented in Figure II-9 showing an example of an arbitrary annual profile dat_{year} being transformed to a reduced one dat_{red}. Selecting days from distinct parts of the year will cause a discontinuous profile as shown in Figure II-9. TypSS does not work on limiting those discontinuities but rather replicate the performance of the model despite their presence. This specificity is not detrimental for different evaluated case studies like the one in this example since the envelope has the capability to filter its effect by the internal temperature but could be possible for some cases (specific controller...). However considering averaged values could help tackle those discontinuities if needed and smoothens the obtained curves but it is not in the scope of this study.

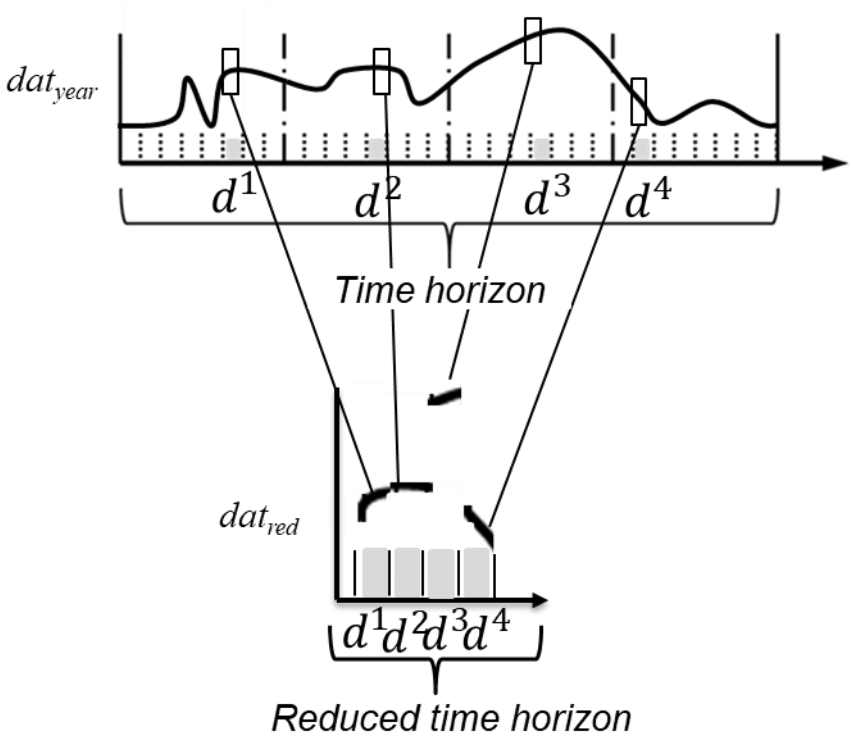


Figure II-9. The process of generating the reduced profile of the initial sequence starting from the annual one.

The process followed in the *initialization* phase is presented in Figure II-10 showing the annual target criteria of each individual $crit_{year_{indv}}$ and the initial short sequence $dat_{red_{START}}$ as outputs that will be used in the following steps of the algorithm.

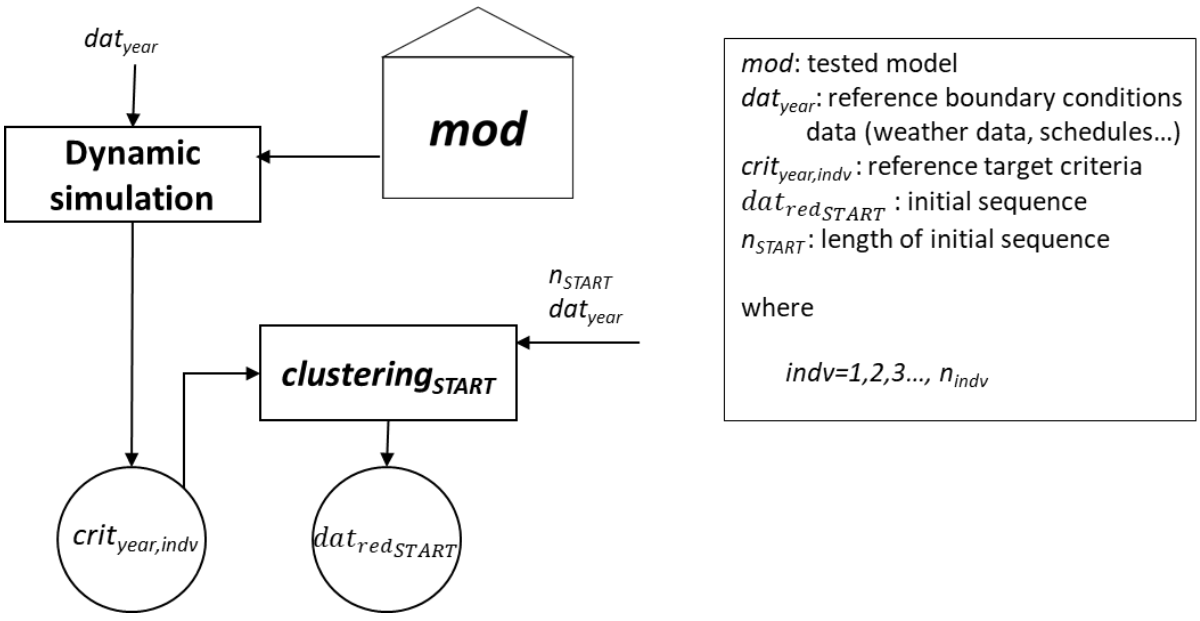


Figure II-10. The general process of the initialization phase.

II.2.4. Period setting phase

Figure II-11 presents the iterative process followed by *the Period setting* phase starting from an initial sequence $dat_{red_{START}}$ till the output sequence $dat_{red_{STOP}}$.

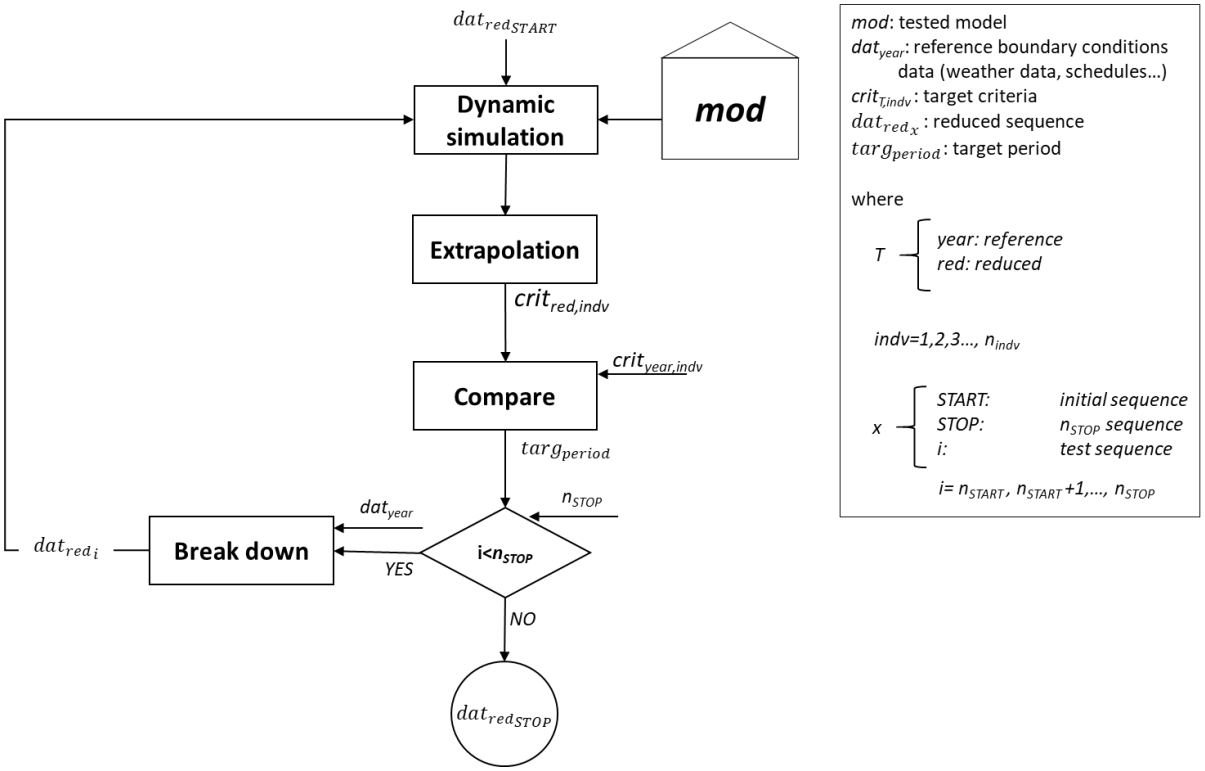


Figure II-11. The general process of the Period setting phase.

At this stage of the algorithm, reduced dynamic simulations are applied. Using the physical model mod, individuals I_{indv} are characterized with unique parameters par_{indv} and the initial reduced sequence $dat_{red_{START}}$. n_{indv} simulations are carried out on them. The data of the last day of the sequence $dat_{red_{START}}$ is used to initialize the simulation process. $dat_{red_{START}}$ includes a short input data file containing all climate and schedule inputs (occupants, electric loads…) after extracting them from the annual data files dat_{year} in addition to the last day of the reduced sequence, duplicated and placed in the beginning of the sequence to define the model's initial state.

The reduced simulations are very fast. Using a short sequence (of four days for instance) is much faster than running the same simulation on a 365 days sequence. In order to evaluate the performance of the simulated short sequence, a comparison is performed between the periodic value of the annual target criteria $crit_{year,indv_n}$ and those obtained by the reduced dynamic simulation $crit_{red,indv_n}$.

Based on the size of the period, the periodic reference values $crit_{year,indv_n}$ are calculated from the complete annual simulation performed previously for each of the tested individuals. If the criterion is extensive, the periodic reference value will be the sum of the criterion performance of each day in the period (eq. 2-1). If the criterion is intensive, the criterion average of the period is considered (eq. 2-2).

$$crit_{year,indv_n} = \sum_{j=day_{n_{start}}}^{day_{n_{last}}} crit_{year,indv_{n,j}} \quad n \in [1, n_{STOP}] \tag{2-1}$$

Where

$crit_{year,indv_{n,j}}$ daily value obtained by reference annual simulation

n_{STOP} length of generated sequence

$day_{n_{start}}$ first day of period n

$day_{n_{last}}$ last day of period n

$$crit_{year,indv_n} = \frac{\sum_{j=day_{n_{start}}}^{day_{n_{last}}} crit_{year,indv_{n,j}}}{n_{size}} \quad n \in [1, n_{STOP}] \tag{2-2}$$

Where

$crit_{year,indv_{n,j}}$ daily value obtained by reference annual simulation

n_{STOP} length of generated sequence

$day_{n_{start}}$ first day of period n

$day_{n_{last}}$ last day of period n

n_{size} size of period n

In order to compare the reference periodic values with the ones found by the short sequence, extrapolation of the reduced values is needed. The values of each period is predicted by extrapolating the criteria of the selected days obtained from the short simulation by the size of the period n_{size}, i.e. weight of the period. This method of extrapolation gives a fair representation of the characteristic days in which day representing bigger periods are given a bigger weight than the ones of the smaller ones. The method of extrapolation also depends on the nature of the studied criteria. If the criterion is extensive, then the evaluated value is the period cumulative sum and therefore the reduced value is multiplied by the number of days of the period. If the

value is intensive, no need for extrapolation since the selected day will represent the average value of the entire period as explained previously. In this example, the backup energy is considered extensive while the energy stored in the tank and the internal room temperature are considered intensive where the daily average temperature is used.

The difference between the periodic reference values and the extrapolated predicted ones is then calculated and the worst performing period showing the highest difference is detected.

In case of multiple target criteria like in the one given in this section, the global difference of all the criteria is calculated. The aim is to find the best compromise for all target criteria. To do that, the values of the criteria are normalized based on the minimum/maximum of the reference values as shown in equations (eq. 2-3), (eq. 2-4), (eq. 2-5) and (eq. 2-6). This transforms all criteria into a scale ranging between [0,1] and therefore allows adding them in a single periodic scalar called $Score_{p_n}$, p standing for periodic and n for the tag of the period, to be evaluated. To calculate this scalar, the difference, between the normalized reference $\widetilde{crit}_{year,indv_n}$ and extrapolated predicted $\widetilde{crit}_{red,indv_n}$ values, Δ_{crit_n} is calculated (eq. 2-7). For each criterion, periods are ranked by assigning a score $Score_{crit_n}$ to each period n. The score classifies the periods in order from worst to best performing with respect to each criterion (eq. 2-8). Finally, for each period, the scores $Score_{crit_n}$ are added to a single scalar $Score_{p_n}$ (eq. 2-9) and the period n showing the least $Score_{p_n}$ is described to be the worst performing in compromise to all target criteria. In case of having several periods with the same minimum value of $Score_{p_n}$, the preceding period is considered since it has an influence on the following ones.

In case of multiple tested individuals like the one given in this example (five tested individuals I_5), the global difference between all the individuals Δ_{p_n} is evaluated instead of Δ_{crit_n}. Therefore, before denoting the scores of each period as previously explained in the case of a single individual and multiple target criteria, the differences Δ_{crit_n} of all individuals are added for each period right after normalization of criteria values, each with its own min/max values, and calculation of differences (eq. 2-10). This will give a single global difference for each period compromising both the target criteria and the tested individuals denoted Δ_{p_n}, p standing for periodic and n for period tag.

$$crit_{year,indv_{min}} = minimum(crit_{year,indv_n}) \quad n \in [1, n_{STOP}] \tag{2-3}$$

Where

$crit_{year,indv_n}$ period value obtained by reference annual simulation

n_{STOP} length of test sequence

$$crit_{year,indv_{max}} = maximum(crit_{year,indv_n}) \quad n \in [1, n_{STOP}] \tag{2-4}$$

Where

$crit_{year,indv_n}$ period value obtained by reference annual simulation

n_{STOP} length of test sequence

$$\widetilde{crit}_{year,indv_n} = \frac{crit_{year,indv_n} - crit_{year,indv_{min}}}{crit_{year,indv_{max}} - crit_{year,indv_{min}}} \quad n \in [1, n_{STOP}] \tag{2-5}$$

Where

$crit_{year,indv_n}$ period value obtained by reference annual simulation

$crit_{year,indv_{max}}$ maximum periods value obtained by reference annual simulation

$crit_{year,indv_{min}}$ minimum periods value obtained by reference annual simulation

n_{STOP} length of test sequence

$$\widetilde{crit}_{red,indv_n} = \frac{crit_{red,indv_n} - crit_{year,indv_{min}}}{crit_{year,indv_{max}} - crit_{year,indv_{min}}} \quad n \in [1, n_{STOP}] \tag{2-6}$$

Where

$crit_{red,indv_n}$ period value obtained by short sequence

$crit_{year,indv_{max}}$ maximum periods value obtained by reference annual simulation

$crit_{year,indv_{min}}$ minimum periods value obtained by reference annual simulation

n_{STOP} length of test sequence

$$\Delta_{crit_n} = |\widetilde{crit}_{red,indv_n} - \widetilde{crit}_{year,indv_n}| \quad n \in [1, n_{STOP}] \tag{2-7}$$

Where

$\widetilde{crit}_{red,indv_n}$ normalized period value obtained by short sequence

$\widetilde{crit}_{year,indv_n}$ normalized period value obtained by reference annual simulation

n_{STOP} length of test sequence

$$Score_{crit_n} = ranking\ of\ the\ period\ compared\ to\ the\ other\ periods\ for\ a\ unique\ criterion\ crit\ based\ on\ \Delta_{crit_n} \qquad n \in [1, n_{STOP}] \qquad (2\text{-}8)$$

Where

n_{STOP} length of test sequence

$$Score_{p_n} = \sum_{crit=1}^{n_{criteria}} Score_{crit_n} \qquad n \in [1, n_{STOP}] \qquad (2\text{-}9)$$

Where

$Score_{crit_n}$ period score for a unique criterion *crit*

n_{STOP} number of days in test sequence

$n_{criteria}$ number of target criteria

$$\Delta_{p_n} = \sum_{indv=1}^{n_{indiv}} (\Delta_{crit_n})_{indv} \qquad n \in [1, n_{STOP}] \qquad (2\text{-}10)$$

Where

$(\Delta_{crit_n})_{indv}$ difference between normalized reference and predicted criteria values for period *n* and individual indv

n_{indiv} number of initial individuals

n_{STOP} length of test sequence

Figure II-12 shows an example of the output of the previously explained process. For an initial sequence of four days and three target criteria, the reference data has been collected in four periods (eq. 2-1, eq. 2-2) and normalized (eq. 2-5) in blue. The short simulation is performed on a sequence of four days; values are extrapolated, normalized (eq. 2-6) and plotted in orange. The algorithm evaluates the differences between the two curves for each criterion (eq. 2-7) and classifies the periods' performances for each criterion (eq. 2-8). Period 2 was found to be the worst performing having the highest differences when compromising the three target criteria (eq. 2-9). In our example of five tested individuals, after calculating the differences between the period values of the two curves, we obtain five values per period, therefore, the differences are added per period to obtain a single difference assigned to each period (eq. 2-10) then equations (eq. 2-8) and (eq. 2-9) are applied to detect the worst performing period

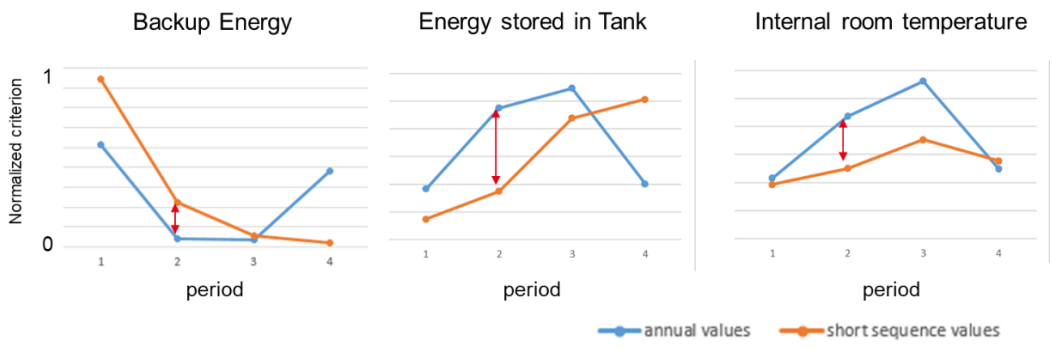


Figure II-12. Comparison between the annual and extrapolated short sequence values for each criteria after being normalized and detecting the worst performing period (period 2).

Worst performing period means that its representative day was the worst in regenerating its period's performances fairly. This means that despite performing clustering algorithm inside the period, which gives the center of the period with respect to all target criteria, not all points in the cluster, were covered. This could be caused by many reasons but one of the main and obvious ones is that the period is too big and witnesses many different instances with great deviations that a single day cannot represent them all. Considering more days for where it is not precise would solve this insufficiency of supplied data. Therefore, the period should be broken to try separating those very distinct instances into smaller more compact groups and new characteristic days can represent fairly these performances.

The algorithm targets the worst performing period, denoted $targ_{period}$, removes its characteristic day $d^{targ_{period}}$, breaks the period into two equal halves and assigns two new centers for the two new halves as explained in the previous part. This process is shown in Figure II-13. Breaking the period generates two smaller more compact periods. The points in these periods are less (less days) than the original bigger period. In this case, the selected days should be more able to represent those days because of having closer performances to their neighboring days. Due to this dividing process, the short sequence is now one day longer where a period has been replaced by two smaller ones with each having its unique representative day as shown in Figure II-13. The figure also shows that the periods are now of different sizes. Extrapolation will take place based on the size of each group and therefore assigning higher weights for representative days d^n of bigger periods n.

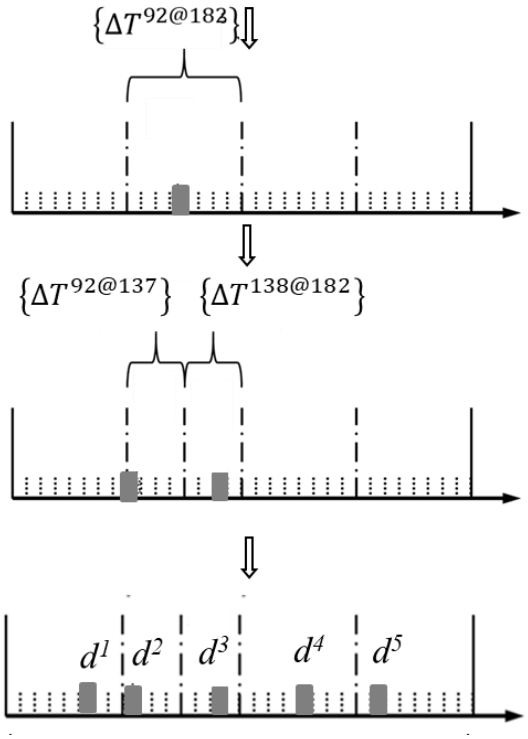


Figure II-13. The process of detecting and dividing the worst performing period.

The preceding steps of detecting periods, dividing them and replacing typical days are repeated iteratively until ending with a sequence of days, dat_{red_i}, of i equal to n_{STOP} as specified by the user, denoted $dat_{red_{STOP}}$. Giving the user the right to define n_{STOP} gives him the decision for the length of the final sequence he is interested in generating. However, it is in the perspective of this work to add an option leaving it to the algorithm to decide the length of the sequence based on the values of the performance indicators.

II.2.5. Typical days' enhancement phase

While the *Period setting* phase seems legitimate in the terms of dividing the year into unique groups with specific days' characteristics in addition to selecting days which are representatives of their periods, it doesn't take into consideration the global performance and the influence of the periods on each other. For that, the *Typical days' enhancement* phase was added. The function is of an iterative aspect as shown in Figure II-14. It starts from the reduced sequence generated by the *Period setting* phase $dat_{red_{STOP}}$ and terminates by the final sequence of the algorithm $dat_{red_{final}}$. $dat_{red_{final}}$ is a sequence of same length and period sizes as $dat_{red_{STOP}}$ but with modified typical days d^n.

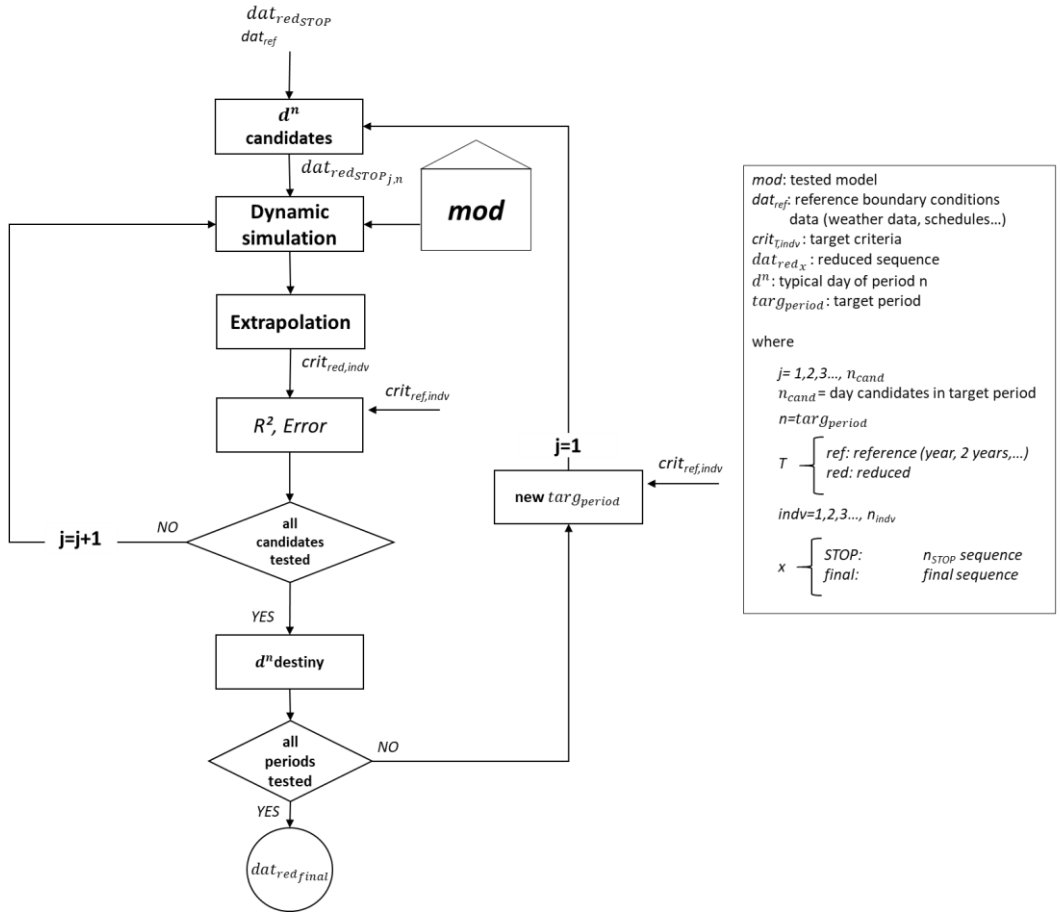


Figure II-14. The general process of the Typical days' enhancement phase.

Modification of d^n is based on the whole performance of the sequence using two global values:

- The global coefficient of determination R^2_{Global} (eq. 2-11), the product of the coefficient of determination $R^2_{crit_{all}}$ of the used target criteria with data points of all individuals simultaneously (eq. 2-12), and
- The global annual sum error E_{Global} (eq. 2-13), the sum of the criteria annual sum errors E_{crit} (eq. 2-14). In case of multiple tested individuals, E_{Global} is calculated for each individual and the maximum between them E_{max} is taken into consideration (eq. 2-15).

The annual sum error is directly linked to the main goal of the test sequence i.e. can the short sequence estimate the annual sums of the model $crit_{year,indv_{sum}}$? However, it is not enough. So, the regression coefficient is used to express how well the short sequence results describe the reference results at each time step for all criteria. Is the short sequence appropriate?

$$\mathrm{R}^2_{Global} = \prod_{crit=1}^{n_{criteria}} R^2_{crit_{all}} \quad (2\text{-}11)$$

Where

$R^2_{crit_{all}}$ coefficient of determination of a target criterion with the data points of all individuals

$n_{criteria}$ number of the target criteria

$$R^2_{crit_{all}} = 1 - \frac{\sum_{j=1}^{365}(crit_{red,indv_j} - crit_{year,indv_j})^2}{\sum_{j=1}^{365}(crit_{year,indv_j} - \overline{crit}_{year,indv})^2} \quad (2\text{-}12)$$

Where

$crit_{red,indv_j}$ daily value of the criterion obtained by the short sequence

$crit_{year,indv_j}$ daily value of the criterion obtained by the reference annual sequence

$\overline{crit}_{year,indv}$ mean value of the criterion in the reference annual sequence

$$\mathrm{E}_{\mathrm{Global}} = \sum_{crit=1}^{n_{criteria}} E_{crit} \quad (2\text{-}13)$$

Where

E_{crit} annual sum error of a target criterion

$n_{criteria}$ number of the target criteria

$$E_{crit} = \frac{\left|crit_{red,indv_{sum}} - crit_{year,indv_{sum}}\right|}{crit_{year,indv_{sum}}} \times 100 \quad (2\text{-}14)$$

Where

$crit_{red,indv_{sum}}$ annual sum of criterion obtained by the short sequence

$crit_{year,indv_{sum}}$ annual sum of criterion obtained by the reference annual sequence

$$\mathrm{E}_{\max} = max(E_{Global})_{indv} \qquad indv \in [1, n_{indv}] \quad (2\text{-}15)$$

Where

E_{Global} global annual sum error

n_{indv} number of tested individuals

The phase starts by the generated sequence of the *Period setting* phase $dat_{red_{STOP}}$. It targets a period $targ_{period}$. However, rather than breaking the period in two parts, it replaces iteratively the representative day d^n, with the tag $n = targ_{period}$, by a list of days of the same period, denoted d^ncandidates in the figure. The other selected days of the generated sequence are kept unchanged. Rapid dynamic simulations of each individual are applied with the new updated reduced sequence $dat_{red_{STOP_{j,n}}}$in each iteration and the reduced criteria profiles $crit_{red,indv}$ are obtained after extrapolation. This is accompanied by calculating the global coefficient of determination R^2_{Global} and the final annual sum error E_{Global} with respect to the reference $crit_{year,indv}$.

When all days in the d^n candidate list of the target period are tested, the previously selected day is then replaced by the day showing the highest R^2_{Global}, and a new updated reduced sequence is found achieved in the block d^n destiny in the figure. Since the global coefficient of determination is now higher, the global performance of the sequence became better than the previous one. If the R^2_{Global} has not improved the day will not be modified. The algorithm then goes to a new target period and repeats the previous loop until all periods were targeted and no more possible improvements in the reduced sequence. In that case, the final sequence $dat_{red_{final}}$is produced, the final output sequence of the TypSS algorithm.

An example of this process is shown in Figure II-15. The year appears to be divided into nine periods (n_{STOP}=9) by the *Period setting* phase giving therefore the sequence dat_{red_9} as shown in the figure. Periods' size of dat_{red_9} are not equal and are remained unchanged through the process of the *Typical days' enhancement* phase as stated previously. The function targets the first period of the sequence and starts making a day modification as shown with $dat_{red_{STOP_{1,1}}}$ $dat_{red_{STOP_{2,1}}}$ and $dat_{red_{STOP_{3,1}}}$. This is accompanied by the calculation of R^2_{Global} and E_{Global} for each reduced sequence. On the other hand, the other typical days $d^n\ with\ n \in]1, n_{STOP}\ ]$ are not changed and remain as they were returned by the *Period setting* phase. When all candidate days are tested in the target period, the new sequence $dat_{red_{STOP_{j,targ_{period}}}}$ with the highest R^2_{Global} is returned and the function goes to another target period.

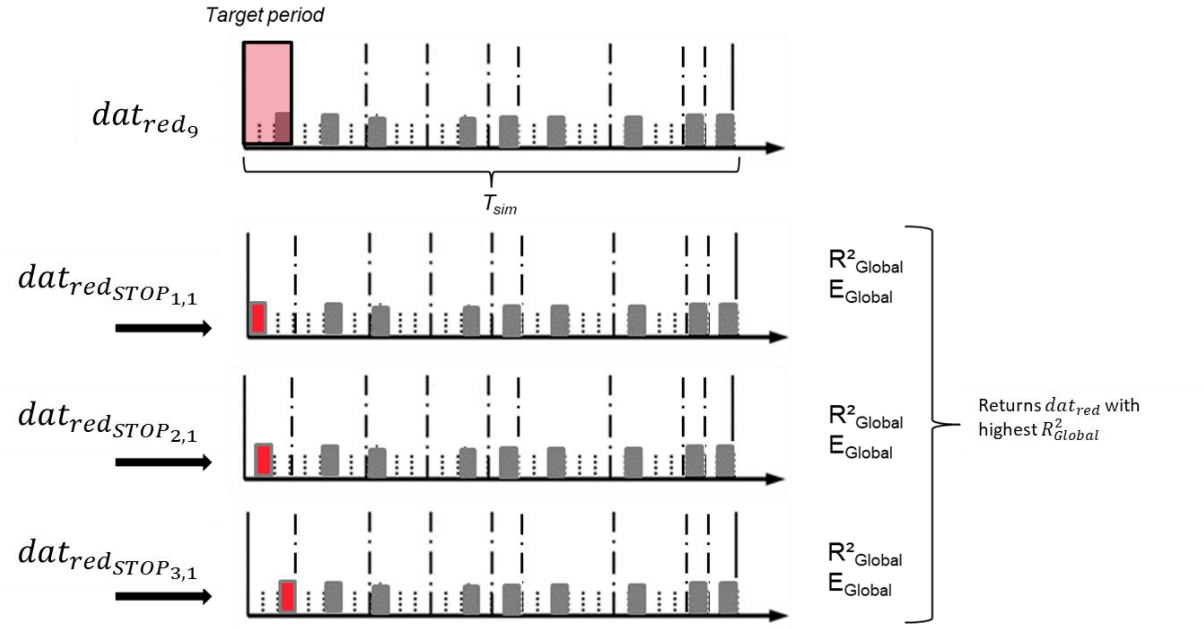


Figure II-15. The process of target period day modification.

Those previous steps are then repeated iteratively until passing through the whole sequence period by period. When all periods are tested, the algorithm finally outputs the compromise between the global coefficient of determination R^2_{Global} (eq.2-11) and global annual sum error E_{Global} (eq. 2-13). This is fulfilled by tracing all the test sequences tested through the course of the *Typical days' enhancement* phase and selecting the sequence showing the highest R^2_{Global} and the lowest E_{Global} in the case of a single individual. In case of multiple individuals, lowest E_{max} (eq. 2-15), the maximum global error recorded by the tested individuals, is considered. This will be the final short sequence. Figure II-16 shows the scatter of all test sequences obtained through the course of the phase with respect to the global coefficient of determination R^2_{Global} and global annual sum error E_{Global}, in addition to the final selected sequence.

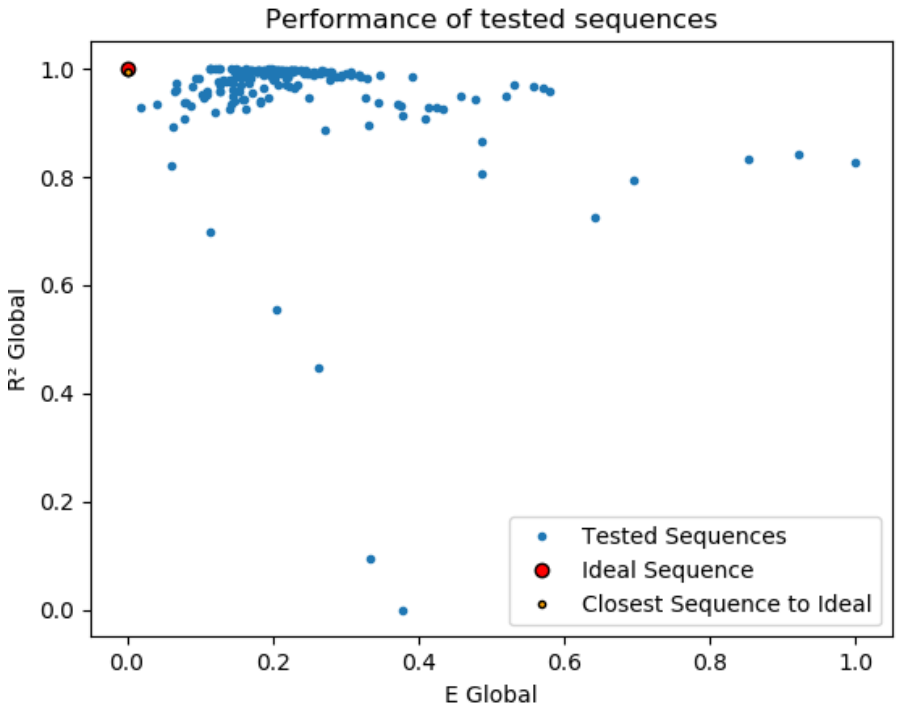


Figure II-16. Scatter of the tested sequences (blue) with respect to the global coefficient of determination R^2_{Global} and the global annual sum error E_{Global} and the selected sequence (orange).

Finally, and in order to speed up the algorithm process and produce more consistent results, two options were integrated in the algorithm in the *Typical days' enhancement* phase that can be activated by the user.

- The first option is related to the d^n candidate list the phase starts in which instead of testing all days of the target period $targ_{period}$, clustering can be applied on the set of period days. The period can thus be divided into a small number of groups where a single day is selected from each. Clustering gives the center of the group therefore running the dynamic simulation on this day will lead to performances close to the other days n the group it is representing. This approach decreases the number of reduced simulations and focuses on simulating days with different performances.
- The second option is related to the way of detecting target periods $targ_{period}$. Instead of targeting periods consecutively for possible typical day d^n modification starting from period $n = 1$, the period targets worst performing periods in the same way it detects them in the *Period setting* phase (eq. 2-1 to eq. 2-10) and starts with them first. If the new detected worst performing period

was the same as the previous even after the day adjustment, repeating the previous steps will not make any changes. Therefore, the algorithm replaces iteratively the typical days starting from the first period of the sequence where there is an influence of the previous periods on the following ones as in the basic *Typical days' enhancement* phase approach. The algorithm skips the periods that were already manipulated and there were no day modifications to avoid useless iterations. However, incase this period was detected later as a worst performing after a day modification, it will be retested since the previous initial conditions have changed now due to a change of a day in the sequence. This option might be more time consuming than the basic consecutive period targeting method since a period could be tested more than once but leads to a more consistent sequence since it takes into consideration the changes in the boundary conditions due to the continuous updating of the reduced sequence.

II.3. Conclusion

In this chapter, the Typical Short Sequence Algorithm (TypSS) for determining a series of typical days allowing to obtain a short simulation sequence has been presented and explained through an example. The approach is of an iterative aspect with an embedded grouping algorithm. It employs averaged and cumulative values of target criteria to evaluate both temporal performances per period and annual performances as a complete simulation.

The algorithm, divided into three parts requiring,

- a physical model, its parametric characteristics and external data profiles such as weather data are essential for the algorithm since it is a simulation-based algorithm that performs dynamic simulations inside its loops. Therefore, previously calculated output files are inconvenient for this type of algorithms.
- trigger and break parameters to be specified by the operator in order to control its functions which include the length of the initial reduced sequence, the length of the final reduced sequence and the number of tested individuals.

The different parts of the algorithm are:

- *Initialization* where the reference target criteria are defined. An initial sequence is also developed starting from those criteria through dividing the year into

periods of equal sizes and selecting a representative day by k-medoids clustering.

- *Period setting* phase where fast dynamic simulations of the individuals take place applying directly the short sequence on the detailed model. Comparison then is done between the annual reference data and the extrapolated reduced ones and the least performing period is detected, i.e. showing the highest difference between reference and predicted criteria. The algorithm then removes the period's typical day, breaks it into two equal halves and assigns two new centers for the two new halves by clustering. Due to this dividing process, the short sequence is now one day longer where a period has been replaced by two smaller ones with each having its unique representative day. This process of detecting and breaking down the worst periods repeats until reaching a length of a sequence of days as precised by the user.
- *Typical days' enhancement* phase was added to take into consideration the global performance and the influence of the periods on each other. It replaces iteratively the representative day of a target period by a list of days of the same period searching for a new day combination that will improve the global coefficient of determination. This step is repeated iteratively until passing through the whole sequence period by period. When all periods are tested, the algorithm finally outputs the compromise between the global coefficient of determination R^2_{Global} and global annual sum error E_{Global}. Two options for the operator are included at this stage to help the algorithm converge faster to more consistent results.

The results obtained by TypSS on the presented case study in Section II.1.2. are presented in the next chapter while activating the previously presented options in the *Typical days' enhancement* phase. The chapter also evaluates the sensitivity of the method on its input parameters, mainly the number of initial individuals, number and type of target criteria, influence of length of both the initial and final sequences. This study will evaluate the generalization ability of the algorithm preparing it for an optimization study.

Chapter III Application of TypSS and sensitivity analysis on its input parameters

III.1. Introduction

After presenting in the previous chapter the algorithm and the process it follows during its search for the typical sequence on a case study, the following chapter is divided into two sections:

- The first section introduces the results obtained after running the algorithm on a case study. Investigating the main objectives of the algorithm. Is it able to estimate accurately, after succeeding in reducing the calculation time, the annual performances and temporal profiles? Is it applicable to studies that include continuous parametric modifications such as optimization studies? This part is divided into two subsections
  - The first subsection shows the results obtained on a single individual I_1 of this case study with a unique parametric configuration for its components and compares them to results obtained by other approaches used in the literature.
  - The second subsection shows the results obtained on a set of individuals of different parametric configurations I_{indv} and therefore evaluating the generalization potentials of the algorithm.

  The parametric inputs of the algorithm in this section continued as in the previous chapter

  - Number of days in the initial sequence $n_{START} = 4\ days$ representing roughly the four seasons of the year.
  - Number of days in the final sequence $n_{STOP} = 12\ days$ to compare the obtained results with SCSPT, a method used in the literature that selects typical days based on monthly values.
  - Number of tested individuals n_{indv} fixed to one in the first subsection with simulating only one individual I_1 and five in the second subsection for five individuals I_5.
  - The modified parametric configurations that define individuals par_{indv} are the surface of the solar collector SCOLL, volume of the storage tank VST and thickness of the insulation material INS.
- The second section examines the questions stated at the end of the previous chapter regarding the sensitivity of the algorithm upon input data modifications.

What influence does the number of tested individuals, number and type of target criteria, length of the final and initial sequence have on the quality of the obtained results?

III.2. Simulation results of the case study

III.2.1. Single individual I_1

As previously shown in Chapter 2, the case study is a building connected to a combined solar thermal and heat pump system (Figure II-2). The system is made of solar thermal collectors of a surface of 9.28m², a storage volume of 0.763m³ and a heat pump. The envelope is of an effective floor area of 70m² and the net floor area (first plus second floor) is 140 m². The model is run using the weather file of Chambery, France.

The model performances used for the day selection are the same target criteria mentioned previously:

- the daily-integrated backup energy (in kWh),
- the daily-integrated energy stored in the tank (in kWh) and
- the daily-averaged internal room temperature of the building (in °C).

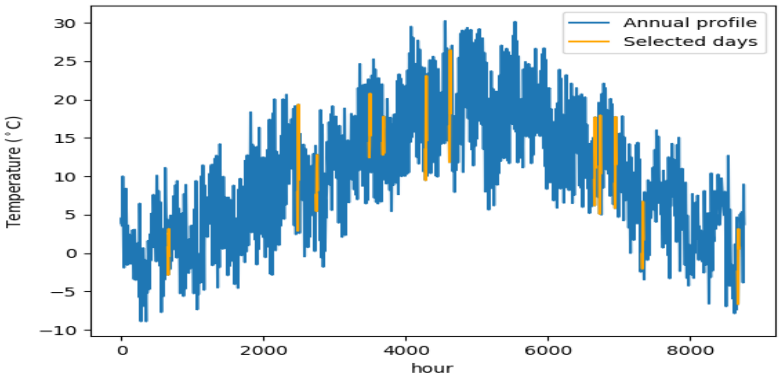
III.2.1.1. Algorithm output

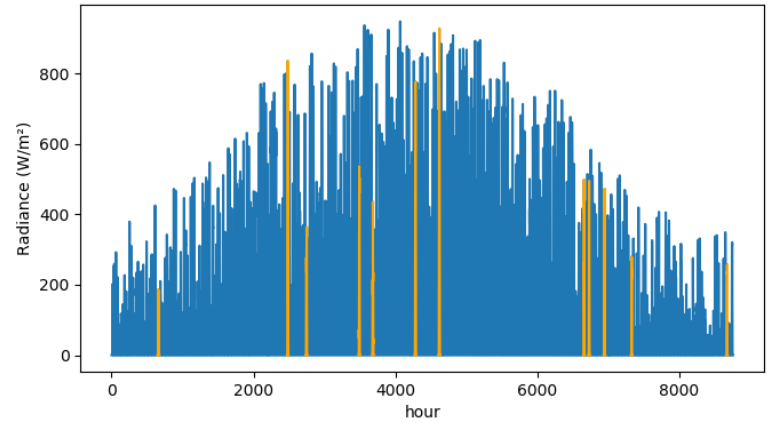
Table III-1 presents a 12 days sequence in addition to the length of each period obtained by TypSS. It is noticed that the algorithm chose to break the inter-seasonal periods of the year into smaller ones and therefore considering more days for these parts of the year. This is influenced by the nature of the used target criteria, which present high variations during this time of the year. This result appears in the table in periods 4,5,8 and 9 where the periods had between 5 and 12 days only. On the other hand, the seasonal periods, or the periods when the weather witnesses stable hot or cold weather remained of big sizes reaching up to 91 days. This was predicted since the performances of the target criteria during this time of the year have a sort of consistent profiles.

Period	1	2	3	4	5	6	7	8	9	10	11	12
Number of days	91	22	23	11	12	23	91	5	6	11	23	46
Selected Day	27th Jan	12th April	23rd April	24th May	1st June	26th June	10th July	3rd Oct	6th Oct	15th Oct	31st Oct	26th Dec

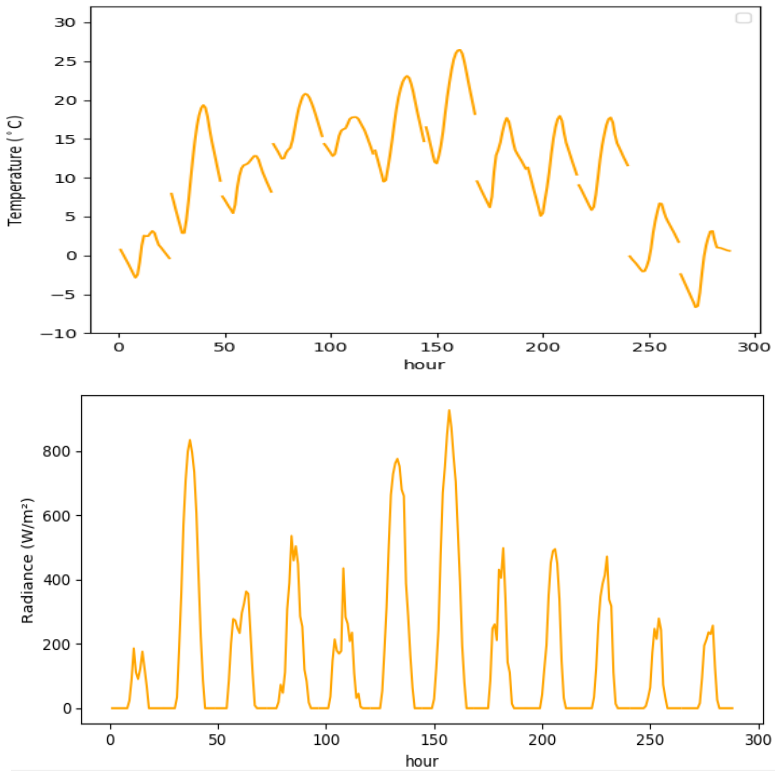
Table III-1.Typical short sequence of 12 days and the number of days in each period.

Figure III-1 shows the 12 days ambient temperature and horizontal solar irradiation profiles (Figure III-1(b)) compared to the annual reference ones (Figure III-1(a)) as an example of two of the data profiles that will be introduced to the model to be simulated on. Figure III-1(a) also shows the selected days in Table III-1 as they are distributed in the year (in orange). The values are per hour, therefore, profiles in the subfigure (a) show 365x24=8760 data while those subfigure (b) 12x24=288 data. The discontinuities between the selected days appear clearly in the reduced ambient temperature profile. Those discontinuities are noticed more around the representative days of the big-sized periods, i.e. periods 1, 7, 11 and 12. The discontinuities are due to the fact that the days are taken from different parts of the year. Since the days are distinct as shown in Table III-1, some climate characteristics, including the ambient temperature, will be discontinuous. This is not visible in the global horizontal profile in the same figure because the daily values of this characteristic always start and end with zero no matter its position in the year. As mentioned Chapter 2, TypSS does not work on limiting those discontinuities but rather regenerate the performance of the model despite their presence.





(a)



(b)

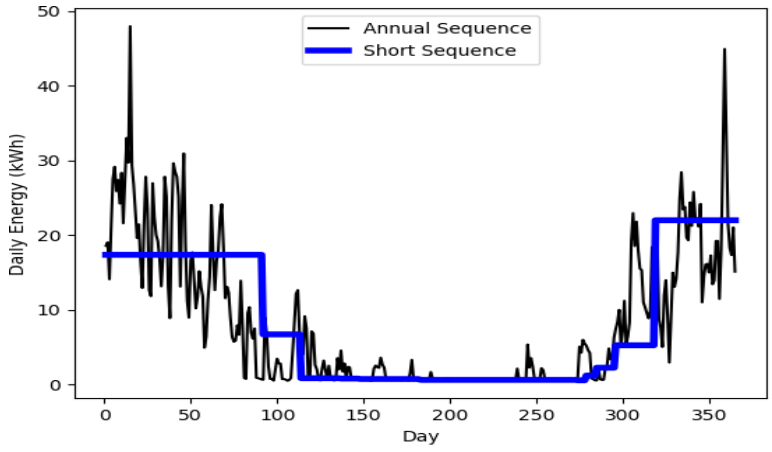
Figure III-1. The hourly ambient temperature and global horizontal radiation profiles: (a) reference annual profile (in blue) and the 12 selected days (in orange), (b) 12 selected days profile.

III.2.1.2. Temporal profiles of the target criteria

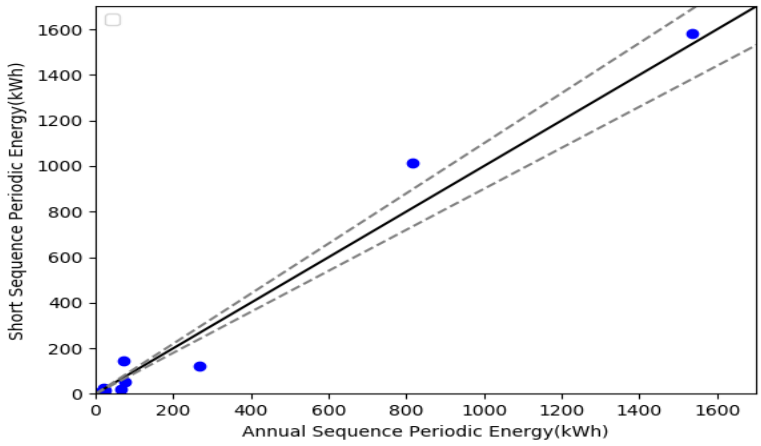
Simulating the model on the sequence of 12 days was about 25 times faster than the annual one. It took about 40secs for this case study while it takes 19mins for a full simulation with the complete sequence on the same computer configuration. Figure III-2 to Figure III-4 show the results obtained for each period when simulating the model

with the short sequence comparing them to the reference values obtained when running a full year simulation.

Regarding the temporal profiles (Figure III-2(a), Figure III-3(a) and Figure III-4(a)), the figures show profiles of daily values, and therefore the reduced simulation generates step-like profiles. Each step is the repetitive performance of the selected day through the period it is representing. The plots show that the output of the short simulation sequence are of the same profile of the annual ones. The backup energy curve from the short sequence simulation decreases gradually with time until it reaches its minimum during the summer period. Then it starts increasing again following the same profile as the reference profile. This evolution is reversed in the case of the other two criteria. The curves start by their minimal values before they increase gradually through the year until reaching their maximum in summer after which they start decreasing gradually through the year; still following the evolution of the reference profiles.

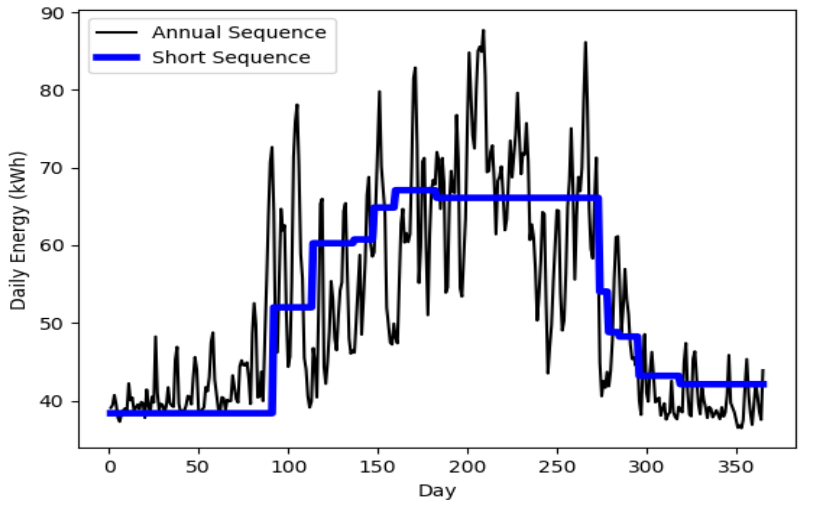


(a)

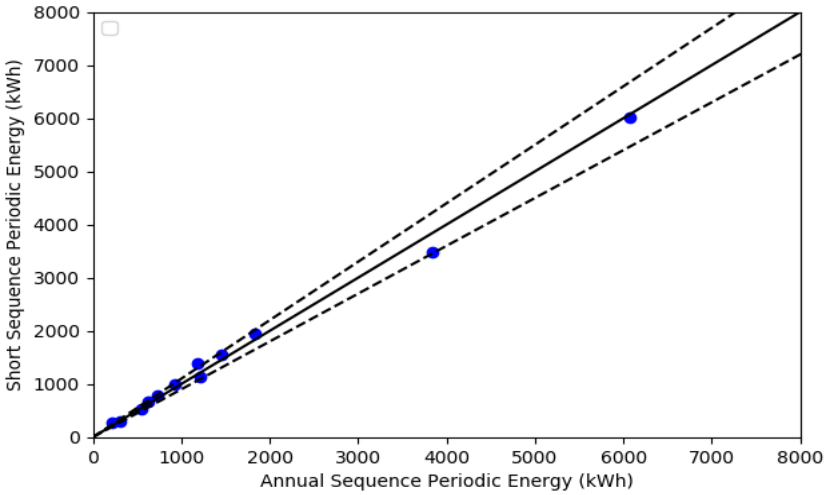


(b)

Figure III-2. Comparison between reference and extrapolated predicted backup energy: (a) temporal daily profile, (b) integrated values per period.

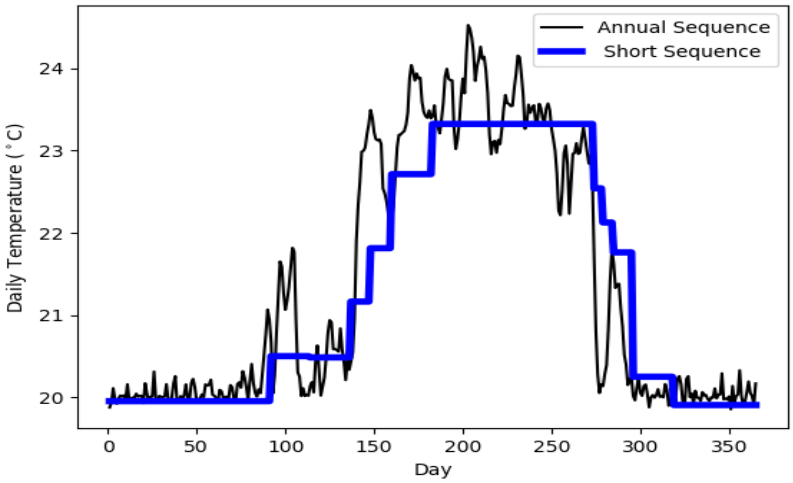


(a)

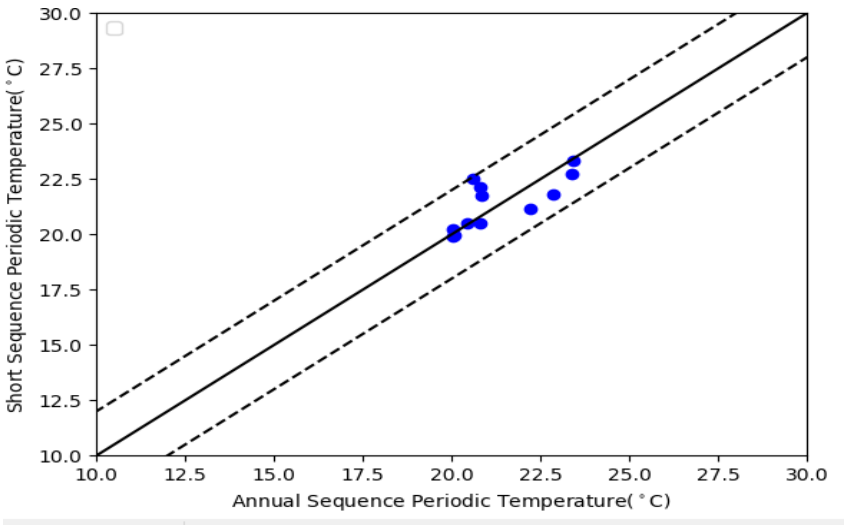


(b)

Figure III-3. Comparison between reference and extrapolated predicted energy stored in the tank: (a) temporal daily profile,(b) integrated values per period.



(a)



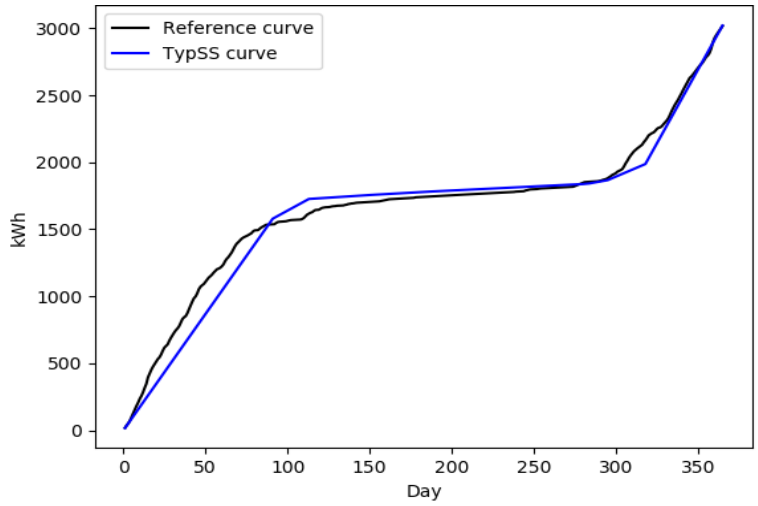
(b)

Figure III-4. Comparison between reference and predicted internal room temperature: (a) temporal daily profile, (b) averaged values per period.

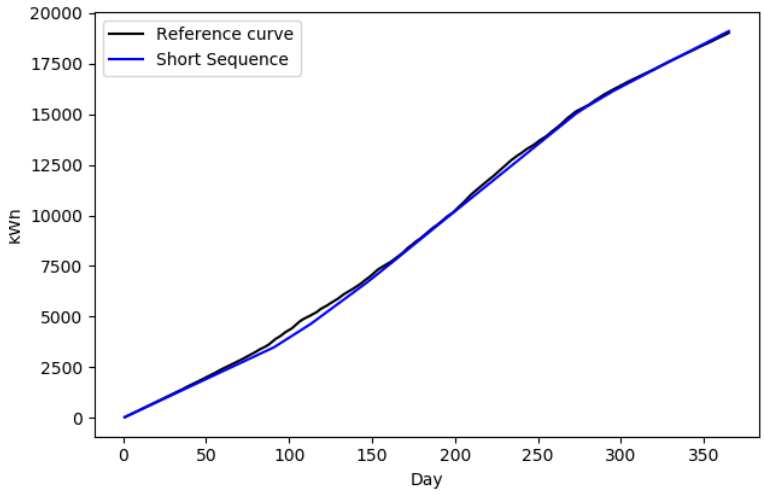
Periodic values are plotted in Figure III-2(b), Figure III-3(b) and Figure III-4(b). The comparison between the reference and predicted values is applied through plotting one with respect to the other. Each point in the plot is a period value of the corresponding criterion taking the annual value as the abscissa (x) and the predicted value as the ordinate ($f(x)$). If the prediction was accurate, predicted values will have similar or very close values to the annual ones falling therefore on the line of equation $f(x) = x$ known as identity line. Figures show high correlations around the identity line with points scattered around it. For most of the periods, values are within the 10% error limits for the energetic criteria and the ±2°C limits for the internal room temperature when comparing the predicted to the reference values. .

III.2.1.3. Annual values and cumulative profiles of the target criteria

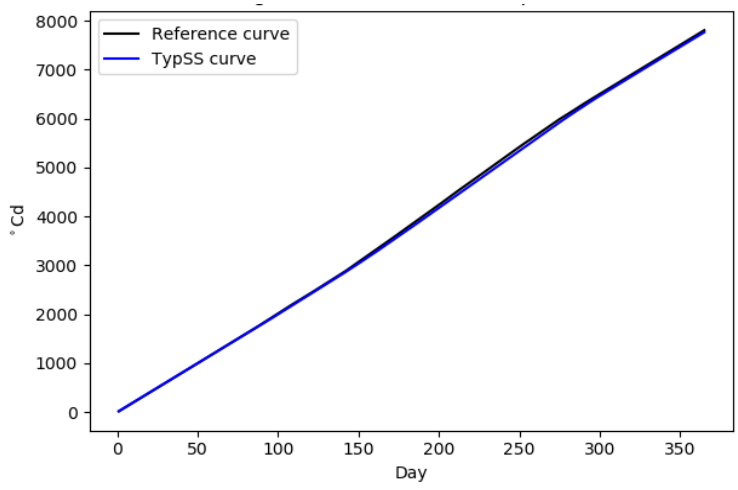
In addition to temporal profiles, the annual sum of the studied criteria and the cumulative profiles are another way of representing the temporal values and very helpful to allow direct reading and comparing of the system performance through the year until reaching the final annual value. Therefore, it is important for the predicted curves to reflect the annual reference ones. Figure III-5 shows the cumulative profiles of the backup energy (Figure III-5(a)), the energy stored in the tank (Figure III-5b)) and the internal room temperature (Figure III-5(c)) as obtained by both, the reduced sequence (in blue) and the reference case (in black).



(a)



(b)



(c)

Figure III-5. Annual and extrapolated cumulative profiles of the target criteria: (a) integrated backup energy, (b) integrated energy stored in the tank, (c) integrated internal room temperature.

With a minor deviation between the annual and the reduced sequence curves, the figure shows that the curves are with high correlation with the reference ones as previously shown in temporal profiles. The coefficients of determination (eq. 3-1) of the backup energy R^2_{elec}, energy stored in the tank $R^2_{Q\ stored}$ and internal room temperature R^2_{Tint} are 0.97, 0.99 and 0.99 respectively.

$$R^2_{crit} = 1 - \frac{\sum_{n=1}^{365}(crit_{red,indv} - crit_{year,indv})^2}{\sum_{n=1}^{365}(crit_{year,indv} - \overline{crit}_{year})^2} \quad (3\text{-}1)$$

Where

n_{period} number of periods

$crit_{red,indv}$ daily value of the criterion obtained by the short sequence

$crit_{year,indv}$ daily value of the criterion obtained by the reference annual sequence

$\overline{crit}_{year}$ mean value of the criterion in the reference annual sequence

In addition to the coefficient of determination, the statistical measure that allows quantifying this comparison is the Coefficient of Variation of Root-Mean Squared Error or CVRMSE. While the Root-Mean Square Error RMSE (eq.3-2) indicates the absolute fit of the model and shows how close the predicted values are to the actual data points giving an objective representation of the predictive accuracy of the model; CVRMSE (eq.3-3) takes this metric one-step further, by normalizing it by the average dependent variable value.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=1}^{365}(crit_{red,indv} - crit_{year,indv})^2}{365}} \quad (3\text{-}2)$$

Where

$crit_{red,indv_n}$ daily value of the criterion obtained by the short sequence

$crit_{year,indv_n}$ daily value of the criterion obtained by the reference annual sequence

$$\mathrm{CV(RMSE)} = \frac{1}{\overline{crit}_{year,indv}} \times RMSE \quad (3\text{-}3)$$

Where

$\overline{crit}_{year,indv}$ mean value of the criterion *crit* in the reference annual sequence

RMSE Root-Mean Square Error

According to ASHRAE Guideline 14 [92], a CVRMSE of and below 25% indicates a good model fit with acceptable predictive capabilities. Table III-2 shows the CVRMSE (considering daily values of the cumulative profiles) with values recorded between 1.1 and 5.8%. In addition to that, the relative annual sum error E_{crit} (eq. 2-14) is of 0.1% for the backup energy, 0.4% for energy stored in the tank and 0.5% for the internal room temperature.

Criteria	Reference annual sum	Predicted annual sum	Error	CVRMSE
Backup energy	3017 kWh	3020 kWh	0.1%	5.8%
Energy stored in the tank	19011 kWh	19095 kWh	0.4%	1.8%
Internal room temperature	7804 °Cd	7763 °Cd	0.5%	1.1%

Table III-2. Comparison between reference and predicted annual sum of the target criteria.

III.2.1.4. Comparison with other approaches

In order to assess the value of the new developed method; the obtained results were compared to two approaches used by researchers and applied on the same case study. The first one is the clustering algorithm by k-medoids [82]. In this case, the algorithm is simply applied on the results of the annual/reference results. No simulations are run during the search of the typical days so it does not take into consideration the influence of the selected days on the model dynamics during the target. Literature showed that it is a practical approach but results will show the importance of using simulations during the target process. The second approach is the iterative approach SCSPT that was developed to reduce the time consumed during the dynamic simulation of detailed solar combisystem models [79]. The two approaches work in the following manner:

Clustering Algorithm, K-medoids

K-medoids clustering algorithm is an exclusive algorithm which lies under the partitional clustering approaches. It divides data segments into a pre-determined number of clusters in which the elements of a cluster are unique and therefore not shared by other clusters. The difference between K-means, presented in Chapter 1, and K-medoids clustering is that the former assigns as a group center the exact mean of the group, which might not be an existing element, while the latter searches for an existing element closest to the mean and assigns it as the group center, as shown in Figure III-6. This method of clustering is more realistic in the case of searching for a real typical representative day.

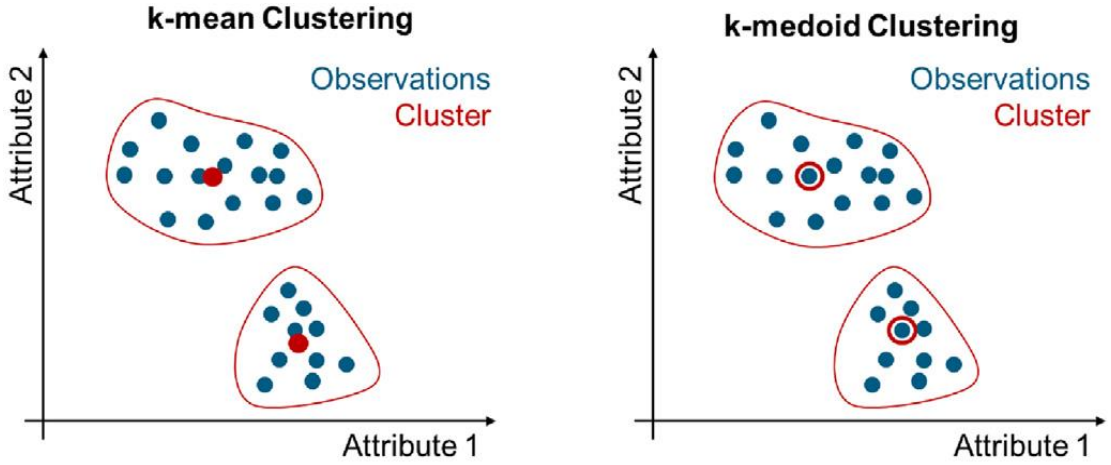


Figure III-6. Principle of partitional clustering, Kotzur et al.[82]

Short Cycle System Performance Test (SCSPT)

In order to build a 12 days short sequence, SCSPT uses monthly climate data in addition to several monthly performance criteria as attributes for the calculation of "Target" criteria by empirical equations. The three target criteria are:

- A "Target Ambient Temperature" T'_{amb} calculated from the monthly ambient temperature, the monthly space heating energy consumption for the heating season and the monthly internal temperature for the cooling season.
- A "Target Irradiation Sum" G'_{coll} calculated from the monthly total solar irradiation, the monthly energy stored in the tank, the electrical backup energy needed for the heating season and the energy need for cooling in the cooling season.

- A “Target Horizontal Irradiation Sum” G'_{hor} calculated from the monthly total horizontal irradiation and the monthly internal room temperature.

The algorithm then starts from a random initial 12 days sequence and searches the typical days that would have the closest weather data characteristics to those three “Target” weather criteria by calculating a global error ΔE_{SCSPT} and limiting it to a threshold μ_{SCSPT}.

These steps are repeated iteratively for each month until constructing the 12 days sequence. In a connected research work, Sayegh et al. [93] evaluated the performance of the approach and its efficiency upon generalization. The approach showed some limitations regarding this issue due to its functionality that cannot be applied except on a single individual at once; therefore, the obtained sequence is not adapted to great number of parametric modifications. In addition to that, it has limitations regarding its flexibility to different case studies since it depends on empirical equations and final sequence is highly dependent on the initial, randomly selected, sequence.

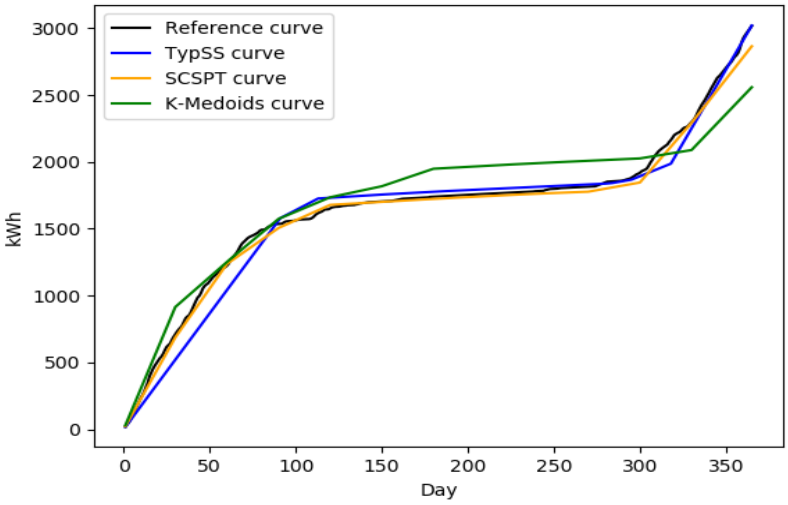
The three sequence reduction methods are applied on the same solar combisystem model and therefore three sequences of 12 days have been generated based on the same target criteria used in the previous part and coefficient of determination, annual error and CVRMSE. Those criteria were directly used by the TypSS and clustering approaches and indirectly (through the empirical equations) by the SCSPT approach. The time consumed by each method to find its own sequence varies significantly between them. While the clustering algorithm being the fastest with couple of minutes, it took about 2.5hrs for the SCSPT method and 3hrs for the TypSS algorithm to converge to their final sequences due to the repetitive simulations of the model with the test sequences. Simulating the model on the final sequences of each method gives the results in Figure III-7 and detailed in Table III-3.

Criterion		K-Medoids	SCSPT	TypSS
Backup Electrical Energy	*R^2*	0.88	0.99	0.97
	Annual Sum Error (%)	15	4.5	0.1
	CVRMSE (%)	10	3	5.8
Energy Stored in Tank	*R^2*	0.98	0.98	0.99
	Annual Sum Error (%)	4.5	4	0.4
	CVRMSE (%)	6	8.2	1.8
Room Temperature	*R^2*	0.99	0.99	0.99
	Annual Sum Error (%)	0.6	2.1	0.5
	CVRMSE (%)	3	2.4	1.1
Calculation time		seconds	2.5 hours	3 hours

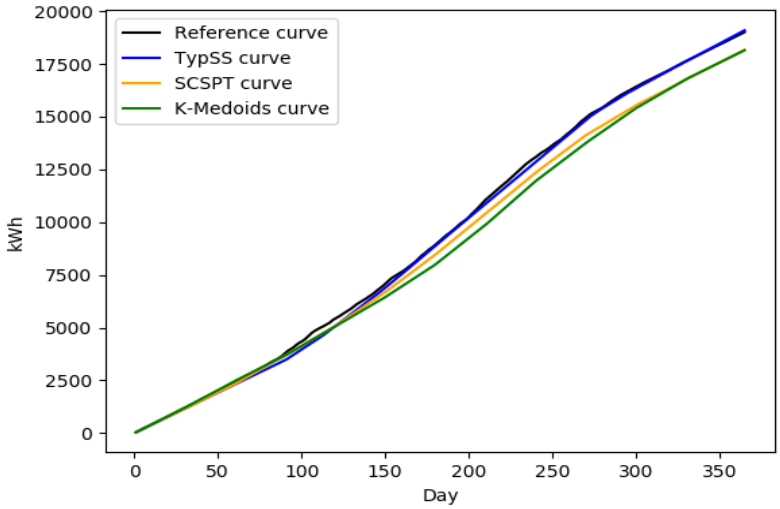
Table III-3. Comparison between the three time reduction methods results.

Figure III-7 shows that the cumulative profiles were globally better generated using the developed TypSS approach in comparison with the other two approaches. In the case of backup energy curve (Figure III-7(a)), the TypSS curve (in blue) followed with a good correlation the reference one (in black), recording an R^2 of 0.97. At the end of the winter period, where the electrical need decreases, the curve overestimates the electrical consumption. However, this overestimation is then corrected at the beginning of the heating season, leading to a final value close to the reference one (0.1% difference as shown in Table III-3). The SCSPT curve (in orange) showed the best cumulative profile of this criteria almost replicating the reference curve. However, the curve deviates at the end posing an error of 4.5% on annual sum estimation. This proximity in covering the variations of the relative curve is due to the size of the periods. The winter season is represented by only one day in the TypSS case (while three in the SCSPT case)

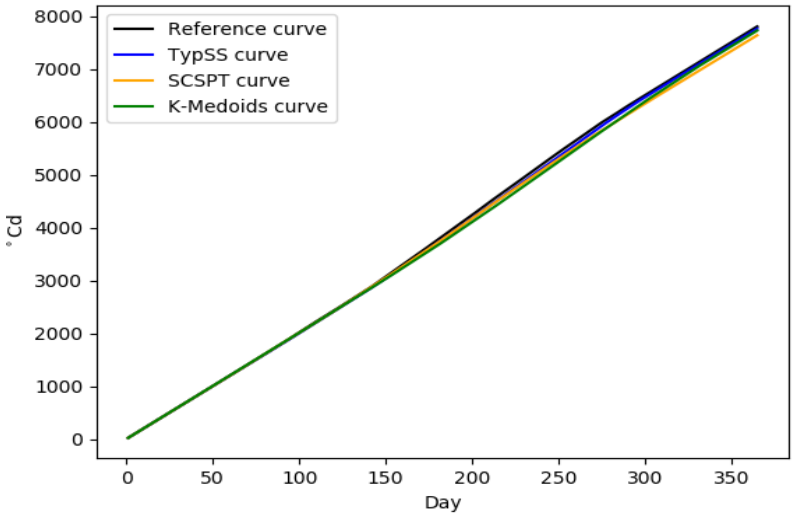
which makes this period very sensitive to estimation error. However, those errors are then compensated since more days are used to describe inter-seasonal periods. Finally, the clustering curve (in green) was the least performant where it not only failed attaining the annual sum with a 15% error, but the curve evolution did not reflect well the reference one recording an R^2 of 0.88. This is due to the fact that the clustering approach does not take into consideration the simulation process and the effect of representative days on each other when constructing the short sequence unlike the other two approaches. On the other hand, the evolution of the curves was better for the energy stored in the tank (Figure III-7(b)) and the internal room temperature (Figure III-7(c)) with the TypSS method performing the best regarding annual sum errors while the worst performance was recorded by the basic clustering algorithm.



(a)



(b)



(c)

Figure III-7. Annual and extrapolated cumulative profiles as obtained by the three methods: (a) backup energy, (b) energy stored in the tank, (c) internal room temperature.

III.2.2. Multiple tested individuals

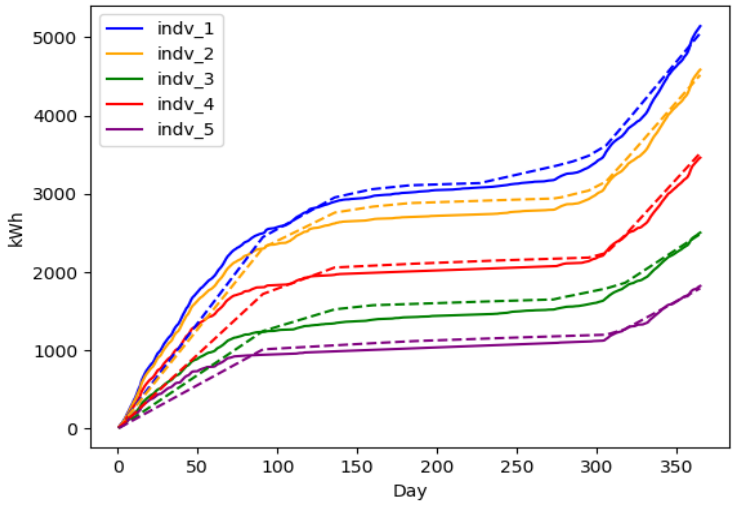
III.2.2.1. Simulation results

The five individuals (Table II-1) selected by clustering based on their performance from the 50 data base individuals (listed in Appendix B) are run on TypSS simultaneously and a single sequence of 12 days was generated. The obtained sequence is presented in Table III-4, which appeared to be different from the sequence obtained with a single individual and presented in Table III-1. This is expected since the initial conditions have now changed taking into consideration the data of five individuals at the same time.

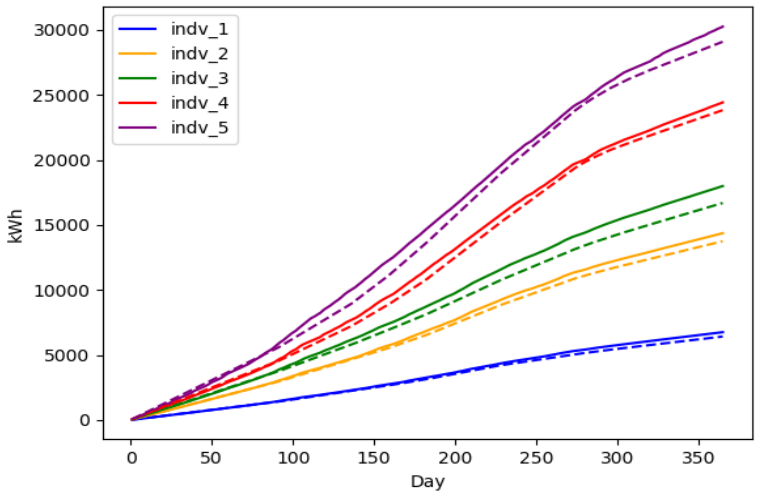
Period	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Number of days	91	22	23	23	23	45	46	11	11	11	12	46
Selected Day	**6th Feb**	**13th April**	**4th May**	**20th May**	**20th June**	**6th July**	**17th Sep**	**2nd Oct**	**20th Oct**	**1st Nov**	**7th Nov**	**20th Dec**

Table III-4.Typical short sequence of 12 days and the number of days in each period obtained on five individuals.

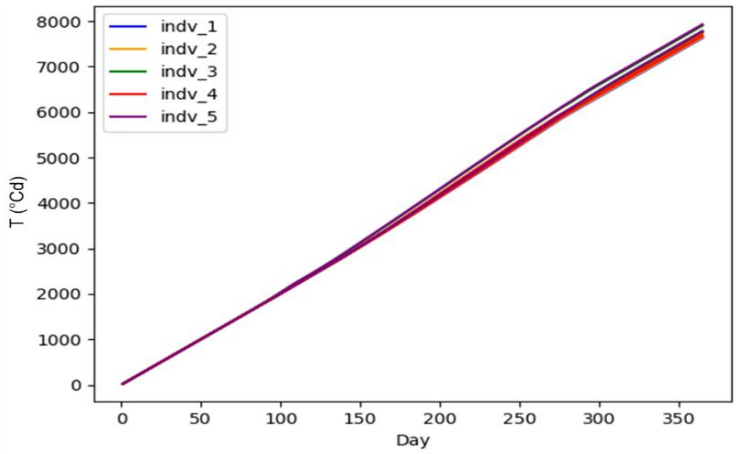
Figure III-8 shows the cumulative profiles of the target criteria as obtained by the reduced sequence (dashed line) and the reference case (solid line). Each individual is given a unique color for a better visualization of the obtained results. In addition to that, the coefficients of determination and CVRMSE of each curve are recorded in Table III-5. The figure shows that despite using a single short sequence of 12 days for simulating five different individuals of the model, each having unique parametric characteristics and different performances, the generated curves follow in a very good correlation the reference annual ones, for all individuals and all target criteria. Regarding the backup energy Figure III-8(a)), the sequence generated almost identical curves for individuals 1 and 2, and very close curves for individuals 3, 4 and 5 with minor deviations.



(a)



(b)



(c)

Figure III-8. Annual (solid) and extrapolated cumulative (dashed) profiles as obtained by the five individuals: (a) backup energy, (b) energy stored in the tank, (c) internal room temperature.

This appears in Table III-5, which shows the recorded R^2 for this criteria ranging between 0.92 and 0.97 and CVRMSE inferior to 15.2%. In addition to that, the curves generated for the other target criteria, energy stored in the tank (Figure III-8(b)) and internal room temperature (Figure III-8(c)), were also in high correlation to the reference ones. The former showing R^2 ranging between 0.98 and 0.99 and CVRMSE inferior to 10.3% and the latter R^2 equal to 0.99 and CVRMSE inferior to 3.5%. Leading therefore for a global R^2 of 0.98 for all target criteria of all individuals.

Criteria	**Backup energy**		**Energy stored in the Tank**		**Internal room temperature**	
Individuals	R^2	CVRMSE %	R^2	CVRMSE %	R^2	CVRMSE %
1	0.97	10.2	0.99	6.4	0.99	2.5
2	0.96	8.4	0.99	4.2	0.99	2.4
3	0.91	15.2	0.98	6.8	0.99	3.3
4	0.94	13.3	0.99	6.1	0.99	2.7
5	0.92	11.4	0.99	10.3	0.99	3.5
Global R^2	**0.98**					

Table III-5. The global and individual coefficient of determination of the three target criteria.

Moreover, Table III-6 shows that the annual values of the target criteria were estimated with a good precision of a relative error not exceeding 2% for the backup energy and internal room temperature, and 8% for the energy stored in the tank between the reference values (AN) and the predicted ones (TS).

Criteria	**Backup energy**			**Energy stored in the tank**			**Internal Room Temperature**		
Individuals	AN (kWh)	TS (kWh)	Error (%)	AN (kWh)	TS (kWh)	Error (%)	AN (°Cd)	TS (°Cd)	Error (%)
1	5142	5052	1.7	6748	6414	4.9	7676	7633	1
2	4586	4518	1.5	14361	13746	4.3	7697	7648	1
3	2504	2486	1	17988	16685	7.2	7903	7770	1.6
4	3461	3520	1.7	24423	23824	2.4	7745	7676	1
5	1822	1793	1.6	30251	29085	3.9	7931	7785	1.8

Table III-6. The reference (AN) and predicted (TS) annual values and their relative errors of the target criteria per individual.

To assess its generalization quality, the sequence based on five individuals is used for the simulation of the 50 individuals they are selected from. The obtained results appear in Figure III-9 (in blue) in comparison with single individual based sequence (in orange). The curves show that while the relative errors did not exceed 10% for all target criteria in the case of five individuals based sequence (in blue), the backup energy was badly estimated for the most of 50 individuals in the case of a single individual based sequence recording up to 45% error. On the other hand, while the influence of number of individuals did not seem to be noticeable in the case of energy stored in the tank, the performance of a single individual was slightly better than that of five in the case of internal room temperature with a minor difference up to 1%. However, in the case of temperature, the performance value usually taken into consideration is the temporal estimation within the limits of ±2°C when considering temperature in Celsius. In the following sections, the annual sum error is still considered even for the temperature criterion to keep consistency of the study.

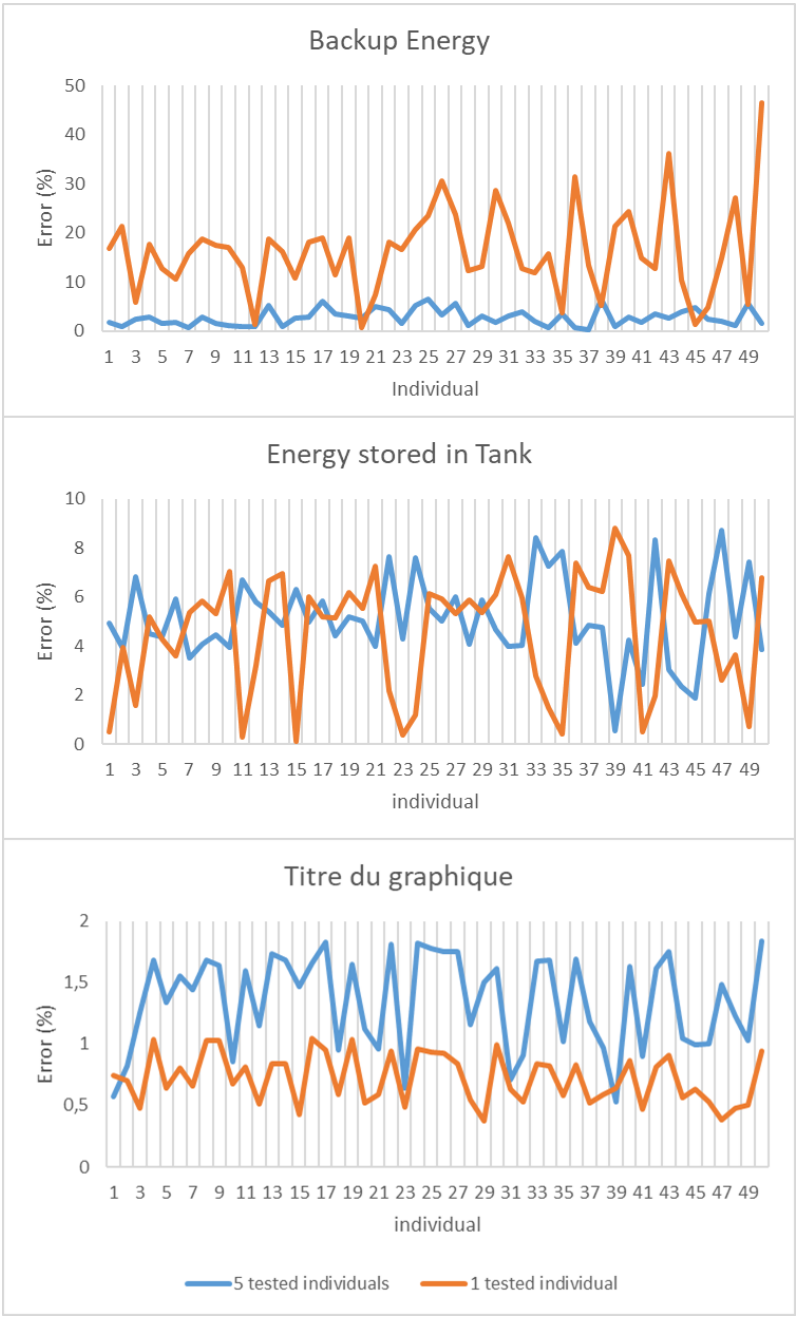


Figure III-9. Annual sum errors of the target criteria of all 50 individuals obtained after simulation with the typical day sequences obtained with one individual (orange) and five individuals (blue).

III.3. Sensitivity of the TYPSS algorithm to its main parameters

In the following section, sensitivity of the algorithm to its main parameters has been evaluated. Each parameter is evaluated by its own with no crossing between them to evaluate their influence separately. The same study is executed on several parametric

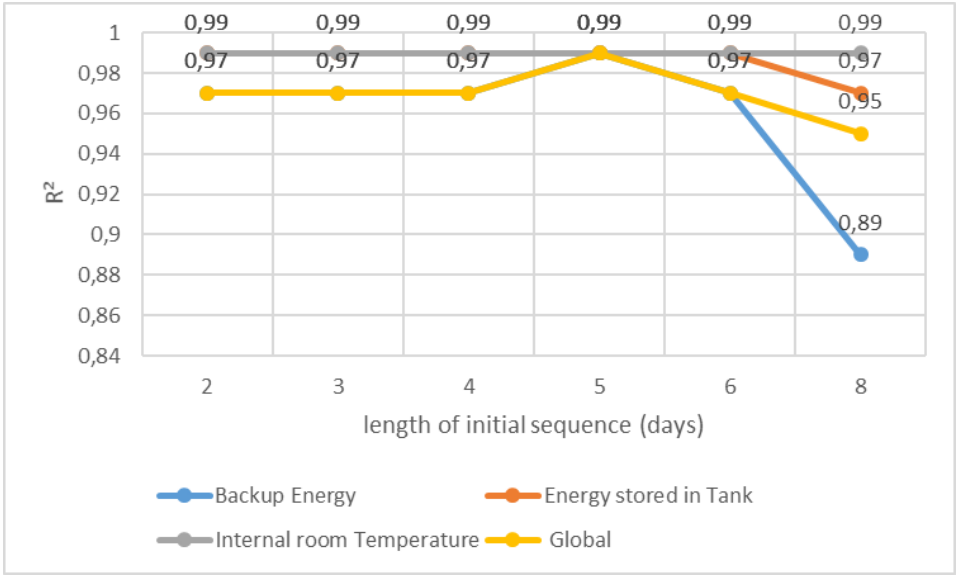
inputs. The algorithm requires several inputs from the user to start its day selection process and therefore there is an uncertainty on each input and questions regarding their influence on the final results. The main parameters that initialize the algorithm and can be modified by the user are four:

- Length of the initial sequence or number of days in initial sequence
- Length of the generated sequence or number of days in final sequence
- Number of tested individuals
- Number and type of the target criteria

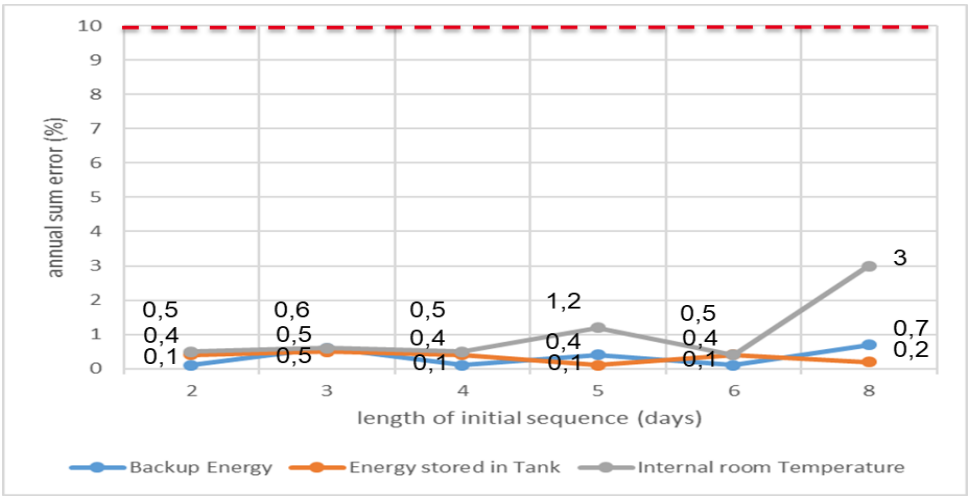
These parameters will influence the quality of the output results where they are directly dependent on them. It was shown in the literature that generating a longer sequence does not necessarily implement better prediction [82]. Rather, it is related to the case study and the evaluated criteria. In addition to that, the initial sequence was found to be influential on the final output as in the case of SCSPT iterative approach. Therefore, these inputs should be evaluated to give the user certainty about his choice of initial data. Moreover, the way the algorithm works suggests that giving more data will help in having more generalized results. This was shown in the previous part in Figure III-9 where giving more individuals would help in generating a generalized sequence applicable on further studies. However, giving a lot of data will make it difficult for the algorithm to find a single representative day for a huge set of distant points therefore it is important to know until what extent adding more individuals is helpful or starts to affect negatively the obtained results. In addition to that, TypSS algorithm is a simulation based algorithm, therefore, adding more individuals will require additional simulations for each test sequence through the course of the algorithm and therefore causing more computation time by the algorithm to converge. This is also applicable on the case of target criteria where trying to estimate many criteria at the same time will diverge the focus of the algorithm and might lead to less accurate results and therefore it is important to know the limit of the algorithm regarding this aspect. As a result to all those remarks, a sensitivity analysis is essential and has been applied and presented in this section to evaluate the stability and consistency of the algorithm and its results.

III.3.1. Length of the initial sequence

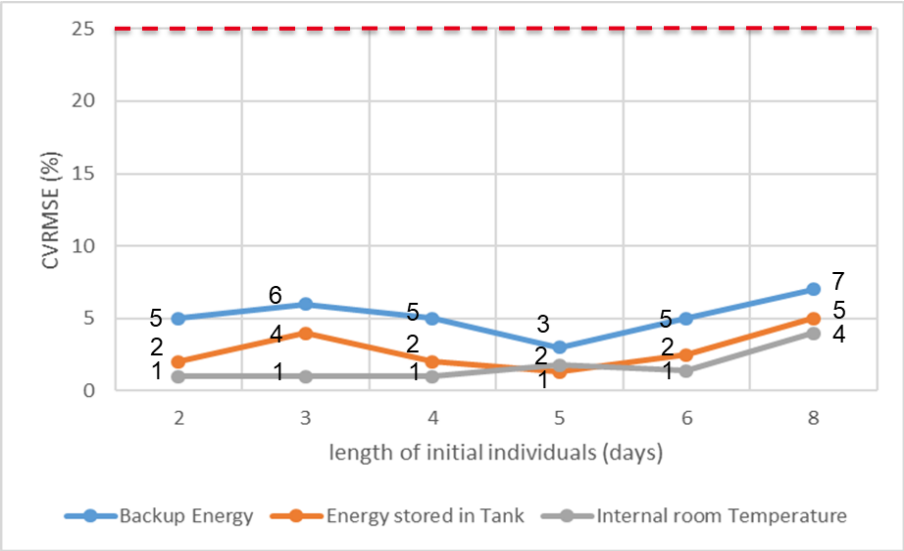
Considering the same solar combisystem model, the influence of the initial sequence has been evaluated. Previously, the initial sequence was formed of four periods representing the four quarters of the year and roughly the four seasons. Five new sequences have been generated starting from 2, 3, 5, 6 and 8 initial periods (numbers are chosen randomly) and ending with a 12 days sequence. The global coefficient of determination, annual sum error and the CVRMSE were examined and plotted in Figure III-10.



(a)



(b)



(c)

Figure III-10. Performances of generated sequences by TypSS starting from different initial sequences regarding each target criterion: (a) coefficient of determination, (b) annual sum error, (c) CVRMSE.

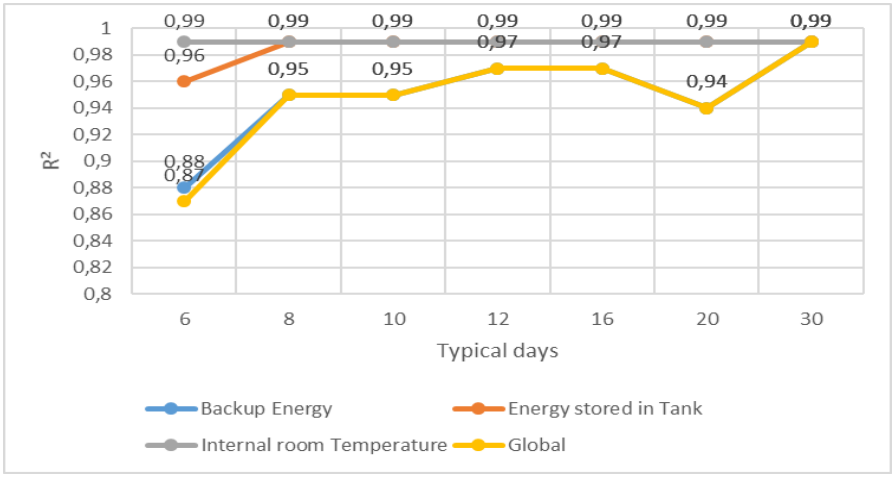
The figures showed that there was no remarkable influence due to the modification of this input for a sequence of 12 days. The global coefficient of determination (the product of the criteria coefficients of determination) stayed high despite the modifications with a global R^2 ranging between 0.95 and 0.99 and criteria values almost stable except for when considering eight initial days. The annual sum error stayed almost stable with errors not exceeding 1.2% while the CVRMSE remained inferior to 5% except for the case of eight initial periods yet still recording values much smaller than the acceptable limits (10% for annual sum error and 25% for the CVRMSE denoted in dashed red in the figure).

In addition to that, it was noticed that starting with a two days initial sequence terminated with the same sequence of the 4 days sequence. Therefore, the algorithm divided the two periods into four equal ones and then continued its course as if it started by four equal initial periods and it is therefore recommended not to start with a very short initial sequence since it will be cut equally in the first iterations by the algorithm. In addition to that, the values recorded by the eight initial periods sequence showed that it is not favorable to divide initially the year into many small equal parts if the final number of days is relatively small. In this study, starting by 8 initial days left only four more steps for the algorithm to reach the 12 days sequence. Therefore, it is better to

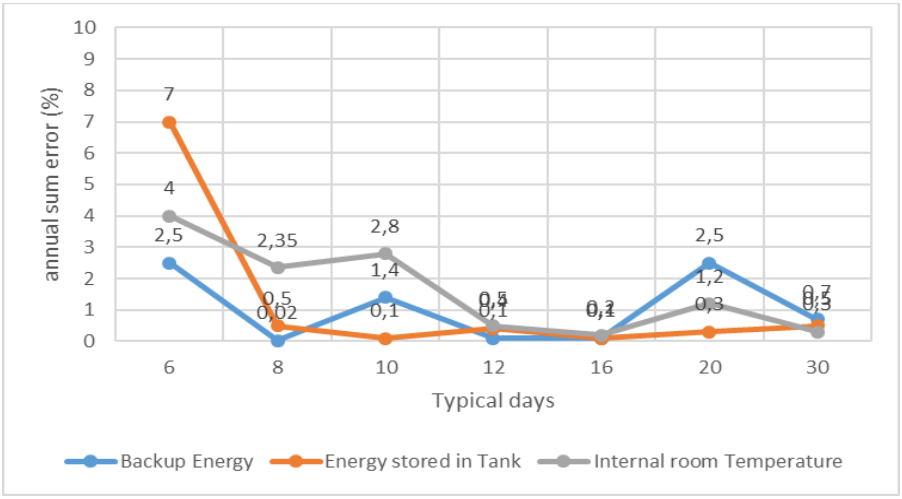
leave the *Period setting* phase in the algorithm do the breaking down of the sequence based on the performance of the simulation. In this case, an initial sequence of 4,5 or even 6 days seem adequate to avoid unnecessary iterations by the algorithm and still leave space for this phase in the algorithm to perform.

III.3.2. Length of the generated sequence

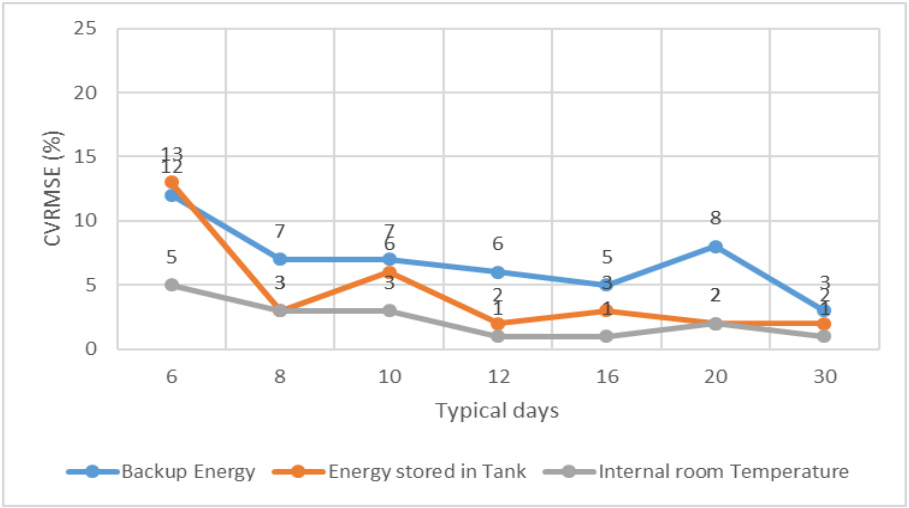
Starting from an initial sequence of four typical days representing the four quarters of the year, seven sequences were generated by TypSS applied on the same case study as before with three considered target criteria. The sequences are of different lengths ranging between 6 and 30 typical days (numbers were chosen randomly). The performances of those sequences regarding the global performance values, i.e. global R^2, annual sum errors and CVRMSE, of each of the target criteria have been analyzed and traced in Figure III-11. Curves show that achieving good results is still applicable even with very short sequences. However, increasing the number of days will indeed help in achieving better performances. Regarding the backup energy, global coefficient of determination R^2 increased up to 0.99 with a 30 days sequence while being 0.88 with a 6 days sequence. This was accompanied by an oscillating recording of the annual sum error ranging between 0.02% and 2.5% and CVRMSE decreasing from 12% to 3%. On the other hand, the performance of the two other criteria showed an almost stable recording for the global R^2 around 0.99 and a decreasing annual sum error and CVRMSE as number of typical days increase. In addition to that, the performance of a 20 days sequence was noticeable where it didn't follow the trend of the curves giving less accurate performances than shorter sequences. This supports the idea in the literature indicating that longer sequences do not forcely mean better performances. Rather, the choice of sequence length is directly related to the case study and the initial conditions. The temporal, periodic and cumulative profiles are shown in Appendices C, D and E respectively. They help in better understanding for the recorded values where for instance the high value recorded by the 20 days sequence in the beginning of the autumn season shows the reason the annual values were over estimated therefore affecting the recorded results.



(a)



(b)



(c)

Figure III-11. Performances of generated sequences by TypSS of different sizes regarding each target criterion: (a) coefficient of determination, (b) annual sum error, (c) CVRMSE.

Finally, the time recorded by TypSS for each case is recorded and plotted in Figure III-12. Computational time increases as the number of days in a target sequence is increased reaching 6 hours for a sequence of 30 days while it was 3 hours for the sequence of 12 days which was sufficient to predict all performances of the model as presented in section III.2.1. Therefore, the length of the generated sequence is a parameter that should be considered by the operator since it has an influence on both the quality of the obtained results and the computational time spent by TypSS to converge. For the following sections, sequences of 12 days will be considered since it proved it is efficient for this case study.

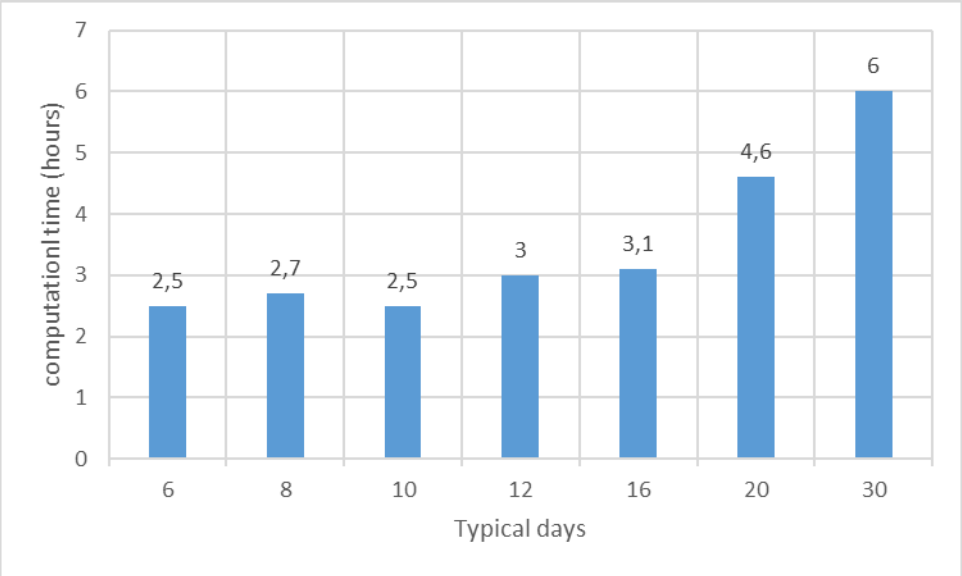


Figure III-12. Time recorded by the algorithm to converge to its final sequences of different sizes.

III.3.3. Number of tested individuals

In this section, the influence of the number of tested individuals is evaluated. Sequences of 12 days have been generated starting from a four days initial sequence and applied on different number of individuals. Sequences were formed on 1, 3, 5, 7 and 10 individuals (numbers were chosen randomly) and the prediction of the three previous target criteria was examined by calculating the annual sum error, global coefficient of determination and CVRMSE. In the same way adopted in section II.2.2. , individuals were selected by clustering based on their performances from the 50 individuals data base generated by LHS. The data base was divided into groups, clusters, in which one of each is selected to obtain different performing individuals. For

example, to obtain 3 individuals, the 50 individuals were classified into 3 groups by clustering and the center of each cluster is selected obtaining 3 centers to be considered as 3 individuals.

Before examining the outputs, the algorithm computational time was recorded and plotted in Figure 3-13. The recorded time includes both the time spent while calculating the annual reference data and the time taken by the algorithm through its course to converge to the final sequence. The data shows that the time increases proportionally as the number of individuals increase. Therefore, it is essential to consider a reasonable number of individuals to avoid a high computational time expense.

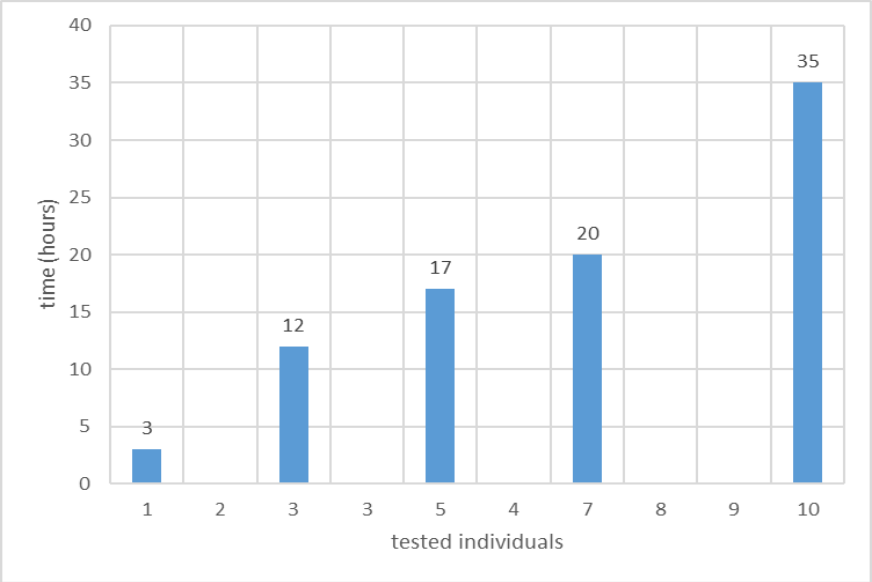


Figure III-13. Time recorded by the algorithm to converge to its final 12 days sequenc regarding different number of tested individuals.

On the other hand and as mentioned previously, the main aim of considering multiple individuals is generating a consistent sequence still applicable upon generalization. Therefore, the influence of number of individuals was not only evaluated on the used individuals but also on the 50 individuals presented in Figure II-5. Figure III-14 shows the global coefficient of determination as recorded after applying the generated short sequence on the used individuals and the 50 individuals.

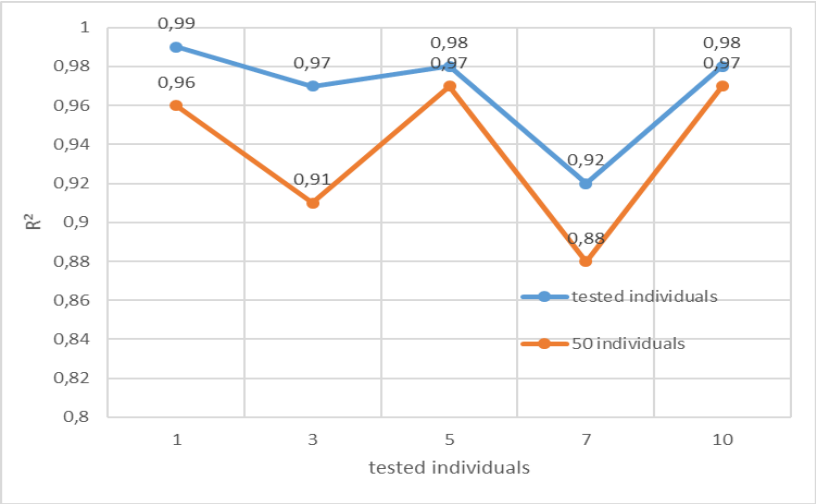
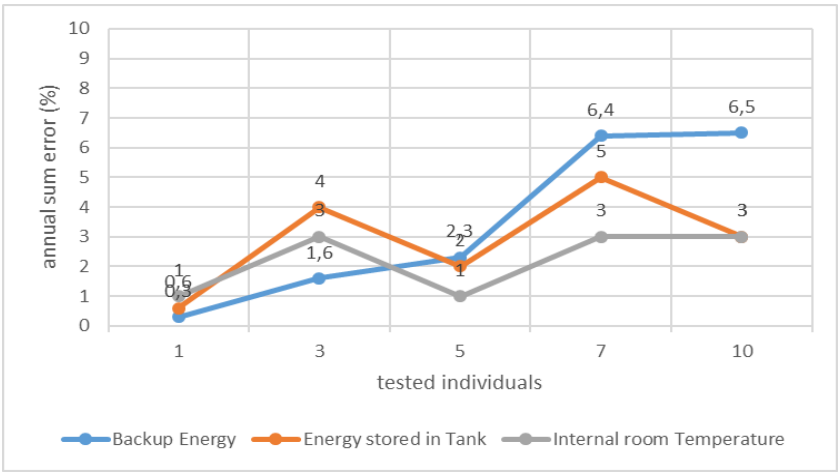


Figure III-14. Global coefficient of determination recorded applying the generated sequences on their corresponding individuals (blue) and original 50 individuals (orange).

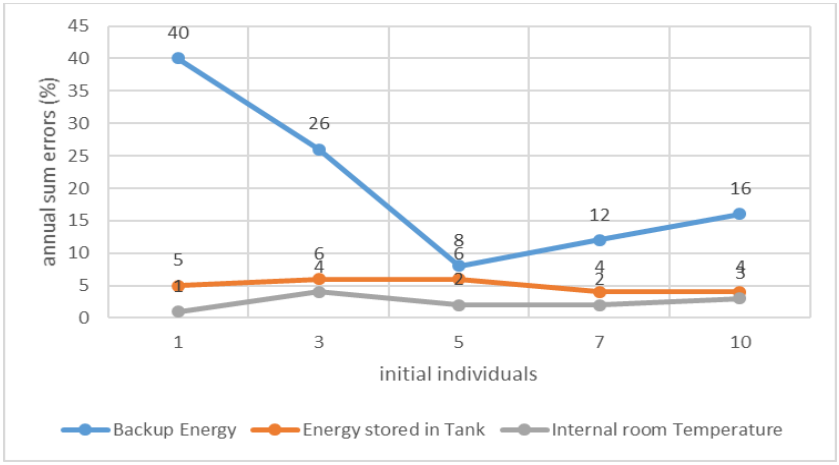
The figure shows that adding more individuals to the algorithm did not affect negatively the obtained results. Considering only the individuals of the algorithm, it was expected that adding more individuals will make the results less accurate since the algorithm now works on a wider data range at the same time. On the other hand, the coefficient of determination of the fifty individuals was expected to be better as the number of individuals increase. This is because more data are now taken into consideration and therefore the generated sequence has covered a wider range of instances that include the performances of those individuals thanks to LHS method of sampling that selects samples from the entire specified space. On the other hand, the figure shows that the results were consistent with high coefficient of determinations ranging between 0.88 and 0.99 unlike the expected trend. However, this could be reasoned to the way TypSS works on in which it tries to improve this parameter as much as possible to propose finally a sequence with a high coefficient of determination.

For further analysis, the annual sum errors of the target criteria were calculated for each sequence. The results are plotted in Figure III-15. Due to better visualizing concerns, not all individuals values are plotted. Rather, only the maximum recorded errors were taken since it covers the performance of the whole set of individuals. The aim is to have all individuals falling under the acceptable limits so examining the individuals showing the maximum errors will cover the rest. Figure III-15(a) shows the maximum annual sum errors recorded for each target criteria applying the generated

sequence on their corresponding individuals while Figure III-15(b) on the original 50 individuals. The trend of the curves followed an expected trajectory. In the case of applying the sequences on their corresponding individuals (Figure III-15(a)), the annual sum error increases as more individuals are included. This is expected since more data are now taken into consideration and therefore clusters are bigger and less compact. This leads to a cluster center not fully capable of representing all data points. Even though the errors were increasing as the number of individuals increase, they stayed inferior to the 10% limit. The backup energy showed the highest increase from 0.3% to 6.5% while the other two criteria did not exceed the 5% error.



(a)



(b)

Figure III-15. Maximum annual sum errors recorded for each target criterion applying the generated sequences on: (a) their corresponding individuals and (b) the original 50 individuals.

On the other hand, adding more individuals improved the quality of generalization results as shown in Figure III-15(b). While the energy stored in the tank and the internal room temperature showed a consistent behavior, the backup energy decreased from recording a maximum of 40% error in the case of a single individual to 8% with five individuals. However, the maximum error went back increasing with adding more individuals reaching up to 16% with ten individuals. This reflects the idea proposed previously that adding more individuals would increase the possibility of obtaining distant performing instances per period and therefore would make less compact periods leading to less performing representing days. It is therefore essential to be reasonable with the number of individuals. Adding individuals will help the algorithm in considering more performances for the selected days. However considering too many individuals would diverge the focus of the algorithm. Finally, observations are similar regarding the CVRMSE and results are present in Appendix F.

III.3.4. Number and type of the target criteria

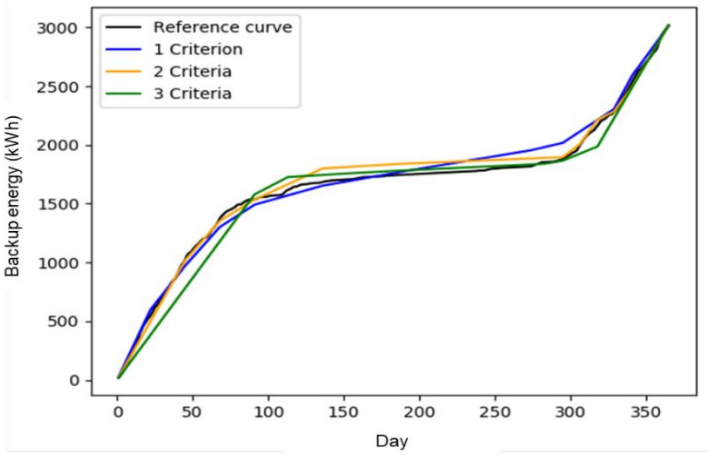
The aim of TypSS is to regenerate fastly specific performances that are under interest by the user. The algorithm was developed in a way that it can take whatever number or type of a performance, denoted target criterion, as long as it has an annual profile. However, giving many criteria to regenerate will diverge the focus of the algorithm and therefore affect the quality of the results. In such case, the evaluation variables of all criteria are taken into consideration simultaneously and denoted global values. Taking random criteria might thus affect the performance of the algorithm since, for instance, it was noticed in the previous part that the *Period setting* phase focuses more on periods that whiteness great temporal modifications. Therefore, there should be a sort of harmony between the selected target criteria or else there will be a specific focus on a specific part of the year. The choice of the target criteria depends highly on the aim of the study and the interest of the user. If a later optimization will be performed for a specific performance of the model, it is recommended to take into consideration only this criterion for TypSS if possible or with what might help in presenting over time system phenomena such as thermal inertia. This will give the algorithm the ability to focus entirely on this aspect and therefore assure better presentation. To verify that, three different sequences were generated and compared on the individual I_1 of section III.2.1. . The first is based only on the backup energy as the target criterion (for a following interest in optimizing this criterion as will be presented in Chapter IV). The

second is based on the backup energy and the energy stored in the tank. Finally the third is the one presented previously and includes the three initial target criteria, energy stored in the tank, backup energy and internal room temperature as shown in Table III-7.

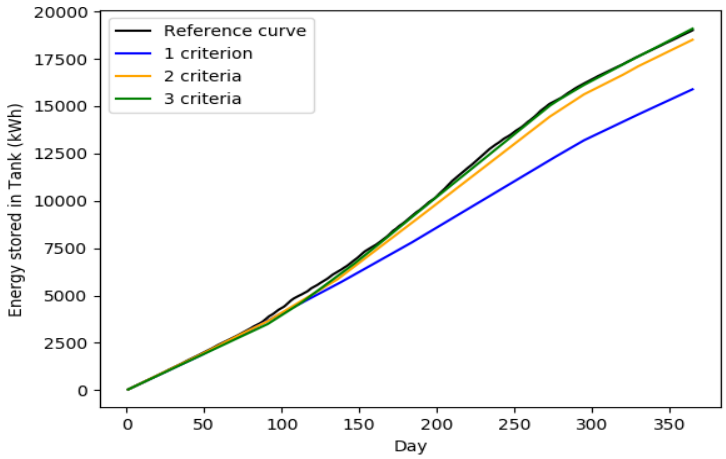
1 Criterion	Backup energy		
2 Criteria	Backup energy	Energy stored in the tank	
3 Criteria	Backup energy	Energy stored in the tank	Internal room temperature

Table III-7. Considered criteria in each case.

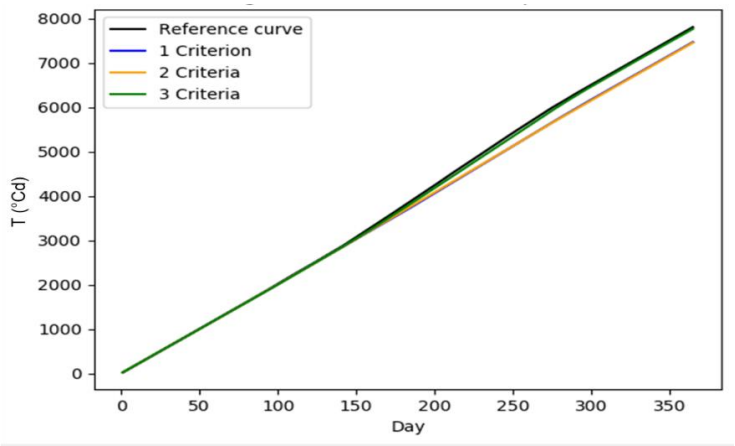
The results are traced in Figure III-16 and detailed in Table III-8. The aim is to evaluate the influence of such input variations.



(a)



(b)



(c)

Figure III-16. Cumulative profiles as obtained with different criteria combination: (a) backup energy, (b) energy stored in the tank, (c) internal room temperature.

Case	Backup energy			Energy stored in the Tank			Internal room temperature		
	R^2	Error %	CVRMSE %	R^2	Error %	CVRMSE %	R^2	Error %	CVRMSE %
1 Criterion	**0.99**	**0.1**	**4**	*0.83*	*16*	*20*	*0.99*	*4*	*5*
2 Criteria	**0.98**	**0.1**	**4**	**0.99**	**2.6**	**4.6**	*0.99*	*4.5*	*6*
3 Criteria	**0.97**	**0.1**	**5.8**	**0.99**	**0.4**	**1.8**	**0.99**	**0.5**	**1.1**

Table III-8. Results obtained with different criteria combination: considered criteria (in bold) and not considered criteria (italic).

The obtained results show the influence of the target criteria on their profiles. When the backup energy was taken alone into consideration, the cumulative profile was replicating the reference profile in the periods of electric energy use, i.e. cold periods of the year. The profile increased in an exact trend as the reference one recording an R^2 of 0.99 before it deviated slightly at the end of the summer period but this deviation was corrected in the following periods to continue exactly as the reference recording with a final annual sum error of 0.1%. Creating a longer sequence might lead to better representation, which appeared with a sequence of 30 days previously presented in Figure III-11. As more criteria were added, the R^2 decreased slightly to 0.97 while the

CVRMSE increased slightly from 4% to 5.8%. This is because the focus of the algorithm currently diverged to other aspects too. The backup energy profile is less identical to the reference one especially in the first heating part of the year because the algorithm has reduced the cuts of the year in this part and applied more period division in the summer period where the internal room temperature witnesses variations. Figures and Table also showed that not including specific criterion, as a target criterion would lead to bad predictions of them as the curves in blue showed in Figure III-16(b) and Figure III-16(c) and the orange curve in Figure III-16(c). Moreover, it was noticeable how the estimation of the energy stored in the tank improved slightly after adding the internal room temperature, which reflects the idea that having a harmony between the target criteria will influence positively the outcome of the algorithm. Table III-9 shows how each of the four initial periods of the year have been divided by the *Period setting* phase for a total of 12 periods as influenced by the criteria change.

	Initial period 1	Initial period 2	Initial period 3	Initial period 4
1 Criterion	4 sub periods	2 sub periods	No division	5 sub periods
2 Criteria	3 sub periods	2 sub periods	No division	6 sub periods
3 Criteria	No division	5 sub periods	No division	5 sub periods

Table III-9. Initial periods' division influenced by the modification of the target criteria.

Finally, the generalization aspect was examined for each case (Figure III-17). It was clear from the recorded errors that not considering a criterion in the algorithm process would not lead to an accurate prediction of it. This appeared in the energy stored in the tank and internal room temperature curves, recording best values when taken into consideration. Moreover, the influence of considering more criteria in diverging the focus of the algorithm appeared in the backup energy figure that recorded least performance when three criteria is considered. However, it was noticed the positive influence of adding the energy stored in tank to the prediction of the backup energy. This is reasoned to the role of the energy stored in the tank in representing the inertia of the system and therefore giving a closer performance to the real case. It is therefore

recommended to consider energy stored in the tank in backup energy studies even though it is not in its direct interest.

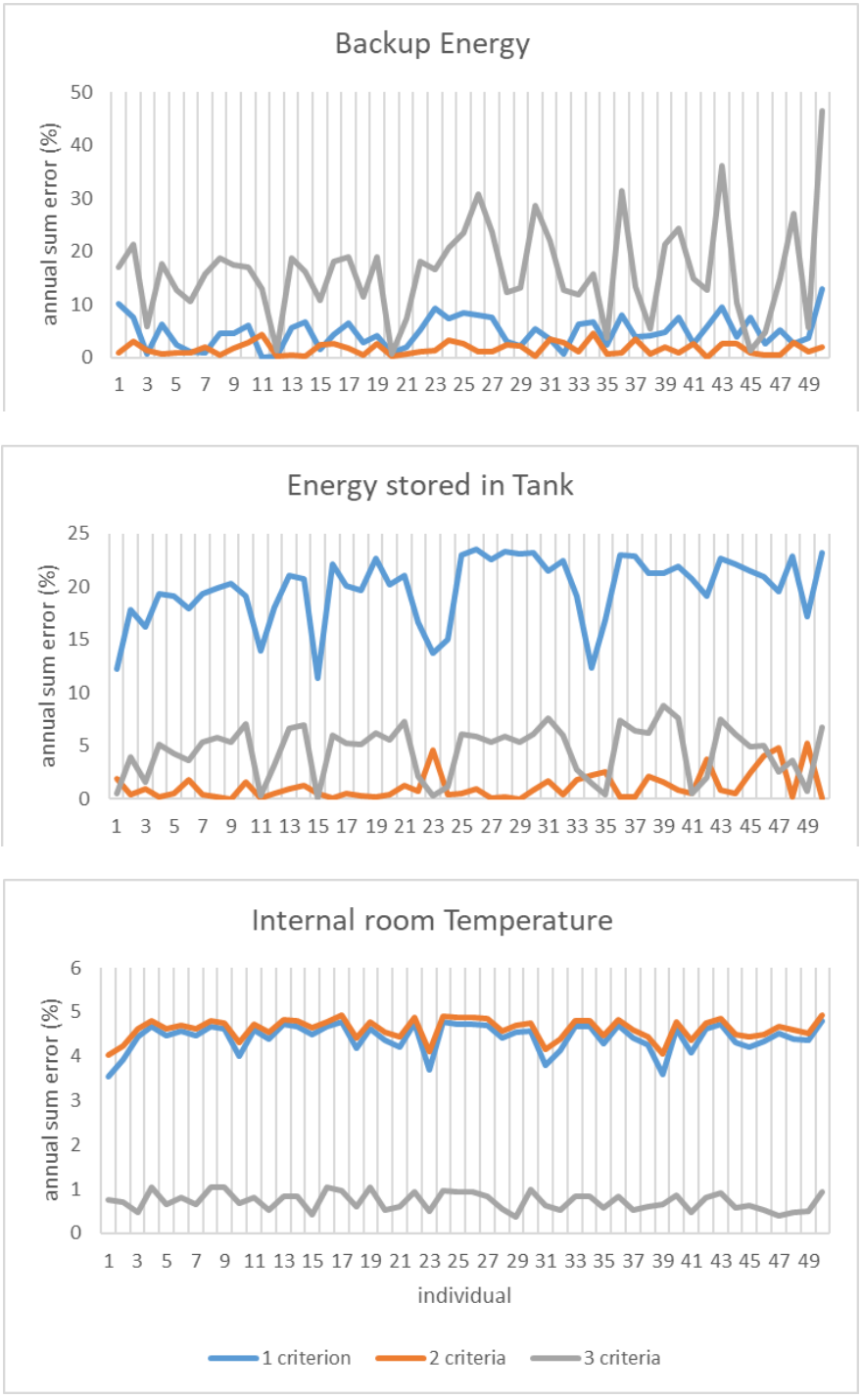


Figure III-17. Annual sum errors of the target criteria of all 50 individuals obtained after simulation with the typical day sequences obtained with one criterion (blue), two criteria (orange) and three criteria (grey).

III.4. Conclusion

Applied on a building model with a solar combisystem the simulation of a sequence of 12 days was about 25 times faster than the annual one using the same computer configuration. In addition to the saved simulation time, results show that the output of the short simulation sequence are of high correlation with the reference ones in addition to annual sum errors not exceeding 1% and daily CVRMSE inferior to 6%. Results were also compared to two approaches used by researchers on the same case study and showed best performance.

In addition to that, despite using a single short sequence of 12 days for simulating simultaneously five different individuals of solar combisystems each having a unique parametric configuration, the generated curves follow in a very good correlation the reference annual ones, for all individuals and all target criteria. Moreover, the annual sums were estimated with a high precision of a relative error not exceeding 2% for the backup energy and internal room temperature and 8% for the energy stored in the tank. The daily CVRMSE values of the target criteria were all inferior to the 25% limit specified by ASHRAE recording a maximum of 15.2% for the backup energy. The same sequence was then tested on 45 other individuals not taken into consideration by the algorithm. The curves show that the sequence succeeded in predicting the annual performances of all target criteria with relative errors not exceeding 10%. The curves also showed that generating a sequence on more than a single individual improves the quality of results and favors the aim of developing a generalized sequence applicable on a wide parametric range.

After analyzing the output of a model simulation, the sensitivity of results to four initialization inputs of the methodology were evaluated (without crossing) in the second part of the chapter to evaluate the sensitivity of the algorithm to the user inputs:

- Length of the initial sequence or number of days in initial sequence
- Length of the generated sequence or number of days in final sequence
- Number of tested individuals
- Number and type of the target criteria

Regarding the length of the initial sequence, figures showed that there was no remarkable influence due to the modification of this input as long as it leaves space for the algorithm to perform until it reaches the final sequence. The global coefficient of

determination stayed high despite the modifications with a global R^2 ranging between 0.95 and 0.99. The annual sum error stayed almost stable with errors not exceeding 1.2% while the CVRMSE remained inferior to the 25% limit. Therefore, it is recommended not to start with a very short initial sequence since it will be cut equally in the first iterations by the algorithm. In addition to that, the values recorded by the relatively long initial sequence (eight initial periods when considering 12 final periods) showed that it is not favorable to divide initially the year into many small equal parts. Rather leave it for the algorithm through the *Period setting* phase to do its breaking down based on the performance of the simulation.

On the other hand, regarding the length of the generated sequence, curves show that achieving good results is still applicable even with very short final sequences. However, increasing the number of days will indeed help in achieving better performances. This supports the idea found in the literature indicating that longer sequences do not forcely mean better performances. Rather, the choice of sequence length is directly related to the case study and the initial conditions. In the evaluated model, 12 days was adequate.

Regarding the number of tested individuals, the data shows that the time increases proportionally as the number of individuals increase. Therefore, it is essential to consider a reasonable number of individuals to avoid a high computational time expense. Figures show that adding more individuals to the algorithm did not affect negatively the obtained results. In the case of applying the sequences on their corresponding individuals, the annual sum error increases as more individuals are included. However, they stayed inferior to the 10% limit. The backup energy showed the highest increase from 0.3% to 6.5% while the other two criteria did not exceed the 5% error. When testing all 50 individuals, considering more individuals improved the quality of results. However, the maximum error went back increasing with adding more individuals reaching up to 16% with ten individuals. This reflects the idea that adding more individuals might lead to less compact groups and therefore less performing representing days. It is therefore essential to be reasonable with the number of tested individuals. Moreover, adding more individuals has led to slight increase in CVRMSE values in case of applying the sequences on their corresponding individuals while a descending behavior in case of the 50 individuals.

Finally and regarding the number and type of target criteria, the study focusing on the backup energy (as it is the main interest for a following optimization study) showed that R^2 decreased slightly to 0.97 while the CVRMSE increased slightly from 4% to 5.8% and annual sum error remained unchanged. Therefore, the number and type of the target criteria is a very crucial input variable and the algorithm is highly sensitive to this parameter. The user should be aware of the criteria he is choosing to regenerate and it is recommended to make several trials to find the best criteria combination since it is directly related to the case study and the boundary conditions of the system. In the presented study, the thermal inertia has an influence for better prediction of the system's performances, therefore, it is recommended to consider criteria that represent this phenomenon (such as energy stored in the tank) even though they might not be in the direct interest of the upcoming study.

On the other hand, the global computational time consumed using the new approach for the simulation of the tested building model is higher than using directly the annual sequence. While the latter took 19mins to execute, the global annual computational time taken by the proposed approach is 40 secs by the reduced dynamic simulation in addition to 3 hours taken by TypSS to converge by a single individual as shown in Figure III-13. Therefore, before improving the performance of TypSS, it is not an interesting measure on the scale of model simulation. However, the results obtained after simulation were accurate and the conclusions from the sensitivity evaluation were promising. Therefore, speeding up a very time consuming study, such as optimization of a detailed model, by applying the short simulation sequence is an interesting field to explore. The value of the approach, in its current version, is in its output sequence. Implementing the reduced sequence in repetitive simulation based studies will show its value in saving time with respect to using the annual time consuming simulation. The next chapter presents the performance of TypSS on a multi-objective optimization of the building model presenting the process OptiTpSS.

Chapter IV Multi-objective optimization using reduced sequences: Introducing OptiTypSS

IV.1. Introduction

IV.1.1. Objective

When it comes to optimization, algorithms run iteratively numerical models constructing sequences of progressively better solutions to a point that satisfies optimal conditions. Because of code features, the results may be non-linear and have discontinuities. The use of special optimization methods that do not require the computation of the derivatives of the function is therefore necessary [27]. For that reason, the building simulation model is usually coupled with an optimization engine, which runs algorithms, and strategies to find an optimal solution [28].

It was cleared in Chapter 1 that in building performance simulation (BPS), multi-objective optimization is more relevant than the single objective approach and there exist many research works that consider this approach while optimization. On the other hand, simulation-based optimization techniques require up to thousands of simulations to evaluate the case study and simulating detailed models is very useful for accurate and credible results. The optimization schemes may therefore become infeasible due to such computationally expensive models. In addition to that, it was explained in the same chapter that model reduction or the use of surrogate models may cause issue due to doubts regarding precision, sensitivity and even computation time in the case of surrogate model validation. The processing time of optimization studies can be severely affected by the balance between the number of variables and their options so usually computer clusters are used for complicated optimization problems with large number of variables.

Therefore, the use of short simulation sequences is another interesting measure in this case. It was shown in Chapter 3 that the sequence obtained by TypSS reduced the simulation time of a detailed building model by 25 times than a full year simulation. Which means that simplifying the model or replacing it by a surrogate one is not necessarily essential to accelerate its simulation. This conclusion can be projected to simulation-based optimization where detailed model simulations will be consecutively repeated but much faster now thanks to reduced data profiles. The obtained results, which could be used to find the optimal solution, were also accurate and validated upon generalization. Consequently, a multi-objective optimization of a detailed building model while using reduced data profiles found by TypSS is applied in this section. There can be two ways of application:

- either the sequence is defined before the optimization (sequential approach)
- or there is an adaptive plan combining identification of the short sequence and optimization (adaptive approach).

The interest of the second approach is to be more efficient through exploring only the individuals around the optimal solutions. The two approaches were tested and analyzed. In addition to that, results of the second approach were compared to a metamodel adaptive approach. Simplifying the case study with the surrogate approach rises concerns regarding the validity of the model. On the other hand, reducing the simulation sequence by TypSS rises concerns regarding the validity of the reduced sequence with respect to the predicted performances. Therefore, a reference optimization study was performed to evaluate the obtained Pareto fronts. It includes a highly time consuming annual simulation of the detailed model and the reference Pareto front was used to be compared with the predicted ones.

IV.1.2. Multi-objective optimization method

Optimization methods are numerous and can be classified into four categories according to [94]:

- *Deterministic methods* based on the derivative of the results with respect to the decision parameters.
- *Enumerative methods* that go through the entire search space.
- *Random methods* that test certain points in space at random.
- *Evolutionary (genetic)* that processes all the solutions evolved in successive stages. These processes are based on Darwinian evolution and work with the evolution of populations over generations.

In the following study, the chosen optimization method is based on genetic algorithms and more particularly on the method NSGA-II (Non-dominated Sorting Genetic Algorithm-II), an improvement made by researcher Deb of the method NSGA [34] which uses the notion of Pareto dominance [95]. The NSGA-II method consists of:

- Creating a random initial population of individuals.
- Identifying N Pareto fronts: differentiate several Pareto fronts and prioritize them. The first Pareto front will include all non-dominated solutions, the second Pareto front will include all the solutions dominated by a single other solution,

the third Pareto front will include the solutions dominated by 2 other solutions and so on...

Figure IV-1 shows an example of point classification into dominated and non-dominated in an optimization study that minimizes two performance functions. The individual denoted “Reference point” in the figure is dominated by two individuals and therefore will be classified in the third Pareto front rank. The five individuals in black were not dominated by any individual and therefore will be classified in the first Pareto front and so on.

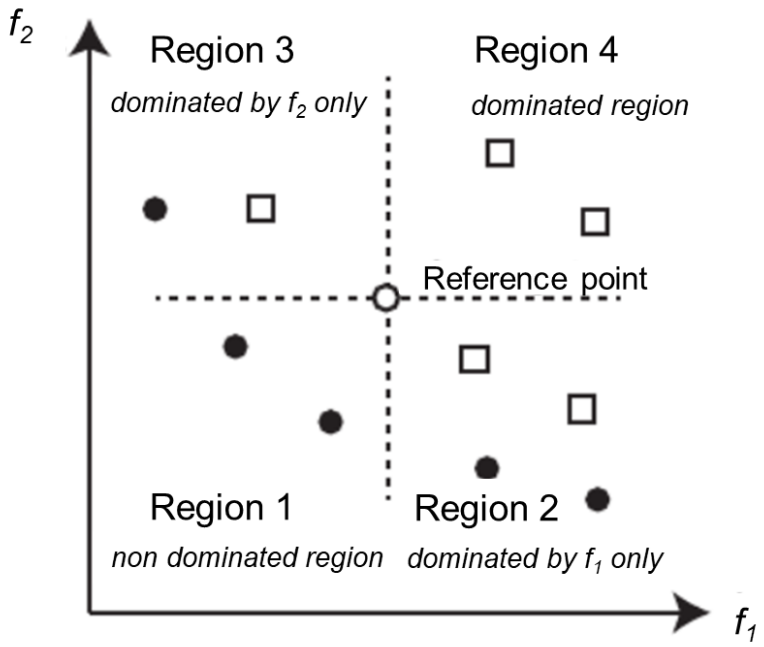


Figure IV-1. Dominated and Non-dominated regions of a reference point [96].

- Performing a “crowding” procedure: calculate for each solution the distance as a function of the perimeter of the hypercube having as vertices the points closest to this solution on each objective as shown in Figure IV-2. This distance makes it possible to rank the solutions on the same Pareto front.

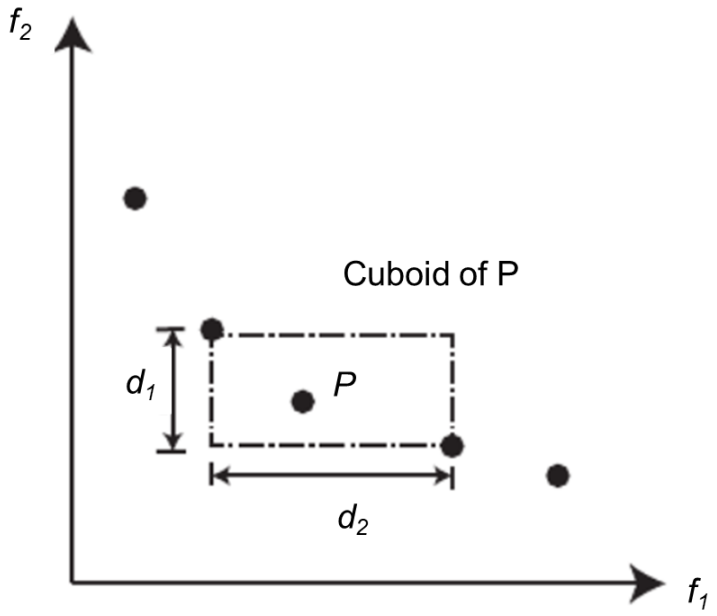


Figure IV-2. Representation of the "crowding" distance [96].

- Selecting the best solutions and create a new population (child population) based on the mutation and crossing of the best solutions from the previous population (parent population). By selecting the best solutions from each population, the population will tend towards the Pareto front over the generations.

  Selection of the best individuals takes the following process:

  - In order to select the best individuals from the parent population, the K-individuals selection tournament is used (Figure IV-3). The process consists of randomly selecting K individuals and keeping the best one, i.e. the highest ranked in the hierarchy of N-Pareto fronts and the "crowding" distance. The process is repeated until the desired number of

individual children is obtained.

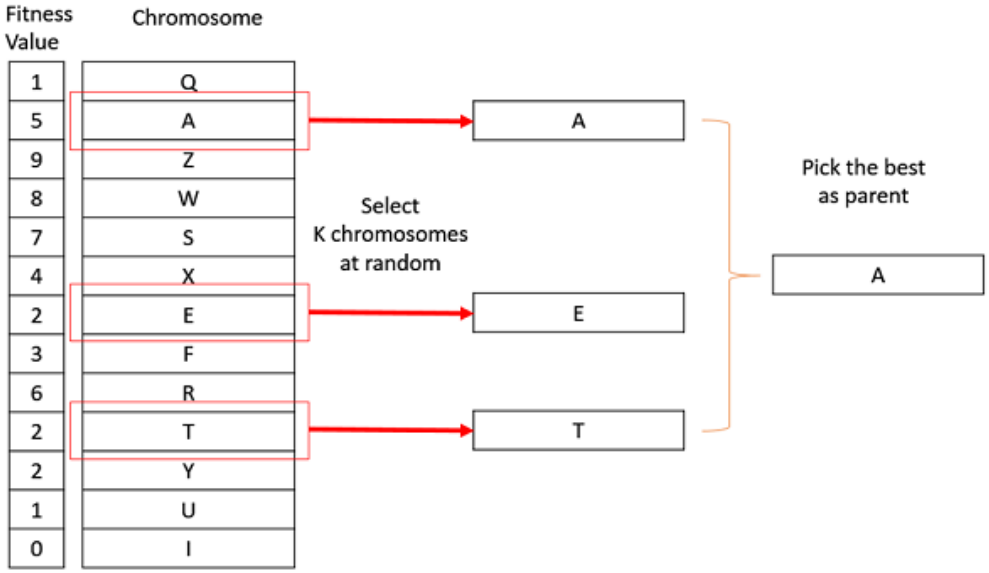


Figure IV-3. Example of a selection tournament for K = 3 and a maximization problem [97].

- Once the best parent individuals have been selected, they are randomly grouped into pairs and their decision parameters are crossed to obtain different child individuals. This crossing defines child individuals $c(i)$ from 2 parent individuals $p_1(i)$ and $p_2(i)$ by the following equation:

$$c(i) = p_1(i) + \beta.(p_2(i) - p_1(i)) \quad (4\text{-}1)$$

where

β random number in the range $[-\alpha, 1+\alpha]$

α percentage of crossing.

- Finally, to avoid obtaining child individuals who are too similar to parent individuals and risk not causing the population to change over the generations, mutations are carried out. Mutation is an operator to maintain genetic diversity, i.e. parametric diversity, from one generation to another. Several genes can be altered under mutation generating therefore different children that may change entirely the obtained solutions. For example, a parameter in the child individual, represented by a gene, can be modified under a specific mutation percentage producing a parameter not inherited from its parents, creating therefore

more diversity. Figure IV-4 presents this process with each parent parameter denoted gene G_{pi_j}, i being the parent indicator and j the gene indicator.

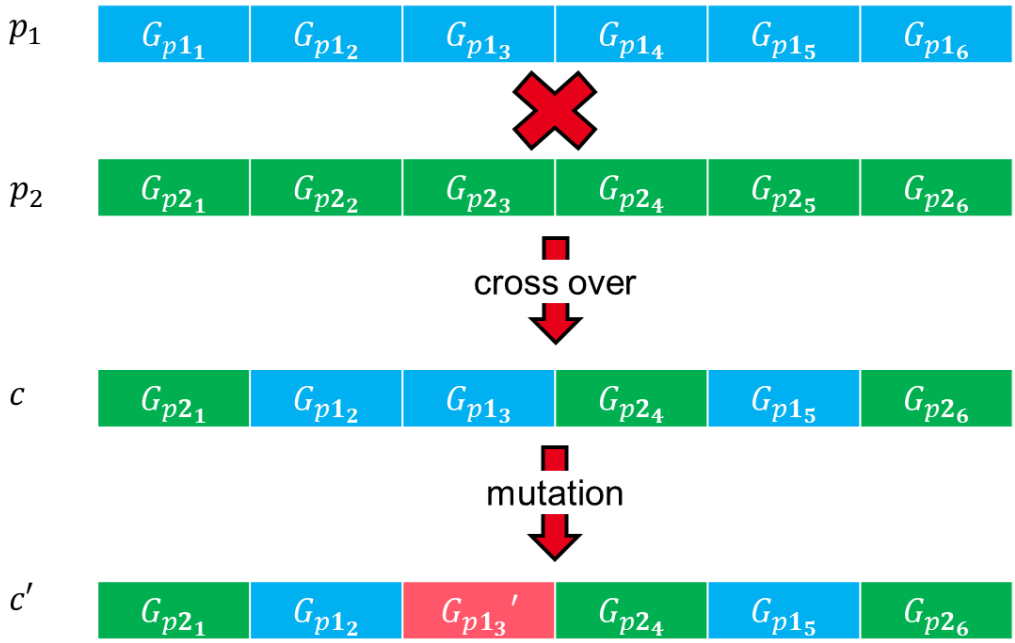


Figure IV-4. Process of cross over between two parents and the mutation of a gene in their obtained child.

The NSGA-II method proceeds by loops, each corresponding to a generation, which improve the population at each iteration. Therefore, unchanging generations can be an indicator for the end of the optimization process and therefore giving the obtained Pareto front as the final output of the optimization algorithm.

IV.1.3. Parametrizing the multi-objective optimization method

The methodology of optimization NSGA-II has been applied on the building model, including both envelop and system, presented previously in Chapter 2 and simulated in Chapter 3. The model was kept detailed and the use of reduced sequences was employed to accelerate the optimization process. The aim of the study was to optimize objective functions of the model while modifying several parameters. The modified parameters are the same presented previously:

- Surface of the solar collector in m² (SCOLL)
- Volume of the storage tank in m^3 (VST)
- Thickness of the insulation material in m (INS)

The optimization algorithm, in a range specified by the operator, modifies those parameters while searching for their optimal combination. The ranges imposed in this study were the ones used previously when generating the 50 individuals by LHS (SCOLL ranging between 6.5-25m², VST ranging between 0.3-1m³ and INS ranging between 0.04-0.3m) to keep consistency of the study. The focus of the study was to optimize:

- The annual backup energy Q_{backup} (in kWh) needed by the system to operate which was previously used as one of the target criteria for TypSS algorithm.
- The investment, material and installation, cost $cost_{Total}$ (in €) of the three modified parameters (eq. 4-2). The cost of each parameter was calculated based on the equations (eq. 4-3) to (eq. 4-7) described in [98].

$$cost_{Total} = cost_{Coll} + cost_{Vol} + cost_{Ins} \quad (4\text{-}2)$$

Where
$cost_{Coll}$ investment cost of solar collectors
$cost_{Vol}$ investment cost of storage tank
$cost_{Ins}$ investment cost of insulation material

$$cost_{Coll} = 900 \times SCOLL \quad (4\text{-}3)$$

Where
$SCOLL$ surface area of the collectors in m²

$$cost_{Vol} = 0.5 \times VST + 1000 \quad (4\text{-}4)$$

Where
VST volume of the storage tank in liters

$$cost_{Ins} = cost_{Ins_{ext\ wall}} + cost_{Ins_{roof}} \quad (4\text{-}5)$$

Where
$cost_{Ins_{ext\ wall}}$ investment cost of external wall insulation
$cost_{Ins_{roof}}$ investment cost of roof insulation

$$cost_{Ins_{ext\ wall}} = 60 + 117 \times (INS - 0.1) \quad (4\text{-}6)$$

Where
INS insulation thickness in m

$$cost_{Ins_{roof}} = 19 + 50 \times (INS - 0.3) \quad (4\text{-}7)$$

Where
INS insulation thickness in m

Those objective functions were chosen for being two opposite yet important performances of building assessment. The annual need for backup energy expresses the energy savings and environmental impact of the model. However reducing energy is of a cost that should be examined since high investment cost is inapplicable even if it was ideal for the environment. Other criteria could have been examined such as energy savings, primary energy, CO_2 or internal room temperature as in [47], [48] but only two criteria are analyzed in the following sections for simplicity.

Regarding the optimization's algorithm parameters and since the optimal α value, used in eq. 4-1, is different for each optimization problem and cannot be known without performing numerous simulations, it is set to 80%. This value is recurrent in the literature and does not influence the final result but the number of generations necessary to obtain the Pareto front [99]. In addition to that, Gaussian mutation is used. It consists in choosing a decision parameter randomly on a number of child individuals and in adding a random value to it according to a Gaussian distribution. If the new value falls outside the range of variation of the decision parameter, it remains unchanged. Mutation rate was set at 10% which is frequently used in the literature. As with the growth rate, this value does not influence the final result [99]. Finally and to ensure getting best individuals, the optimization algorithm stops when the population has not changed for 20 generations.

In order to evaluate the quality of the obtained results in the sequential approach, a reference optimization study was performed. It includes an annual simulation of the detailed model and the obtained 2D Pareto front was used to be compared with the reduced one. Figure IV-5 shows the Pareto front obtained with the complete annual simulation. It was obtained after running the algorithm on a super computer with 30 cores and took 175 hours (around 7.3 days) to converge.

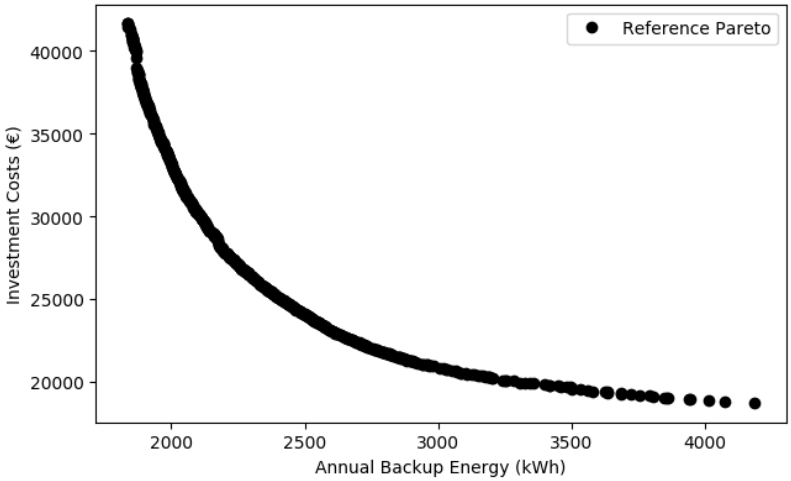


Figure IV-5. Reference Pareto front obtained with an annual simulation.

IV.2. Sequential multi-objective optimization methodology

In this approach, the reduced sequence is generated previously and applied directly in the optimization study of the building model. The 12 days sequence presented in Chapter 3 in Table III-1 and obtained from a single tested individuals was used. The same parameter ranges, as in the reference case, were defined (SCOLL, VST and INS) and the same model aspects/performances (investment cost and annual need of backup energy) were evaluated for comparison with the Pareto front obtained upon an annual simulation. Moreover, the optimization study was applied on the same computer configuration of 30 cores without any modifications of software or hardware. Figure IV-6 shows the Pareto front obtained (in blue) with respect to the reference Pareto front (black).

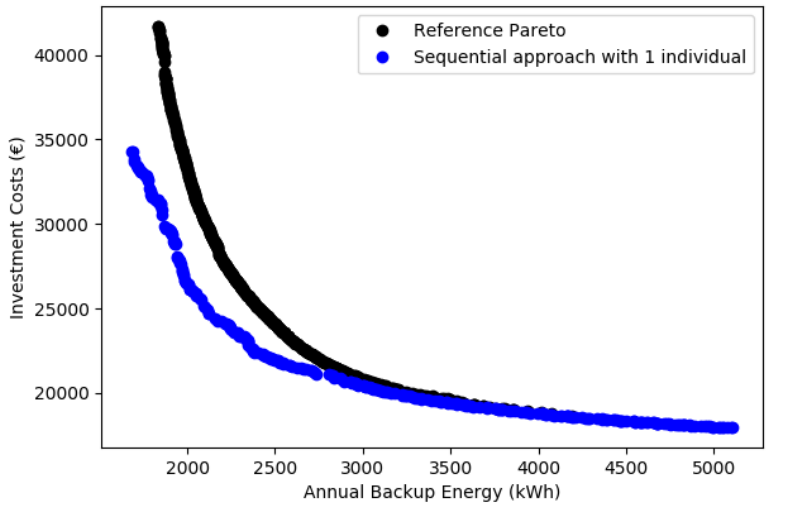


Figure IV-6. Predicted Pareto front with respect to the reference one after applying the short sequence obtained from a single individual and three target criteria.

The optimization algorithm converged much faster than the reference one. It took only 5.7 hours to converge when using the reduced sequence while it took 175 hours when using the annual data profiles. This time reduction by 30 times can be reasoned by the simulation time saved through running 12 days data instead of 365 days data in each “child” simulation (a reduction by 25 times as shown in the previous section). In addition to that, less points were obtained on the final Pareto front (600 individuals on the reference Pareto front while 281 on the reduced one) this means that less iterations were conducted before final convergence. However, and as expected, the generated Pareto front did not reflect accurately the reference one. While the predicted Pareto front was superposing the reference curve when annual backup energy is superior to 3000 kWh per year, the prediction was very bad for individuals of the Pareto front with lower backup energy consumption. This is because not enough data were taken into consideration when generating the reduced sequence on a single individual. This result reflects the conclusion obtained in Chapter 3 and plotted Figure III-9 that showed high errors when using a single individual reduced sequence in estimating the performances of a great number of individuals of various parametric configurations.

Therefore, and following this conclusion, a new Pareto front was calculated using the 12 days sequence presented in Table III-4 and obtained after testing five individuals. The algorithm converged in the same time as the previous of 5.7 hours. The obtained Pareto is presented in and compared with the previous Pareto front in Figure IV-7. The influence of using more individuals was clearly noticed when analyzing the predicted Pareto fronts. The figure shows that the Pareto front obtained with a sequence generated from five tested individuals (in red) is closer to the reference Pareto front than the one obtained from a sequence generated with a single individual (in blue). Therefore considering more individuals when generating the test sequence is very helpful for having a better predicted Pareto front.

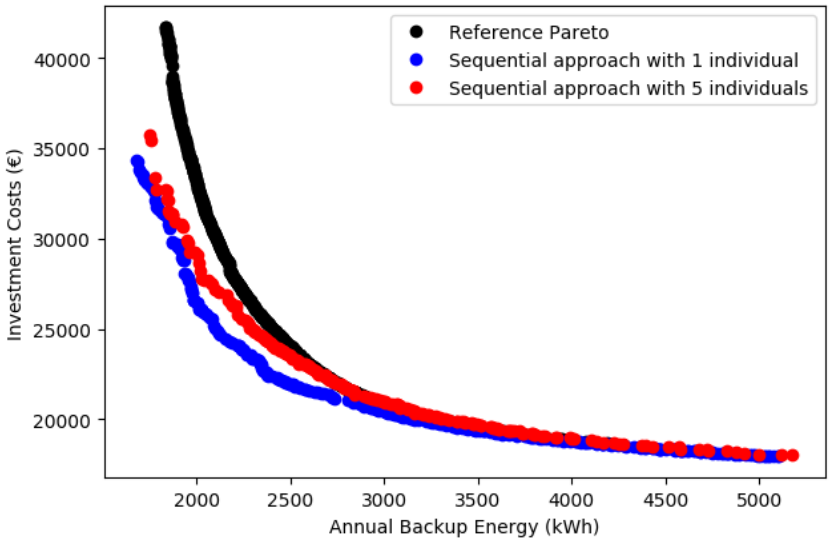


Figure IV-7. Comparison between the reference and the two predicted Pareto fronts after applying the short sequence obtained from a single (blue) and five individuals (red) considering three target criteria.

However, even though the results are better than before, the Pareto front still shows deflection for individuals with lower need for backup energy reaching up to 8% in its maximum deflection. This result is not in contrary with the previous generalization conclusion (Figure III-9) since all points passed the 10% error limit previously, which is still respected in this plot. However, the prediction needs to be enhanced. Therefore, another conclusion from the previous sensitivity analysis in Chapter 3 is used. It is related to the type and number of the used target criteria in the day selection process. It was shown previously that there should be a harmony between the used target criteria or this will affect the results. Therefore, limiting the criteria used for the selection process to only the criteria used for the following optimization study would give more focus by the algorithm on those specific performances and therefore might help in obtaining a more representative sequence. In the current optimization study, the only criterion that TypSS could use for day selection is the backup energy since the investment cost is not a model performance with an annual profile which is required to be considered by the algorithm. As a result, a reduced sequence of 12 days starting from four initial periods was generated on the five individuals by considering only the backup energy as a target criterion. The obtained sequence was used to run an optimization study and the predicted Pareto front is plotted in Figure IV-8.

As shown in the figure, the predicted Pareto front is now much closer to the reference than the previous Pareto fronts. This is in accordance with the conclusion in Chapter 3 regarding the influence of the choice of target criteria for better prediction by TypSS.

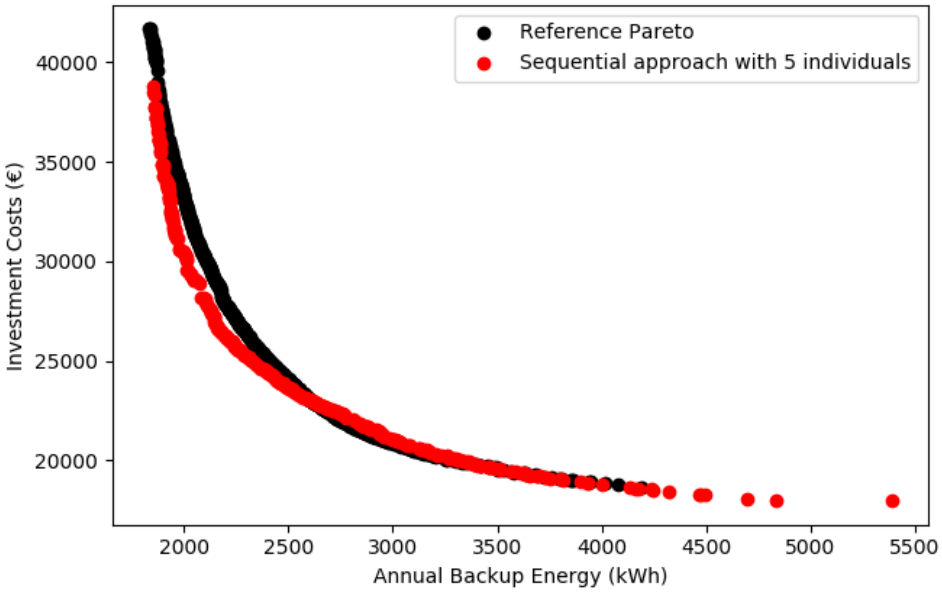


Figure IV-8. Predicted Pareto front with respect to the reference one after applying the short sequence obtained from five individuals and backup energy as the only target criterion.

However, the obtained Pareto front still deflects with an error of max 5% from the reference one, which is less than the previous cases but still not obtaining the real Pareto front. The reason behind these deflections are the tested individuals. If they are not close to an optimized set of parameters, the algorithm will dig into this false solution. So, TypSS has to bring good predictions not for a great set of individuals but the ones close to the Pareto front. Therefore, an adaptive approach, named OptiTypSS, is proposed and evaluated in the next section in an attempt to improve even more the predicted Pareto front. It involves using TypSS, and not only its output sequence, inside the optimization study. The proposed process works on benefiting from the aspects of TypSS to decrease the differences between the predicted and reference Pareto fronts, without even having the latter.

On the other hand, Figure III-17 in the previous chapter showed that considering energy stored in the tank in addition to the backup energy was of positive influence in the prediction of the latter, due to the information the former adds related to the inertia of the system. So a Pareto front considering only those two criteria while day selection

by TypSS is calculated and recorded a maximum of 5%. The curve plot is shown in Appendix G.

IV.3. Adaptive multi-objective optimization methodology (OptiTypSS)

The process of OptiTypSS is presented in Figure IV-9. The strategy searches for the optimal solutions while at the same time identifying the reduced sequence.

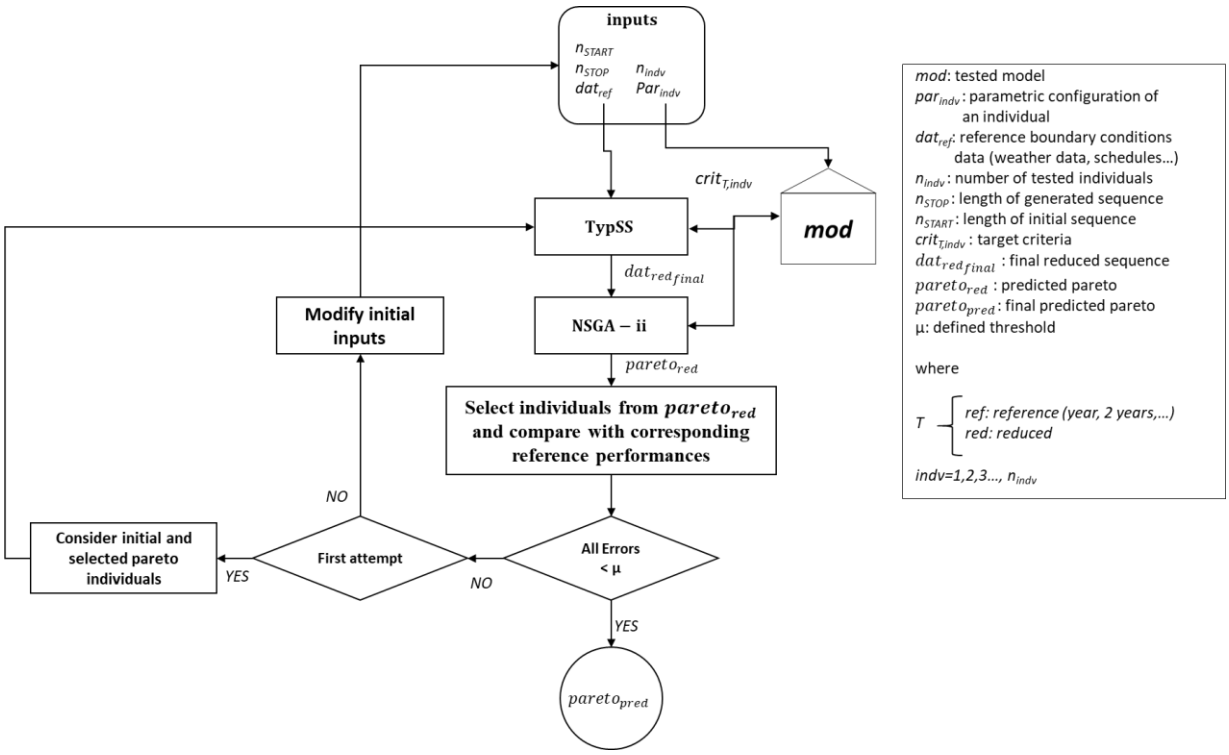


Figure IV-9. Proposed adaptive strategy to enhance the predicted Pareto front (OptiTypSS).

The idea is to check the representation of the predicted Pareto front by testing several individuals taken from different parts of the Pareto front itself and validating its proximity to reference data. To do so, the predicted Pareto front is divided into sections and a single individual is selected from each. There are different sampling processes to do so, random selection is an option but clustering approach is also convenient. The Pareto front is divided into clusters or parts and the center of the cluster (an individual) is selected. The axes of the Pareto front represent the predicted evaluated performances obtained by the reduced sequence. On the other hand, the optimization algorithm also gives as output the parameters of the individuals forming this Pareto front. Therefore, annual simulations could be run on the optimal individuals taking their

outputs as reference data. Comparing this output to the reduced value will give the idea about whether the reference performances were really predicted or not. Therefore it will show if the reference Pareto front was regenerated or not since those reference values will be present on it if the predicted Pareto front was well estimated.

If the obtained errors are very small (inferior to a defined very small threshold μ) this means that the predicted and reference Pareto fronts are very close to each other and the reduced sequence succeeded in regenerating the reference Pareto front. If not, the performance of selected individuals were badly predicted and should be improved.

Since TypSS works in its process on minimizing the differences between the final annual sums, then, including those selected individuals as individuals considered by TypSS will give the algorithm the ability to predict their performances in specific which will increase the chance in predicting the reference data. Therefore, the initial and selected individuals are now all considered by TypSS to generate a new sequence that can predict the performances of all individuals at the same time. TypSS is run again and the obtained sequence is introduced to the optimization algorithm to get a new predicted Pareto front. The obtained Pareto front is validated through the same process as before (selecting individuals, calculating reference data and comparing the results). Having minor errors means that the generated Pareto front is very close to the reference curve. However, if the individuals' errors are still superior to the threshold even after adding the new individuals, this implements that the used input for TypSS are not sufficient and need to be modified to improve the obtained sequence. Adding more individuals from the new Pareto front will have, in addition to increasing the number of heavy annual simulations, the disadvantage of considering too many individuals which was proven in the previous sensitivity analysis of being of bad influence. The proposed strategy was tested in an example.

To avoid having too many tested individuals, the process starts with three initial individuals ($n_{indv_{ini}} = 3$) and adds another three selected from the obtained predicted Pareto front ($n_{indv_{selec}} = 3$). This will give a total of six tested individuals which proved to be convenient in the previous sensitivity analysis of Chapter 3. The tested model (mod) is the same building model and a sequence of 12 days ($n_{STOP} = 12\ days$) was generate by TypSS starting from four initial periods ($n_{START} = 4\ days$). Backup energy was the only considered criterion ($crit_{T,indv}$) by TypSS for day selection as in the

previous example. NSGA-II was run on the reduced sequence and a predicted Pareto front is obtained. It was divided into three parts by clustering and a single individual was selected from each. This is shown in Figure IV-10. The initial individuals are indvs 1, 2 and 3. They appear to be scattered in the space with individuals 2 and 3 not close to the Pareto front. The different clusters of the divided Pareto front are denoted a color where a single individual is selected from each.

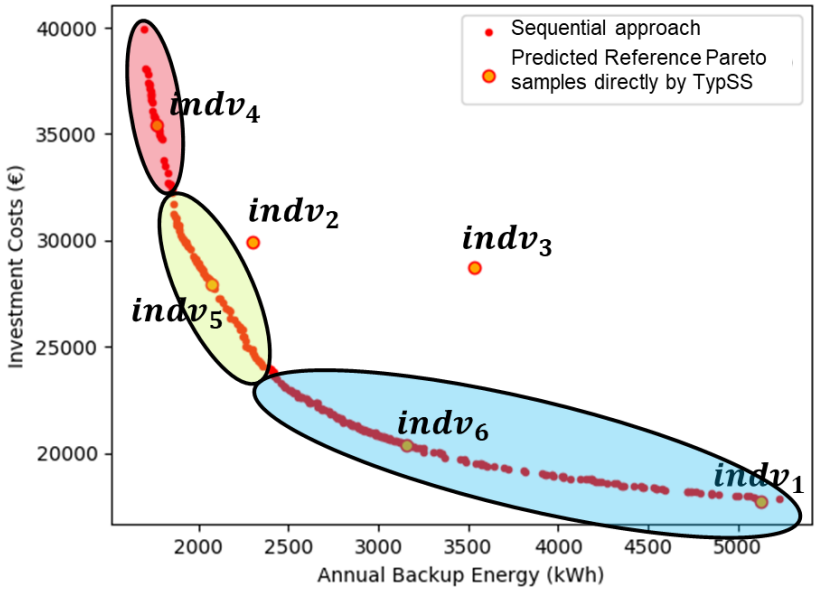


Figure IV-10. Predicted Pareto front obtained from three individuals and a single target criterion. Pareto is divided into three parts showing the initial and selected individuals.

The parametric configurations of the three selected individuals were taken from an output file of the optimization algorithm, annual simulations were applied for the new selected individuals and the obtained annual sum of backup energy were compared to the ones taken from the predicted Pareto front. Results shown in Table IV-1 indicate that the Pareto front actually deflected from the reference one at the first and second part, represented by their corresponding individuals, and it shows relatively high differences. This appears in Figure IV-11.

	Indv	Annual sum (kWh)	Predicted sum (kWh)	Relative error (%)
Initial individuals $n_{indv_{ini}}$	1	5142	5124	3.5
	2	2388	2297	3.8
	3	3461	3531	2
Selected individuals $n_{indv_{selec}}$	4	1956	1769	9.5
	5	2242	2076	7.4
	6	3098	3153	1.7

Table IV-1. Comparison between annual and predicted sums of the initial and selected individuals from the first predicted Pareto front.

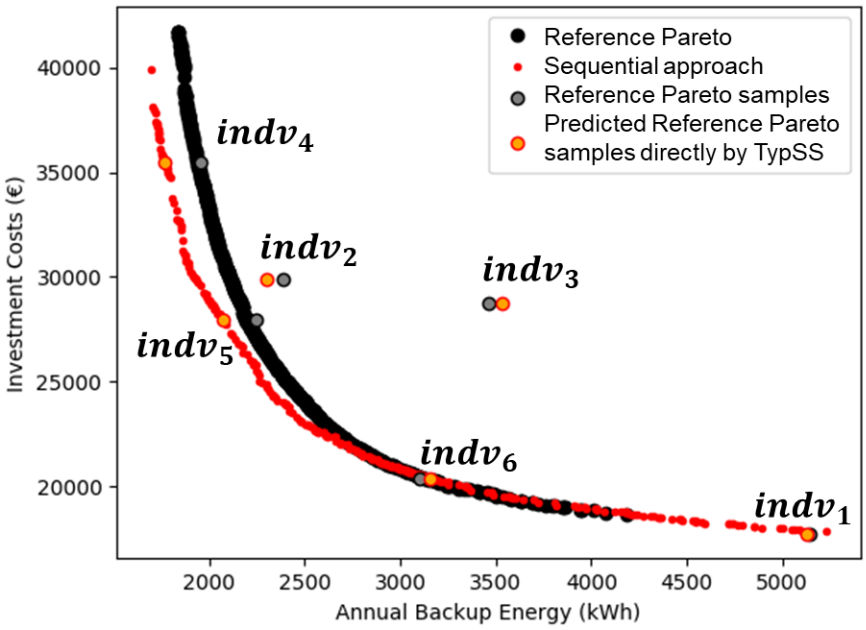


Figure IV-11. Predicted Pareto front with respect to the reference one after applying the short sequence obtained from three individuals and a single target criterion.

Following the previous results, the three selected individuals were added to the three initial ones obtaining a total of six individuals and a new reduced sequence was generated by applying TypSS on all considered individuals. The obtained sequence was introduced to NSGA-II and a predicted Pareto front was obtained. Three new individuals were selected from the new Pareto front (indvs 7, 8 and 9 in Figure IV-12)

and were validated by their corresponding annual performance values. The results are shown in Table IV-2.

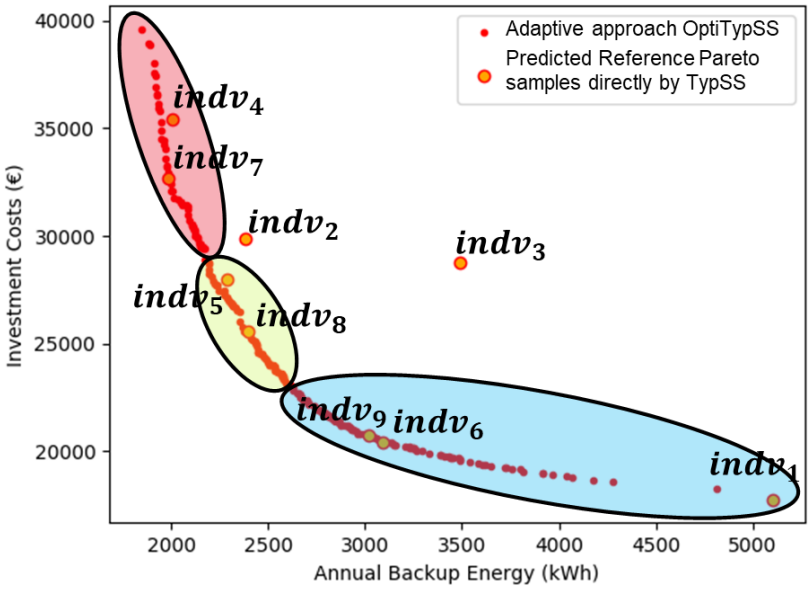


Figure IV-12. Predicted Pareto front obtained from six individuals and a single target criterion. Pareto is divided into three parts showing the initial and selected individuals.

	Indv	Annual sum (kWh)	Predicted sum (kWh)	Relative error (%)
Initial individuals $n_{indv_{ini}}$	1	5142	5106	0.7
	2	2388	2383	0.2
	3	3461	3489	0.8
	4	2242	2288	2
	5	1956	2004	2.5
	6	3098	3088	0.3
Selected individuals $n_{indv_{selec}}$	7	2016	1987	1.4
	8	2373	2395	0.9
	9	3017	3016	0.01

Table IV-2. Comparison between annual and predicted sums of the initial and selected individuals from the second predicted Pareto front.

The predicted values are highly in accordance with the reference values recording minor errors and this appears in Figure IV-13 when plotting the predicted and reference Pareto front. The two Pareto fronts are superposing each other with the predicted Pareto front almost replicating the reference one.

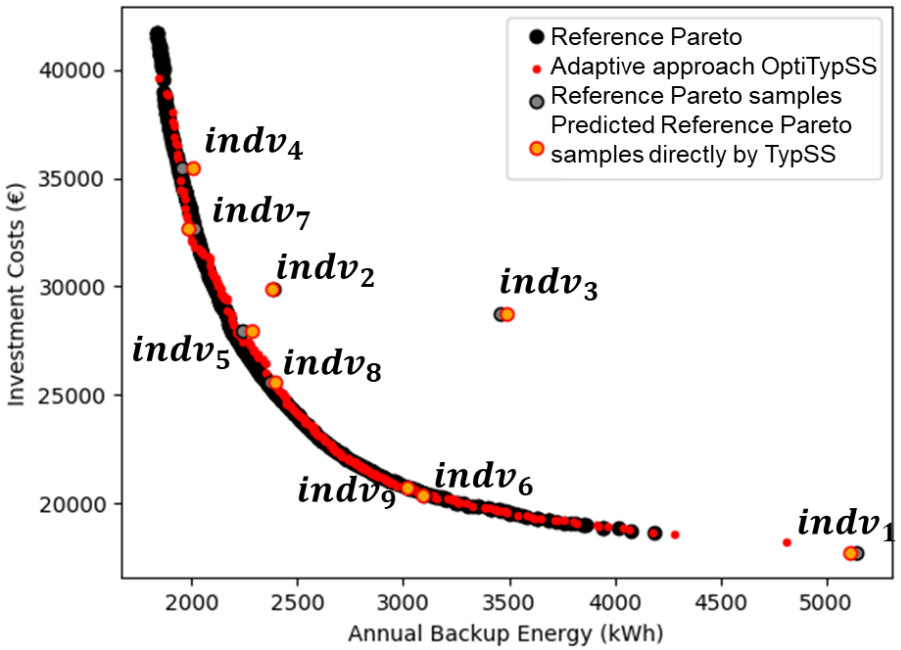


Figure IV-13. Predicted Pareto front with respect to the reference one after applying the proposed strategy.

The same test was run considering only the three individuals selected from the Pareto front and is presented in Appendix H. The prediction was almost good but not as accurate as the previous results recording a maximum of 3% deflection. This could be reasoned by giving a very small number of individuals for TypSS to explore even though they are around the optimal cases. Considering more individuals is helpful since it gives more data for TypSS to generate a well predicting sequence. On the other hand, Appendix I shows the curve obtained by considering 10 individuals (5 initial and 5 added from the predicted Pareto front). The results show a good final Pareto front that slightly deflects from the reference Pareto front with a maximum of 4%. This can be reasoned to considering too many individuals; the algorithm could not predict well the performances of low backup energy need individuals. This supports again the previous remarks and the decision taken for testing a sufficient yet limited number of individuals.

Finally, the optimal individuals' parameters of the reference and predicted Pareto fronts of Figure IV-13 are traced in Figure IV-14 to see if OptiTypSS succeeded in finding the same optimal solutions after succeeding in finding the real Pareto front. The 3D plot shows that in fact the individuals are not identical. The points are close and scattered in the same zone of the space which means that the found individuals are close to the real ones but not exactly the same.

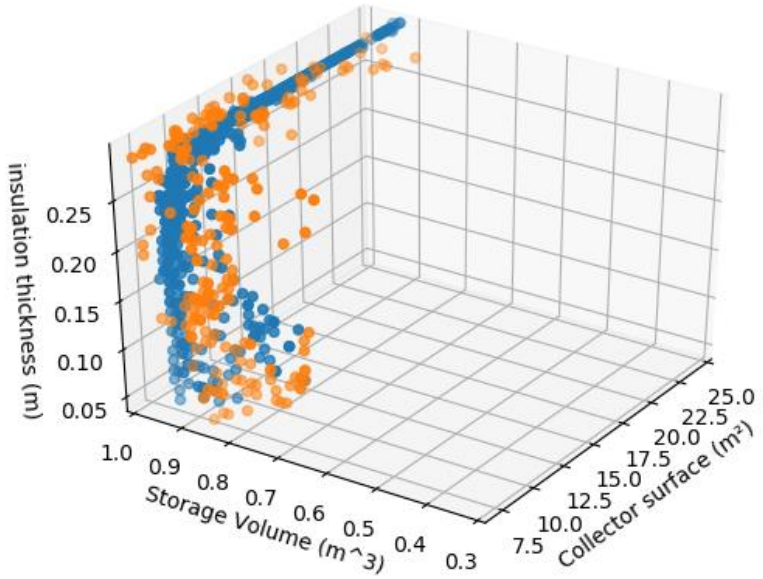


Figure IV-14. Individuals corresponding the predicted (orange) and reference (blue) Pareto fronts as found after applying the proposed strategy.

IV.4. Comparison of OptiTypSS with an adaptive metamodel based approach

In order to evaluate the value of OptiTypSS, the obtained Pareto front was compared to the result obtained from an adaptive optimization approach with metamodels, one of the most used methods in multi-objective optimization of detailed models [66]. A metamodel being an approximation model of the original simulation model that mimics the behavior of the original model to be able to produce the model responses at reduced computational cost as previously stated in Chapter 1.

Following the method developed in the INTENSE project [100], the multi-criteria optimization of the building model was carried out at the same time as the construction of the metamodel to limit computational time expenses. The method consists in adding learning points where the estimation variance is greater, taking into account only the

zone of the optimal solutions (sets of parameters forming the Pareto front). After each point addition, re-evaluation of the metamodel and the Pareto front takes place. Once the metamodel has reached a reasonable precision over all the ranges of variation of the decision parameters, the focus becomes on a small area of the reference function.

This technique allows the reduction of the exploration areas of the decision parameters which leads in reducing the computing time knowing that only the area of optimal solutions is worth being considered in an optimization study. The method is divided into three main steps:

- Step 1: Development of the metamodel from learning points obtained by LHS sampling.
- Step 2: Improvement of the metamodel by sequential addition of learning points where the estimation variance is the highest over the entire range of variation of the decision parameters with re-evaluation of the metamodel after each point addition.
- Step 3: Improvement of the metamodel by sequential addition of learning points where the estimation variance is highest on the area of the decision parameters forming the Pareto front of Step 2 with re-evaluation of the metamodel and the Pareto front after each addition point.

In the tested case study, the approach required 10 learning points in each step for a total of 30 learning points, i.e. 30 heavy annual simulations, in order to create the metamodel. The obtained Pareto front is plotted in Figure IV-15 (in orange) and shows high correlation with the reference one (in black). The Pareto front obtained by OptiTypSS is slightly less performant recording errors inferior to 1.5% (in red) while requiring only nine heavy annual simulations (six for the generation of the Pareto front and three for validation). This result implements that the new developed method is as efficient as other approaches used in the domain and is an interesting field to continue on since it gives reliable results. However, the time of convergence was much faster in the case of the metamodel than the others as shown in Table IV-3. Optimization by metamodel took only 2 hours while using 10 processors (which can't be increased due to considering 10 learning individuals in each step) while it took around 38.4 hours with OptiTypSS (includes running TypSS and applying the optimization twice). The majority of the time consumed to generate the Pareto front of the reduced sequence was mainly consumed in the TypSS algorithm itself during the generation of the short sequence

where it spent a total of 27 hours for both first (with 3 initial individuals) and second (with 6 tested individuals) calls. This time was mainly spent in the *Typical days' enhancement* phase because of the numerous iterations it does while modifying the typical days period by period. Activating the option previously explained in this phase which runs clustering inside the periods and selects a set of days to be tested instead of all days of the period would help in speeding up the phase and therefore the TypSS algorithm and OptiTypSS in global. In addition to that, while NSGA-II was run using 30 processors for parallel simulations, TypSS was applied on a single processor. Performing parallel simulations in TypSS would speed up the process. However, to be relevant, the different approaches should be run several times to generalize those time results since they could also depend on how busy the computer was when doing the study with other processes which should be taken into consideration.

		Reference	OptiTypSS		metamodel
Number of heavy simulations		600	6		30
Number of process calls	Pre-optimization	-	TypSS	2x	-
	NSGA-II	1	2x		1
Process Time (hours)	Pre-optimization	-	1st call	9	-
			2nd call	18	
	NSGA-II	147	1st call	5.7	2
			2nd call	5.7	
Total Time (hours)		147	38.4		2

Table IV-3. Time consumed to obtain the final Pareto fronts of Reference, OptiTypSS and metamodel simulations.

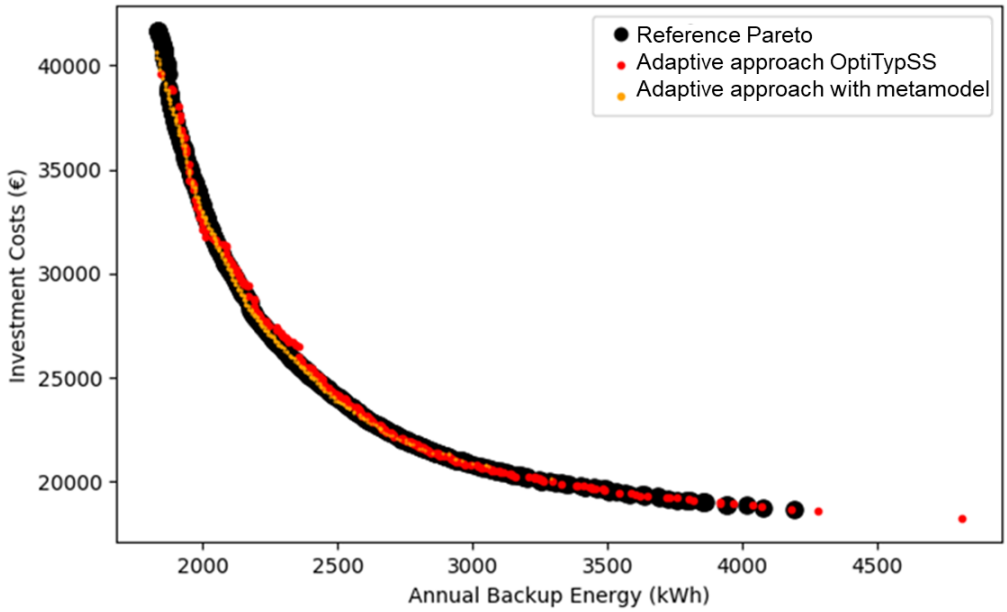


Figure IV-15. The reference Pareto front with respect to the predicted ones by the proposed strategy and metamodel.

IV.5. Conclusion

It was shown in the conclusion of the previous chapter that TypSS is still not an attractive approach in model simulation studies when it comes to global computation time due to time consumed in typical day selection. However, the accuracy of the rapidly obtained results by the generated sequence makes using it in heavy repetitive simulation based studies such as optimization an interesting field to explore. In this chapter, NSGA-II (Non-dominated Sorting Genetic Algorithm-ii) was chosen as an optimization method based on genetic algorithms and uses the notion of Pareto dominance for optimization of some parameters of the model presented in Figure II-2 . The optimized model was kept detailed and the reduced sequences was employed in a sequential and adaptive approach to accelerate the optimization process.

In the sequential approach, the optimization algorithm converged much faster when using the short sequence recording 5.7 hours to converge while it took 175 hours when using the annual data profiles on a 30 processors computer. In addition to that, the influence of the initial inputs was noticed where results improved when considering more test individuals and focusing, when generating the reduced sequence in TypSS, only on the target criteria used in the optimization study.

After that, the adaptive approach OptiTypSS was proposed and tested to improve even more the predicted Pareto front. It involves using TypSS, and not only its output

sequence, inside the optimization study. The proposed process works on benefiting from the aspects of TypSS to decrease the differences between the predicted and reference Pareto fronts, without even having the latter. This is achieved by checking the representation of the predicted Pareto front by selecting and testing several individuals taken from different parts of the Pareto front itself and validating its proximity to corresponding individuals simulated with reference data.

The obtained Pareto front improved and superposed the reference one showing a result as efficient as using metamodels. However, optimization by reduced sequence was not the fastest due to the time consumed by TypSS in the generation of the reduced sequence. However, employing metamodels can be limited for too complex models where there exist too many optimization parameters since the meta-model itself is more complex and needs more points to have good learning. While in the OptiTypSS, no simplification of the model is required and no need for a great number of individuals as previously shown. This is a promising result for more complex case studies like network models. Accelerating TypSS by reducing the number of tested days during the day selection process, parallelizing the simulations or even improving the functions to reach faster to an adequate reduced sequence will be very helpful in achieving fast and accurate optimization studies of models despite their level of complexity.

General conclusions and perspectives

The thesis studies the concept of dynamic building performance simulations by sequences of several typical days. The literature contains various approaches to select a representative set of historical periods. These approaches can be grouped in three main categories: Heuristic Approaches, Iterative Approaches and Grouping Algorithms. The predilection by the researchers into using grouping algorithms rather than other approaches, with a special interest in the K-means clustering approach was noticeable due to its practicality and efficiency. However, efficiency of a method is directly related to the case studied or optimized.

Therefore, the thesis presents and evaluates a new day selection approach called TypSS (Typical Short Sequence) Algorithm to generate robust reduced sequences and can be applied on different detailed models in characterization and optimization studies.

The approach is of an iterative aspect with an embedded grouping algorithm. It employs averaged and cumulative values of target criteria (model performances) specified by the user to evaluate both temporal performances per period and annual performances as a complete year. The algorithm divides the year into different sections, denoted periods, and selects representative days for each period creating a sequence of typical days to be used directly in dynamic simulations of detailed models.

After explaining in details the process followed by the algorithm, it was applied on a building model with a solar combisystem and a sequence of 12 days was generated. Dynamic simulation on the short sequence was about 25 times faster than the annual one using the same computer configuration. In addition to the saved simulation time, results show that the output of the short simulation sequence are of high correlation with the reference ones recording minor annual sum errors. Results were also compared to the ones obtained using sequences from clustering by K-mediods and the iterative approach SCSPT, used by researchers in the literature, on the same case study and showed best performance.

A sequence was also generated using simultaneously five individuals to evaluate the generalization capability of the method and curves follow in a very good correlation the reference annual ones, for all individuals and all target criteria. The sequence was validated on 45 other individuals not taken into consideration by the algorithm. The curves showed that the sequence succeeded in predicting the annual performances of

all target criteria with relative errors not exceeding 10%. They also showed that generating a sequence on more than a single individual improves the quality of results and favors the aim of developing a generalized sequence applicable on a wide parametric range.

After analyzing the output of a model simulation, the sensitivity of results to four initialization inputs of the algorithm were evaluated. The length of the initial sequence, length of the final sequence, number of tested individuals and number and type of the targeted target criteria were examined. The sensitivity to each input was tested separately and without crossing.

Regarding the length of the initial sequence, results showed that it is not favorable to divide initially the year into many small equal parts if the final generated sequence is relatively short. Rather leave it for the algorithm to do its breaking down based on the performance of the simulation. On the other hand, regarding the length of the generated sequence, curves show that achieving good results is still applicable even with very short final sequences. However, increasing the number of days will indeed help in achieving better performances depending on the case study.

Regarding the number of tested individuals, the data shows that the time increases proportionally as the number of individuals increase. Therefore, despite considering more individuals in the day selection process helps in giving a generalized sequence, it is essential to consider a reasonable number of individuals to avoid a high computational time expense. Finally and regarding the number and type of target criteria, the algorithm distributes its focus as more criteria are added. Thus, it is a very crucial input variable and the algorithm is highly sensitive to this parameter. The user should be aware of the criteria he is choosing to replicate in case he is interested in a multi criteria study. In this case, it is recommended to make several trials to find the best criteria combination since it is directly related to the case study and the boundary conditions of the system.

After evaluating the algorithm in a model simulation and its sensitivity to input parameters modifications, it appeared that despite the accurate results it obtains, it is not as attractive due to the time consumed in the day selection process. However, rapidly obtained accurate results by the generated reduced sequence make employing it in repetitive simulation studies interesting such as optimization. Therefore, NSGA-II

(Non-dominated Sorting Genetic Algorithm-II) was chosen as an optimization method based on genetic algorithms that uses the notion of Pareto dominance. A sequential and adaptive approach were evaluated.

In the sequential approach study, the optimization algorithm converged much faster when using the short sequence recording 5.7 hours to converge while it took 175 hours when using the annual data profiles on a 30 processors computer. In addition to that, the influence of the initial inputs was noticed where results improved when considering more test individuals and focusing only on the target criteria used in the optimization study. However, predicted Pareto fronts diverged slightly from the reference one.

In the adaptive approach study, OptiTypSS was proposed to improve the predicted Pareto front. It involves using TypSS, and not only its output sequence, inside the optimization study. The obtained Pareto front superposes the reference one showing a result as efficient as using metamodels. However, optimization by reduced sequence was not the fastest due to the time consumed by TypSS in the generation of the reduced sequence. Accelerating TypSS by activating the options included in its *Typical day enhancement* phase or parallelizing the phases' iterations will be very helpful in achieving fast and accurate optimization studies of models despite their level of complexity.

In perspective, the algorithm can be updated to improve its performance and speed up the time of convergence. Combining the second and third phases in a way that would achieve year dividing and at the same time succeeding in estimating the annual performances by the reduced sequence will help in speeding up the time of convergence. This would also remove the need for a "number of generated days in the final sequence n_{STOP}" input since a shorter good-performing sequence might be achieved before even reaching the specified final number of days. However, the majority of the computational time consumed is in the third phase, *Typical day enhancement* phase, so further work should be done first on this phase to avoid consecutive calling of a high computational time phase.

Moreover, considering other evaluation values, in addition to the coefficient of determination and the annual sum errors, inside the algorithm might improve the quality of the obtained sequence. For instance, considering the Coefficient of Variation of the Root Mean Squared Error (CVRMSE) or the Normalized Mean Bias Error

(NMBE) when day selection would enhance the choice taken by the algorithm. NMBE gives a sense of the total difference between annual predicted energy uses, and annual reference energy use. CVRMSE gives an indication of the model's ability to predict the overall load shape that is reflected in the data, i.e. temporal profiles.

In addition to that, the discontinuities found in the generated sequence may be detrimental towards some case studies such as specific controllers. Modifications can be applied on the algorithm's process to smoothen those discontinuities and become more general. This also includes adding functions that could discover specific instances in the year such as peak demand days which are important in the designing stage.

Furthermore, the algorithms (TypSS and OptiTypSS) are still in their early testing stages. They should be applied on other case studies to validate their generality. Thermal inertia is one of the main challenges to estimate when using reduced sequences. While the model tested in the thesis has thermal inertia in its envelop and storage volume, the algorithms should be tested on cases with higher thermal inertia such as large buildings. Complex heat networks are also one of the main fields of interest for time and even data reduction due to their high computational time expenses. Sensitivity analysis can also be expanded and crossing between input parameters could be applied to examine the sensitivity of the algorithm to all possible input scenarios.

Regarding the optimization strategy, while the obvious choice in the example shown in this thesis was choosing optimization by metamodels due to relatively fast convergence towards accurate results, this conclusion cannot be generalized. Metamodels require specific technical skills to create and may be inapplicable in complex cases such as heat networks. Therefore, OptiTypSS is still an interesting measure that requires more improvement. Enhancing TypSS will definitely help in accelerating OptiTypSS. However, the strategy itself could be improved. The high computational time expense is directly related to the number of individuals involved in the day selection process. Considering a lower number of individuals concentrated only around the optimal Pareto front would help in decreasing the number of useless or misleading iterations. Using special techniques or algorithms that apply learning processes from previous iterations could make convergence much faster. This opens the door towards new fields of using methods such as evolutionary (genetic) processes

in the day selection and exploring the idea of combining several strategies together using the strength points of each.

References

[1] H. Saint-Simon. Catéchisme des industriels. Paris: Imprimerie de Sétier, 1823.

[2] IEA, “Energy Technology Perspectives 2017 - Executive Summary,” Iea, p. 371, 2017.

[3] UN DESA, “World Population Prospects The 2017 Revision Key Findings and Advance Tables,” World Popul. Prospect. 2017, pp. 1–46, 2017.

[4] J. Hansen, R. Ruedy, M. Sato, et K. Lo, « GLOBAL SURFACE TEMPERATURE CHANGE », *Rev. Geophys.*, vol. 48, nº 4, p. RG4004, déc. 2010, doi: 10.1029/2010RG000345.

[5] GIEC, "Le cinquième rapport de l'évaluation" 2013. [En ligne]. Available: https://www.ipcc.ch/home_languages_main_french.shtml. [Accessed le 01 08 2020].

[6] GIEC, "Rapports d’évaluation" [Online]. Available: https://www.ipcc.ch/home_languages_main_french.shtml. [Accessed 01 08 2020].

[7] L. Pérez-Lombard, J. Ortiz, et C. Pout, « A review on buildings energy consumption information », *Energy and Buildings*, vol. 40, nº 3, p. 394-398, janv. 2008, doi: 10.1016/j.enbuild.2007.03.007.

[8] Ü. Ağbulut, İ. Ceylan, A. E. Gürel, et A. Ergün, « The history of greenhouse gas emissions and relation with the nuclear energy policy for Turkey », *International Journal of Ambient Energy*, p. 1-9, janv. 2019, doi: 10.1080/01430750.2018.1563818.

[9] L. Pérez-Lombard, J. Ortiz et C. Pout. «A review on buildings energy consumption information» Energy and Buildings, vol. 40, n° 13, pp. 394-398, 2008.

[10] D. Ürge-Vorsatz, N. Eyre, P. Graham, D. Harvey, E. Hertwich, Y. Jiang, C. Kornevall, M. Majumdar, J. E. McMahon, S. Mirasgedis, S. Murakami et A. Novikova, «Chapter 10 - Energy End-Use: Building. In Global Energy Assessment - Toward a Sustainable Future» 2012. [Online]. Available: http://www.iiasa.ac.at/web/home/research/Flagship-Projects/Global-Energy-Assessment/Chapte10.en.html. [Accessed le 01 08 2020].

[11] H. Ben Ahmed, B. Multan et Y. Thiaux, « Consommation d'énergie, ressources énergétiques et place de l'électricité » Techniques de l'Ingénieur, vol. TIB248DUO, 2011, pp. D 3 900v2-1 - D 3 900v2-18.

[12] European Commission, “Directive 2006/32 EC of the European parliment and the councile of 5 April on energy end-use efficiency and energy services and repealing Council Directive 93/76/EEC,” Off. J. Eur. Union, pp. 12–25, 2006.

[13] EC, “The Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings,” Off. J. Eur. Union, vol. 53, pp. 1–6, 2010.

[14] P. Christoff, « The promissory note: COP 21 and the Paris Climate Agreement », *Environmental Politics*, vol. 25, nº 5, p. 765-787, sept. 2016, doi: 10.1080/09644016.2016.1191818.

[15] K. Negendahl, « Building performance simulation in the early design stage: An introduction to integrated dynamic models », *Automation in Construction*, vol. 54, p. 39-53, juin 2015, doi: 10.1016/j.autcon.2015.03.002.

[16] X. Shi et W. Yang, « Performance-driven architectural design and optimization technique from a perspective of architects », *Automation in Construction*, vol. 32, p. 125-135, juill. 2013, doi: 10.1016/j.autcon.2013.01.015.

[17] P. Sanguinetti, S. Abdelmohsen, J. Lee, J. Lee, H. Sheward, et C. Eastman, « General system architecture for BIM: An integrated approach for design and analysis », *Advanced Engineering Informatics*, vol. 26, nº 2, p. 317-333, avr. 2012, doi: 10.1016/j.aei.2011.12.001.

[18] A. Zarzycki, « Exploring parametric BIM as a conceptual tool for design and building technology teaching », in *Proceedings of the 2010 Spring Simulation Multiconference on - SpringSim ’10*, Orlando, Florida, 2010, p. 1, doi: 10.1145/1878537.1878739.

[19] S. C. Christian, « Uncertainty propagation and sensitivity analysis techniques in building performance simulation to support conceptual building and system design », 2012, doi: 10.6100/IR735575.

[20] T. Banke, *Parametri i praksis: generativ performance i arkitektur : erhvervs-ph.d.-afhandling*. Kbh.: Det Kongelige Danske Kunstakademis Skoler for Arkitektur, Design og Konservering, Arkitekstskolen, 2013.

[21] S. Petersen, Danmarks Tekniske Universitet, et DTU Byg, *Simulation-based support for integrated design of new low-energy office buildings*. Lyngby: DTU Civil Engineering, Technical University of Denmark, 2011.

[22] C. Struck, P. J. C. J. de Wilde, C. J. Hopfe, et J. L. M. Hensen, « An investigation of the option space in conceptual building design for advanced building simulation », *Advanced Engineering Informatics*, vol. 23, nº 4, p. 386-395, oct. 2009, doi: 10.1016/j.aei.2009.06.004.

[23] W. Wang, H. Rivard, et R. Zmeureanu, « An object-oriented framework for simulation-based green building design optimization with genetic algorithms », *Advanced Engineering Informatics*, vol. 19, nº 1, p. 5-23, janv. 2005, doi: 10.1016/j.aei.2005.03.002.

[24] A. Sharag-Eldin, « A parametric model for predicting wind-induced pressures on low-rise vertical surfaces in shielded environments », *Solar Energy*, vol. 81, nº 1, p. 52-61, janv. 2007, doi: 10.1016/j.solener.2006.06.007.

[25] R. Baños, F. Manzano-Agugliaro, F. G. Montoya, C. Gil, A. Alcayde, et J. Gómez, « Optimization methods applied to renewable and sustainable energy: A review », *Renewable and Sustainable Energy Reviews*, vol. 15, nº 4, p. 1753-1766, mai 2011, doi: 10.1016/j.rser.2010.12.008.

[26] M. Wetter et J. Wright, « A comparison of deterministic and probabilistic optimization algorithms for nonsmooth simulation-based optimization », *Building and Environment*, vol. 39, nº 8, p. 989-999, août 2004, doi: 10.1016/j.buildenv.2004.01.022.

[27] V. Machairas, A. Tsangrassoulis, et K. Axarli, « Algorithms for optimization of building design: A review », *Renewable and Sustainable Energy Reviews*, vol. 31, p. 101-112, mars 2014, doi: 10.1016/j.rser.2013.11.036.

[28] S. Attia, M. Hamdy, W. O'Brien, et S. Carlucci, « Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings

design », *Energy and Buildings*, vol. 60, p. 110-124, mai 2013, doi: 10.1016/j.enbuild.2013.01.016.

[29] R. Hooke et T. A. Jeeves, « `` Direct Search'' Solution of Numerical and Statistical Problems », *J. ACM*, vol. 8, nº 2, p. 212-229, avr. 1961, doi: 10.1145/321062.321069.

[30] J. A. Nelder et R. Mead, « A Simplex Method for Function Minimization », *The Computer Journal*, vol. 7, nº 4, p. 308-313, janv. 1965, doi: 10.1093/comjnl/7.4.308.

[31] A.-T. Nguyen, S. Reiter, et P. Rigo, « A review on simulation-based optimization methods applied to building performance analysis », *Applied Energy*, vol. 113, p. 1043-1058, janv. 2014, doi: 10.1016/j.apenergy.2013.08.061.

[32] L. Magnier et F. Haghighat, « Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network », *Building and Environment*, vol. 45, nº 3, p. 739-746, mars 2010, doi: 10.1016/j.buildenv.2009.08.016.

[33] M. Wetter et E. Polak, « A convergent optimization method using pattern search algorithms with adaptive precision simulation », *Building Services Engineering Research and Technology*, vol. 25, nº 4, p. 327-338, nov. 2004, doi: 10.1191/0143624404bt097oa.

[34] K. Deb, A. Pratap, S. Agarwal, et T. Meyarivan, « A fast and elitist multiobjective genetic algorithm: NSGA-II », *IEEE Trans. Evol. Computat.*, vol. 6, nº 2, p. 182-197, avr. 2002, doi: 10.1109/4235.996017.

[35] S. Sette et L. Boullart, « Genetic programming: principles and applications », *Engineering Applications of Artificial Intelligence*, vol. 14, nº 6, p. 727-736, déc. 2001, doi: 10.1016/S0952-1976(02)00013-1.

[36] L. J. Fogel, *Intelligence through simulated evolution: forty years of evolutionary programming*. New York: Wiley, 1999.

[37] N. Hansen et A. Ostermeier, « Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation », in *Proceedings of IEEE International Conference on Evolutionary Computation*, Nagoya, Japan, 1996, p. 312-317, doi: 10.1109/ICEC.1996.542381.

[38] R. Storn et K. Price, « Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces », *Journal of Global Optimization*, vol. 11, nº 4, p. 341-359, 1997, doi: 10.1023/A:1008202821328.

[39] Zong Woo Geem, Joong Hoon Kim, et G. V. Loganathan, « A New Heuristic Optimization Algorithm: Harmony Search », *SIMULATION*, vol. 76, nº 2, p. 60-68, févr. 2001, doi: 10.1177/003754970107600201.

[40] J. Kennedy et R. Eberhart, « Particle swarm optimization », in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, WA, Australia, 1995, vol. 4, p. 1942-1948, doi: 10.1109/ICNN.1995.488968.

[41] M. Dorigo, V. Maniezzo, et A. Colorni, « Ant system: optimization by a colony of cooperating agents », *IEEE Trans. Syst., Man, Cybern. B*, vol. 26, nº 1, p. 29-41, févr. 1996, doi: 10.1109/3477.484436.

[42] S. Kirkpatrick, C. D. Gelatt, et M. P. Vecchi, « Optimization by Simulated Annealing », *Science*, vol. 220, nº 4598, p. 671-680, mai 1983, doi: 10.1126/science.220.4598.671.

[43] R. Evins, « A review of computational optimisation methods applied to sustainable building design », *Renewable and Sustainable Energy Reviews*, vol. 22, p. 230-245, juin 2013, doi: 10.1016/j.rser.2013.02.004.

[44] M. Hamdy, A. Hasan, et K. Siren, « Applying a multi-objective optimization approach for Design of low-emission cost-effective dwellings », *Building and Environment*, vol. 46, nº 1, p. 109-123, janv. 2011, doi: 10.1016/j.buildenv.2010.07.006.

[45] M. Fesanghary, S. Asadi, et Z. W. Geem, « Design of low-emission and energy-efficient residential buildings using a multi-objective optimization algorithm », *Building and Environment*, vol. 49, p. 245-250, mars 2012, doi: 10.1016/j.buildenv.2011.09.030.

[46] J. A. Wright, H. A. Loosemore, et R. Farmani, « Optimization of building thermal design and control by multi-criterion genetic algorithm », *Energy and Buildings*, vol. 34, nº 9, p. 959-972, oct. 2002, doi: 10.1016/S0378-7788(02)00071-3.

[47] C. Diakaki, E. Grigoroudis, N. Kabelis, D. Kolokotsa, K. Kalaitzakis, et G. Stavrakakis, « A multi-objective decision model for the improvement of energy

efficiency in buildings », *Energy*, vol. 35, nº 12, p. 5483-5496, déc. 2010, doi: 10.1016/j.energy.2010.05.012.

[48] F. P. Chantrelle, H. Lahmidi, W. Keilholz, M. E. Mankibi, et P. Michel, « Development of a multicriteria tool for optimizing the renovation of buildings », *Applied Energy*, vol. 88, nº 4, p. 1386-1394, avr. 2011, doi: 10.1016/j.apenergy.2010.10.002.

[49] E. Triantaphyllou, « Multi-Criteria Decision Making Methods », in *Multi-criteria Decision Making Methods: A Comparative Study*, vol. 44, Boston, MA: Springer US, 2000, p. 5-21.

[50] M. Stadler, M. Groissböck, G. Cardoso, et C. Marnay, « Optimizing Distributed Energy Resources and building retrofits with the strategic DER-CAModel », *Applied Energy*, vol. 132, p. 557-567, nov. 2014, doi: 10.1016/j.apenergy.2014.07.041.

[51] E. Merkel, R. McKenna, et W. Fichtner, « Optimisation of the capacity and the dispatch of decentralised micro-CHP systems: A case study for the UK », *Applied Energy*, vol. 140, p. 120-134, févr. 2015, doi: 10.1016/j.apenergy.2014.11.036.

[52] C. Milan, C. Bojesen, et M. P. Nielsen, « A cost optimization model for 100% renewable residential energy supply systems », *Energy*, vol. 48, nº 1, p. 118-127, déc. 2012, doi: 10.1016/j.energy.2012.05.034.

[53] D. Lauinger, P. Caliandro, J. Van herle, et D. Kuhn, « A linear programming approach to the optimization of residential energy systems », *Journal of Energy Storage*, vol. 7, p. 24-37, août 2016, doi: 10.1016/j.est.2016.04.009.

[54] K. Peippo, P. D. Lund, et E. Vartiainen, « Multivariate optimization of design trade-offs for solar low energy buildings », *Energy and Buildings*, vol. 29, nº 2, p. 189-205, janv. 1999, doi: 10.1016/S0378-7788(98)00055-3.

[55] J. Wang, Z. (John) Zhai, Y. Jing, et C. Zhang, « Particle swarm optimization for redundant building cooling heating and power system », *Applied Energy*, vol. 87, nº 12, p. 3668-3679, déc. 2010, doi: 10.1016/j.apenergy.2010.06.021.

[56] M. J. N. Oliveira Panão, H. J. P. Gonçalves, et P. M. C. Ferrão, « Optimization of the urban building efficiency potential for mid-latitude climates using a genetic

algorithm approach », *Renewable Energy*, vol. 33, nº 5, p. 887-896, mai 2008, doi: 10.1016/j.renene.2007.04.014.

[57] J. H. Lee, « Optimization of indoor climate conditioning with passive and active methods using GA and CFD », *Building and Environment*, vol. 42, nº 9, p. 3333-3340, sept. 2007, doi: 10.1016/j.buildenv.2006.08.029.

[58] P. Mancarella, « MES (multi-energy systems): An overview of concepts and evaluation models », *Energy*, vol. 65, p. 1-17, févr. 2014, doi: 10.1016/j.energy.2013.10.041.

[59] C. Milan, M. Stadler, G. Cardoso, et S. Mashayekh, « Modeling of non-linear CHP efficiency curves in distributed energy systems », *Applied Energy*, vol. 148, p. 334-347, juin 2015, doi: 10.1016/j.apenergy.2015.03.053.

[60] M. Adamski, « Optimization of the form of a building on an oval base », *Building and Environment*, vol. 42, nº 4, p. 1632-1643, avr. 2007, doi: 10.1016/j.buildenv.2006.02.004.

[61] W. Marks, « Multicriteria optimisation of shape of energy-saving buildings », *Building and Environment*, vol. 32, nº 4, p. 331-339, juill. 1997, doi: 10.1016/S0360-1323(96)00065-0.

[62] N. A. D'Cruz et A. D. Radford, « A multicriteria model for building performance and design », *Building and Environment*, vol. 22, nº 3, p. 167-179, janv. 1987, doi: 10.1016/0360-1323(87)90005-9.

[63] D. Castro-Lacouture, J. A. Sefair, L. Flórez, et A. L. Medaglia, « Optimization model for the selection of materials using a LEED-based green building rating system in Colombia », *Building and Environment*, vol. 44, nº 6, p. 1162-1170, juin 2009, doi: 10.1016/j.buildenv.2008.08.009.

[64] J. Michalek, R. Choudhary, et P. Papalambros, « Architectural layout design optimization », *Engineering Optimization*, vol. 34, nº 5, p. 461-484, janv. 2002, doi: 10.1080/03052150214016.

[65] W. Wang, R. Zmeureanu, et H. Rivard, « Applying multi-objective genetic algorithms in green building design optimization », *Building and Environment*, vol. 40, nº 11, p. 1512-1525, nov. 2005, doi: 10.1016/j.buildenv.2004.11.017.

[66] T. Hemker, K. R. Fowler, M. W. Farthing, et O. von Stryk, « A mixed-integer simulation-based optimization approach with surrogate functions in water resources management », *Optim Eng*, vol. 9, nº 4, p. 341-360, déc. 2008, doi: 10.1007/s11081-008-9048-0.

[67] K. Klemm, W. Marks, et A. J. Klemm, « Multicriteria optimisation of the building arrangement with application of numerical simulation », *Building and Environment*, vol. 35, nº 6, p. 537-544, août 2000, doi: 10.1016/S0360-1323(99)00046-3.

[68] L. Chen, Q. Fang, et Z. Zhang, « Research on the identification of temperature in intelligent building based on feed forward neural network and particle swarm optimization algorithm », in *2010 Sixth International Conference on Natural Computation*, Yantai, China, août 2010, p. 1816-1820, doi: 10.1109/ICNC.2010.5584480.

[69] B. Eisenhower, Z. O'Neill, S. Narayanan, V. A. Fonoberov, et I. Mezić, « A methodology for meta-model based optimization in building energy models », *Energy and Buildings*, vol. 47, p. 292-301, avr. 2012, doi: 10.1016/j.enbuild.2011.12.001.

[70] E. Gengembre, B. Ladevie, O. Fudym, et A. Thuillier, « A Kriging constrained efficient global optimization approach applied to low-energy building design problems », *Inverse Problems in Science and Engineering*, vol. 20, nº 7, p. 1101-1114, oct. 2012, doi: 10.1080/17415977.2012.727084.

[71] R. Roy, S. Hinduja, et R. Teti, « Recent advances in engineering design optimisation: Challenges and future trends », *CIRP Annals*, vol. 57, nº 2, p. 697-715, 2008, doi: 10.1016/j.cirp.2008.09.007.

[72] F. Boithias, M. El Mankibi, et P. Michel, « Genetic algorithms based optimization of artificial neural network architecture for buildings' indoor discomfort and energy consumption prediction », *Build. Simul.*, vol. 5, nº 2, p. 95-106, juin 2012, doi: 10.1007/s12273-012-0059-6.

[73] M. Turrin, P. von Buelow, et R. Stouffs, « Design explorations of performance driven geometry in architectural design using parametric modeling and genetic

algorithms », *Advanced Engineering Informatics*, vol. 25, nº 4, p. 656-675, oct. 2011, doi: 10.1016/j.aei.2011.07.009.

[74] A. Belderbos et E. Delarue, « Accounting for flexibility in power system planning with renewables », *International Journal of Electrical Power & Energy Systems*, vol. 71, p. 33-41, oct. 2015, doi: 10.1016/j.ijepes.2015.02.033.

[75] M. Haller, S. Ludig, et N. Bauer, « Decarbonization scenarios for the EU and MENA power system: Considering spatial distribution and short term dynamics of renewable generation », *Energy Policy*, vol. 47, p. 282-290, août 2012, doi: 10.1016/j.enpol.2012.04.069.

[76] M. Fripp, « Switch: A Planning Tool for Power Systems with Large Shares of Intermittent Renewable Energy », *Environmental Science & Technology*, vol. 46, nº 11, p. 6371-6378, juin 2012, doi: 10.1021/es204645c.

[77] E. K. Hart et M. Z. Jacobson, « A Monte Carlo approach to generator portfolio planning and carbon emissions assessments of systems with large penetrations of variable renewables », *Renewable Energy*, vol. 36, nº 8, p. 2278-2286, août 2011, doi: 10.1016/j.renene.2011.01.015.

[78] J. Ortiga, J. C. Bruno, et A. Coronas, « Selection of typical days for the characterisation of energy demand in cogeneration and trigeneration optimisation models for buildings », *Energy Conversion and Management*, vol. 52, nº 4, p. 1934-1942, avr. 2011, doi: 10.1016/j.enconman.2010.11.022.

[79] M. Albaric, J. Nowag, et P. Papillon, « Thermal performance evaluation of solar combisystems using a global approach», EUROSUN2008, oct. 2008.

[80] K. Poncelet, H. Hoschle, E. Delarue, A. Virag, et W. Drhaeseleer, « Selecting Representative Days for Capturing the Implications of Integrating Intermittent Renewables in Generation Expansion Planning Problems », *IEEE Transactions on Power Systems*, vol. 32, nº 3, p. 1936-1948, mai 2017, doi: 10.1109/TPWRS.2016.2596803.

[81] M. S. ElNozahy, M. M. A. Salama, et R. Seethapathy, « A probabilistic load modelling approach using clustering algorithms », 2013, p. 1-5, doi: 10.1109/PESMG.2013.6672073.

[82] L. Kotzur, P. Markewitz, M. Robinius, et D. Stolten, « Impact of different time series aggregation methods on optimal energy system design », *Renewable Energy*, vol. 117, p. 474-487, mars 2018, doi: 10.1016/j.renene.2017.10.017.

[83] S. Fazlollahi, S. L. Bungener, P. Mandel, G. Becker, et F. Maréchal, « Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods », *Computers & Chemical Engineering*, vol. 65, p. 54-66, juin 2014, doi: 10.1016/j.compchemeng.2014.03.005.

[84] F. Domínguez-Muñoz, J. M. Cejudo-López, A. Carrillo-Andrés, et M. Gallardo-Salazar, « Selection of typical demand days for CHP optimization », *Energy and Buildings*, vol. 43, nº 11, p. 3036-3043, nov. 2011, doi: 10.1016/j.enbuild.2011.07.024.

[85] D. Menegon, A. Soppelsa, et R. Fedrizzi, « Development of a new dynamic test procedure for the laboratory characterization of a whole heating and cooling system », *Applied Energy*, vol. 205, p. 976-990, nov. 2017, doi: 10.1016/j.apenergy.2017.08.120.

[86] C. Ribault, « Multicriteria optimizatiion method for design assistance for urban densification process », PhD thesis study, nov.2017.

[87] P. Balachandra et V. Chandru, « Modelling electricity demand with representative load curves », *Energy*, vol. 24, nº 3, p. 219-230, mars 1999, doi: 10.1016/S0360-5442(98)00096-6.

[88] S. Fazlollahi, G. Becker, A. Ashouri, et F. Maréchal, « Multi-objective, multi-period optimization of district energy systems: IV – A case study », *Energy*, vol. 84, p. 365-381, mai 2015, doi: 10.1016/j.energy.2015.03.003.

[89] D. Cheze, P. Papillon, A. Leconte, M.Y. Haller, R. Haberl, T. Persson, C. Bales, R. Haberl (2014). «Towards an harmonized whole system test method for combined renewable heating systems for houses ». EUROSUN2014, sep 2014.

[90] Dott R, Haller MY, Ruschenburg J, Ochs F, Bony J. The Reference Framework for System Simulations - Part B: Buildings and Space Heat Load. IEA SHC Task 44 / HPP Annex 38; 2013.

[91] MacSheep Project web page. Available at: <https://www.macsheep.spf.ch/Project.17970.0.html.> [accessed 23.02.2020.

[92] ASHRAE. 2002. ASHRAE Guideline 14-2002, Measurement of Energy and Demand Savings. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

[93] H. Sayegh, A. Leconte, G. Fraisse, E. Wurtz, S. Rouchier (2019). «Determination of a Short Simulation Sequence for the Multi-Criteria Optimization of Buildings: A Case Study ». BS19, sep 2019.

[94] W.madmax42, « Etude des méthodes d'optimisation Multicritères ». Scribd.

[95] Y. Lu et S. Wang, « Optimal Design of Renewable Energy Systems in Low/Zero Energy Buildings », p. 11.

[96] W. Gong, Q. Duan, J. Li, C. Wang, Z. Di, A. Ye, C. Miao, Y. Dai « Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models: MULTIOBJECTIVE ADAPTIVE SURROGATE BASED OPTIMIZATION », *Water Resour. Res.*, vol. 52, nº 3, p. 1984-2008, mars 2016, doi: 10.1002/2015WR018230.

[97]https://www.tutorialspoint.com/genetic_algorithms/genetic_algorithms_parent_selecti on.htm (last visit 16/09/2020).

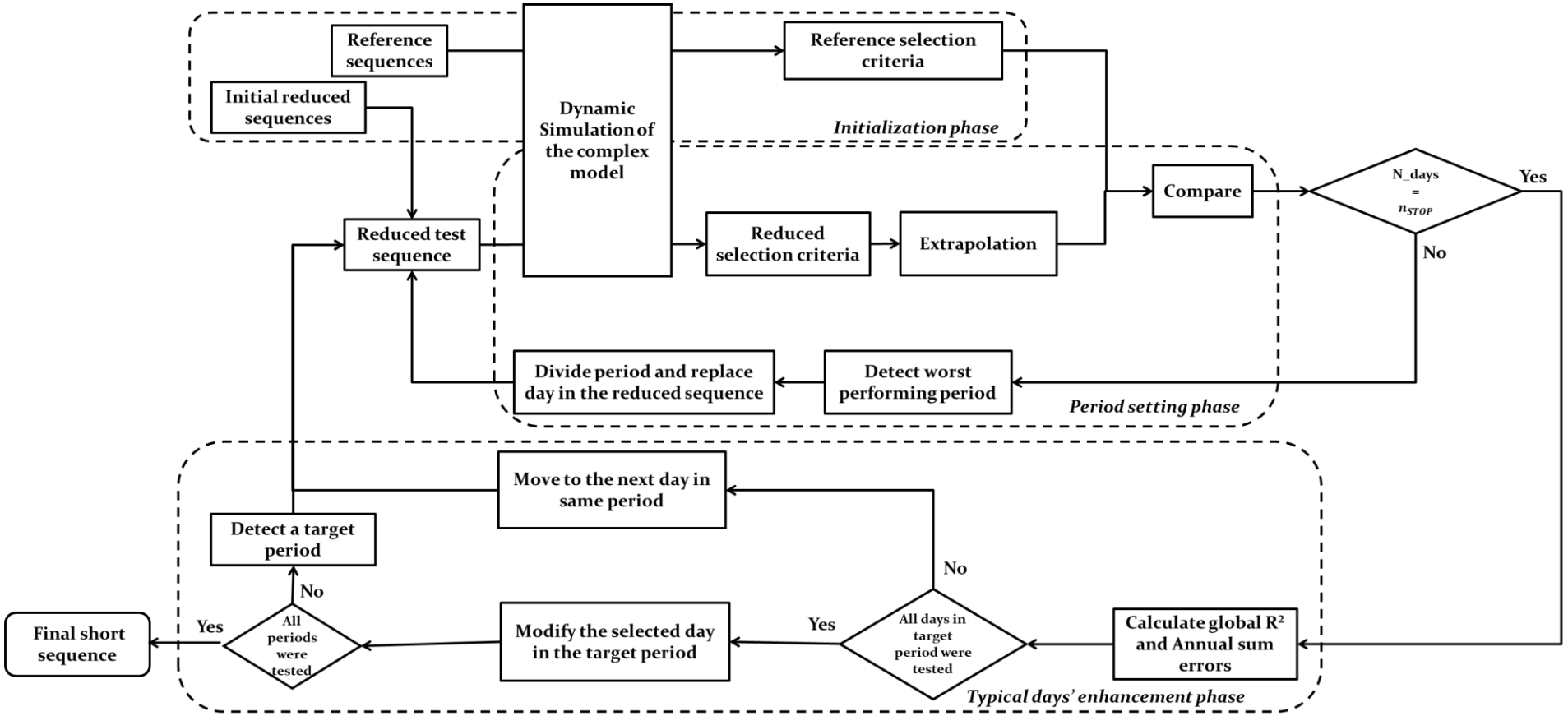
[98] G. Fraisse, M. El Mankibi, B. Peuportier, F. Wurtz, E. François, S. Truchet, J.P. Faure, X. Brunotte. INTégration ENergétique des Systèmes et de l'Enveloppe des bâtiments : développement d'une méthodologie et d'un outil de conception optimisant la performance globale. Compte-rendu de fin de projet INTENSE, Projet ANR-13-VBDU-0003-01. 2018, 33 pages.

[99] Luis M. Camarinha-Matos, Ricardo Almeida, José Oliveira, Technological Innovation for Industry and Service Systems: 10th IFIP WG 5.5/SOCOLNET, May 8-10, 2019.

[100] INTENSE : Intégration énergétique des systèmes et de l'enveloppe des bâtiments. Programme ANR Ville et Bâtiments Durables 2013, Compte rendu de fin de projet. 33 pages. 2018.

APPENDICES

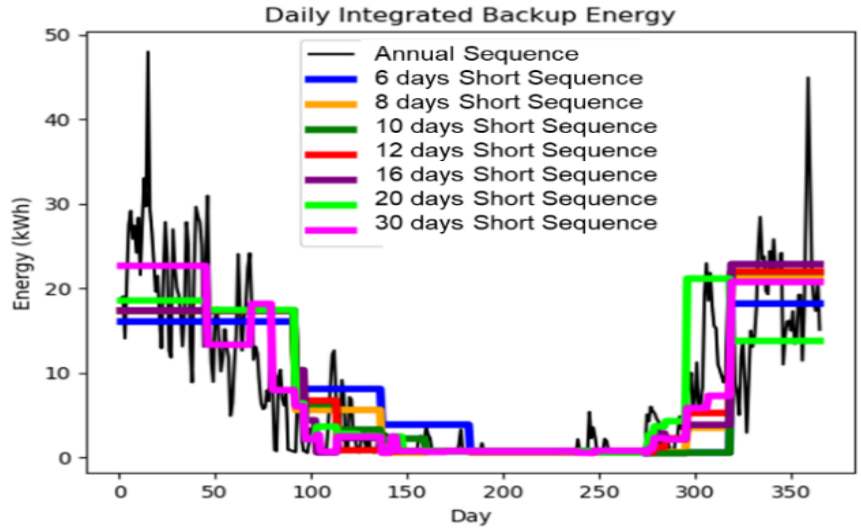
Appendix A. Typical Short Sequence (TypSS) algorithm

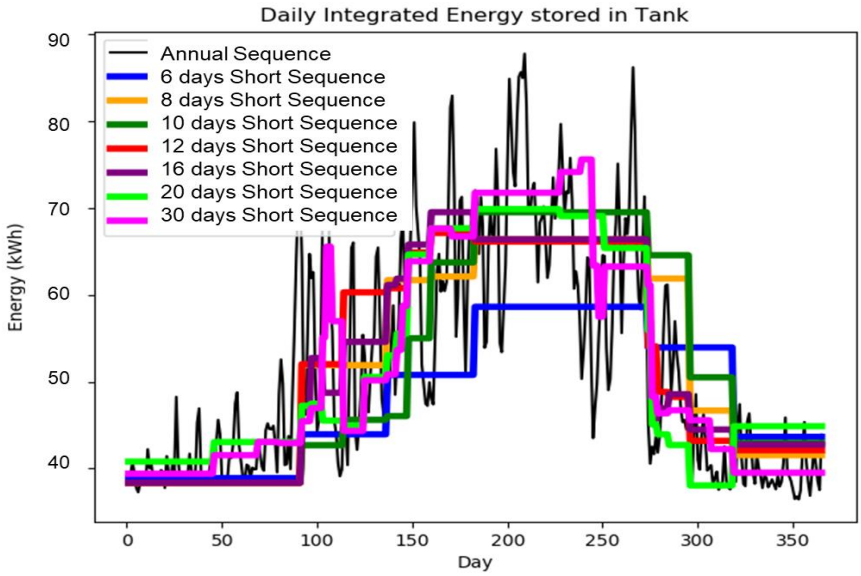


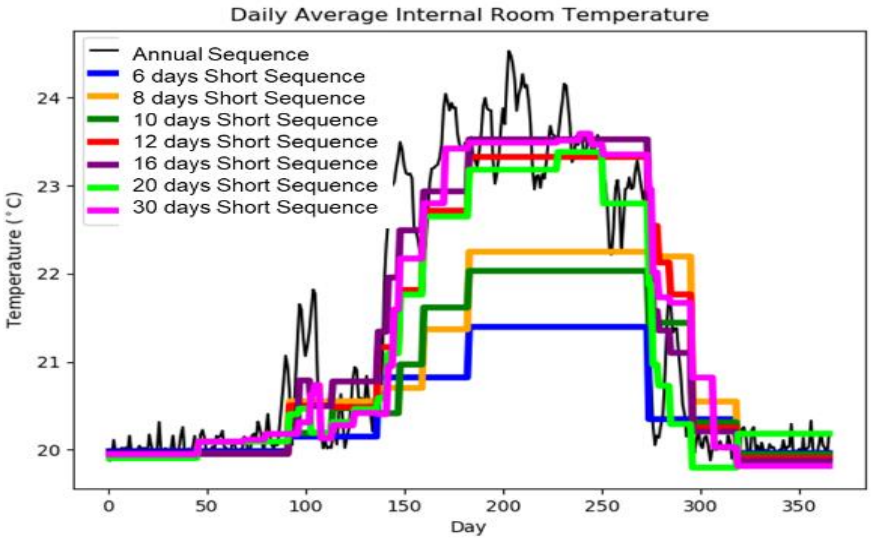
Appendix B. Fifty samples generated by LHS

SCOLL(m^2)	VST (m^3)	INS (m)
6,5	0,3	0,04
15,5573	0,307292	0,0589583
8,61979	0,321875	0,151042
18,6406	0,336458	0,232292
16,7135	0,351042	0,156458
9,77604	0,365625	0,183542
24,8073	0,380208	0,161875
22,8802	0,394792	0,226875
19,7969	0,409375	0,210625
13,2448	0,423958	0,064375
7,46354	0,438542	0,188958
10,5469	0,453125	0,129375
14,401	0,467708	0,259375
14,0156	0,482292	0,243125
6,69271	0,496875	0,167292
20,5677	0,511458	0,221458
12,0885	0,526042	0,297292
12,474	0,540625	0,080625
21,724	0,555208	0,216042
12,8594	0,569792	0,123958
15,1719	0,584375	0,0914583
9,00521	0,598958	0,286458
7,84896	0,613542	0,048125
8,23438	0,628125	0,291875
18,2552	0,642708	0,281042
20,9531	0,657292	0,275625
17,099	0,671875	0,264792
22,1094	0,686458	0,134792
15,9427	0,701042	0,178125
23,2656	0,715625	0,199792
16,3281	0,730208	0,0535417
19,4115	0,744792	0,0752083
10,9323	0,759375	0,237708
7,07812	0,773958	0,248542
9,39062	0,788542	0,102292
21,3385	0,803125	0,253958
17,8698	0,817708	0,140208
14,7865	0,832292	0,096875
23,651	0,846875	0,0427083
19,026	0,861458	0,205208
17,4844	0,876042	0,0697917
11,3177	0,890625	0,194375
24,0365	0,905208	0,270208
24,4219	0,919792	0,107708
22,4948	0,934375	0,0860417
13,6302	0,948958	0,113125
11,7031	0,963542	0,172708
20,1823	0,978125	0,145625
10,1615	0,992708	0,118542
25	1	0,3

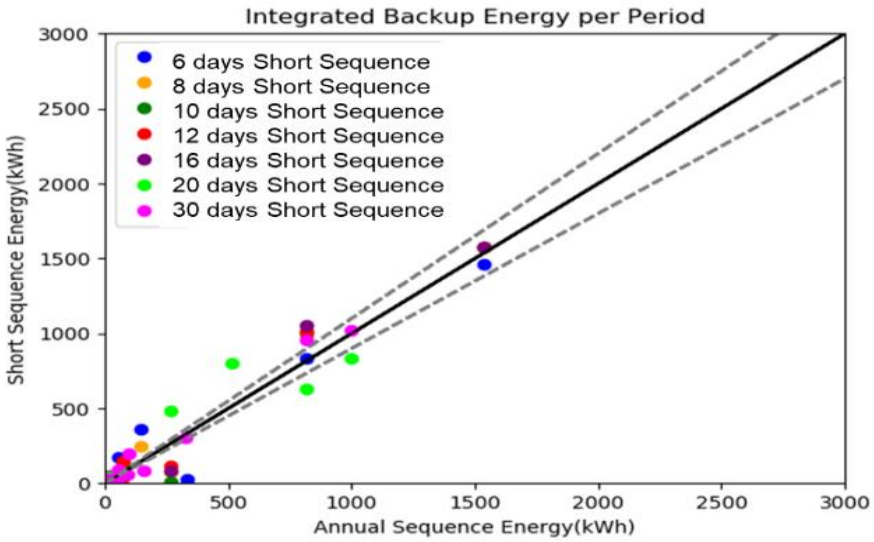
Appendix C. Temporal profiles obtained by reduced sequences of different lengths

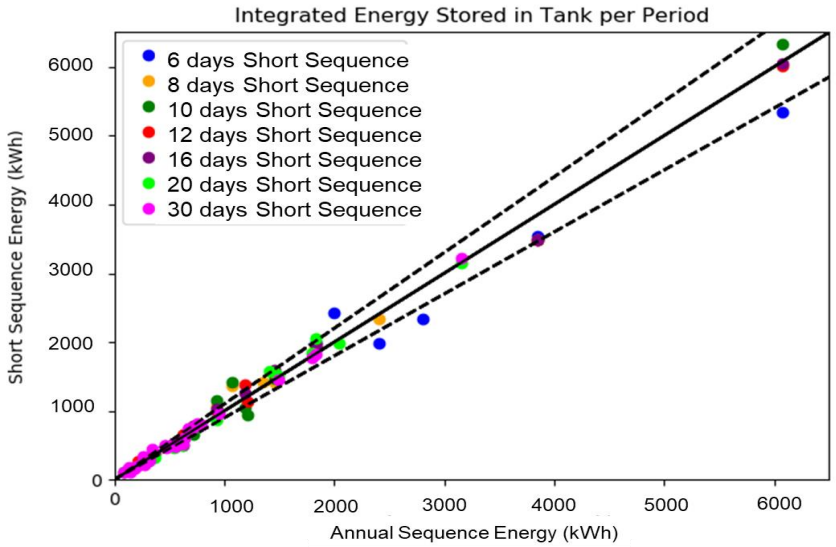


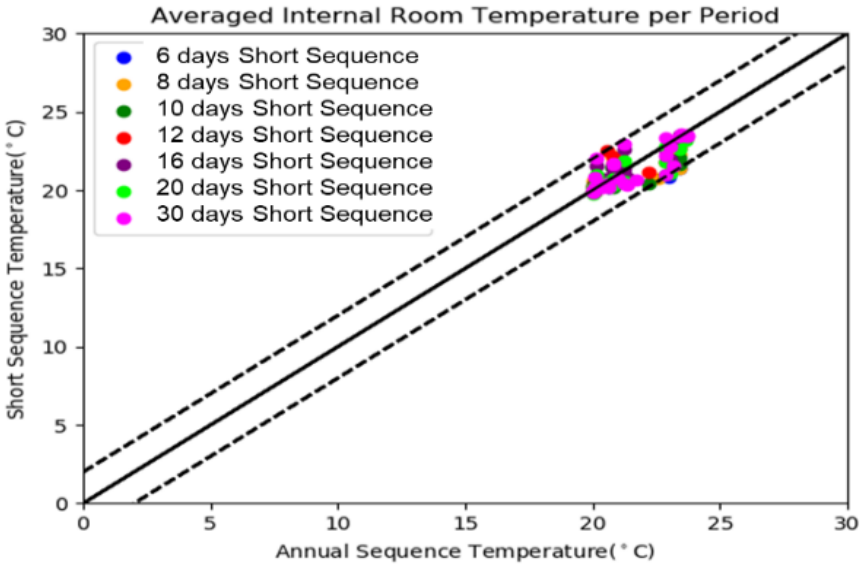




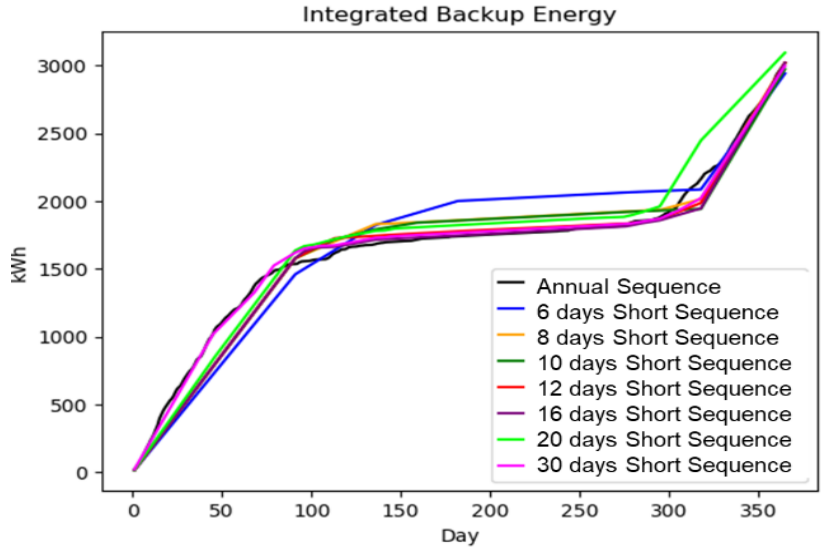
Appendix D. Periodic values obtained by reduced sequences of different lengths

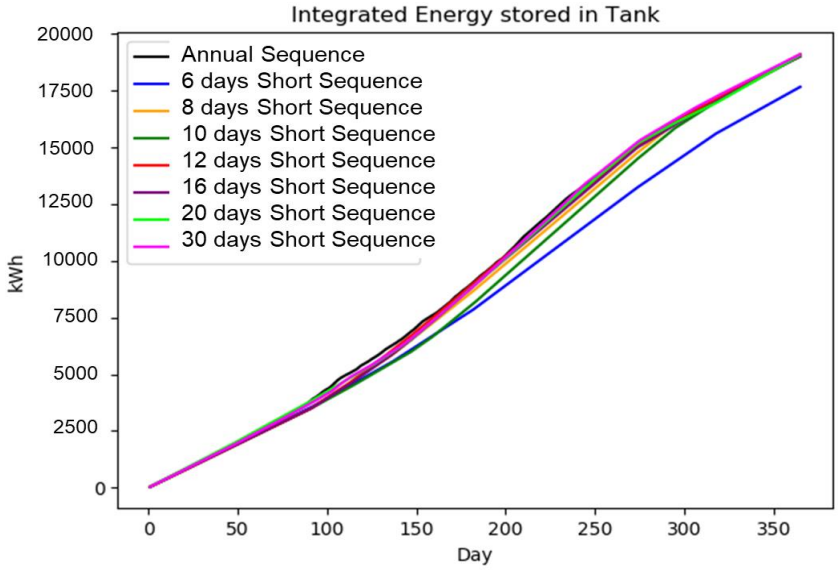


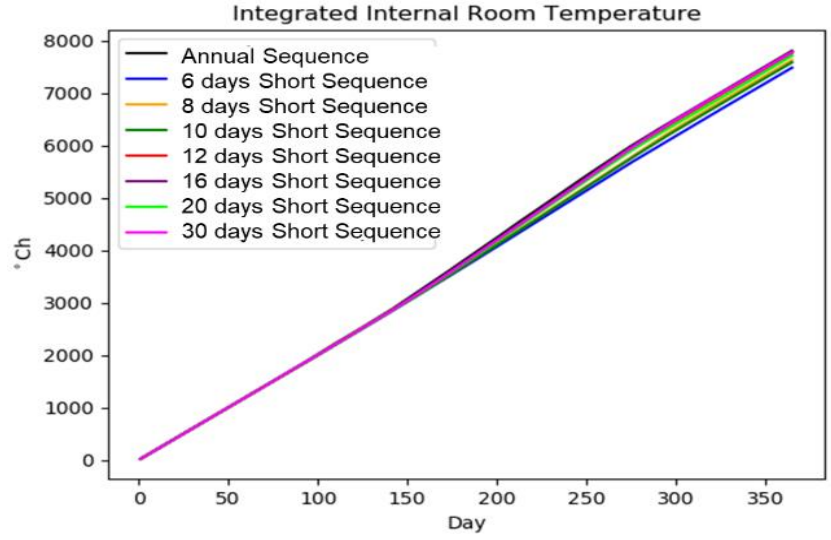




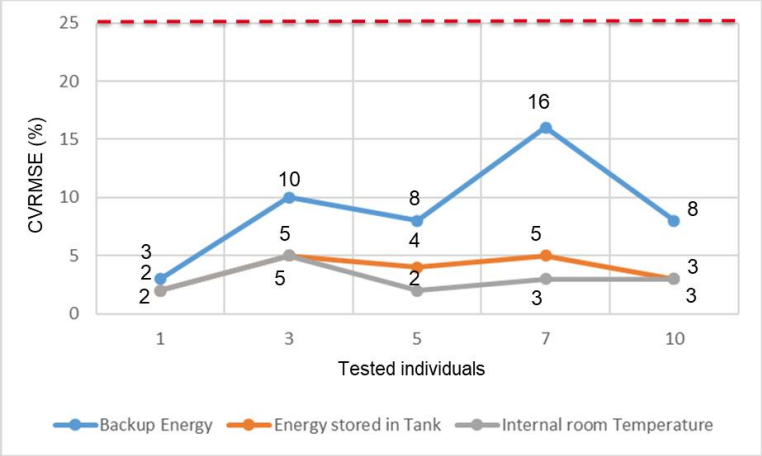
Appendix E. Cumulative profiles obtained by reduced sequences of different lengths



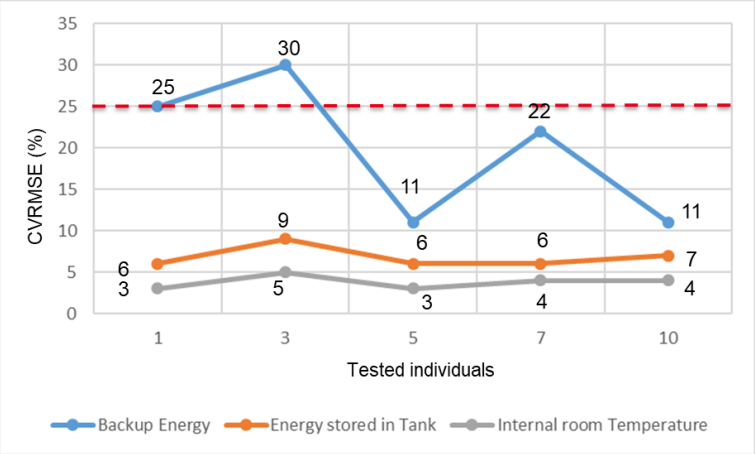




Appendix F. CVRMSE influenced by the number of tested individuals

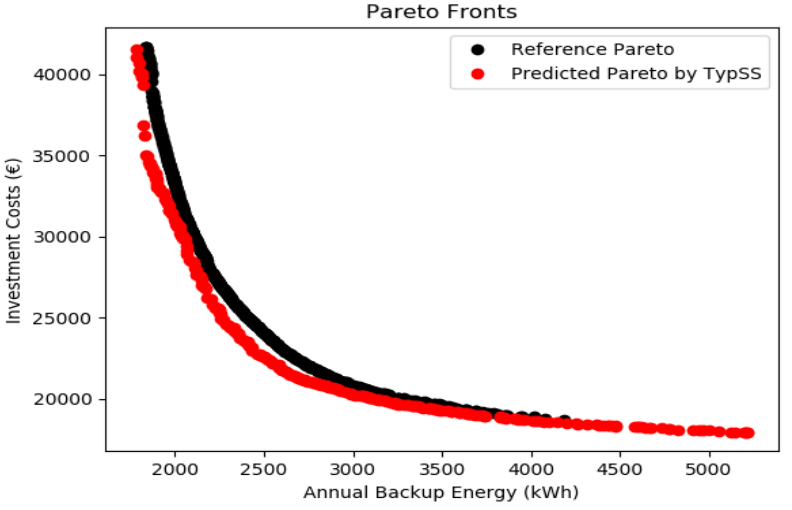


Maximum value recorded between the corresponding individuals

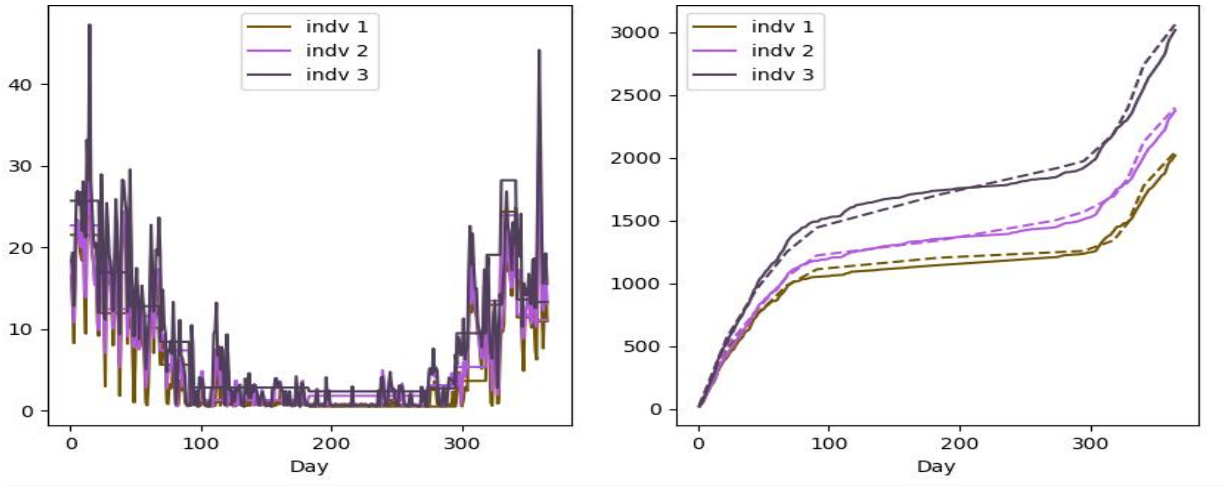


Maximum value recorded between the 50 original individuals

Appendix G. Two target criteria, three individuals Pareto front



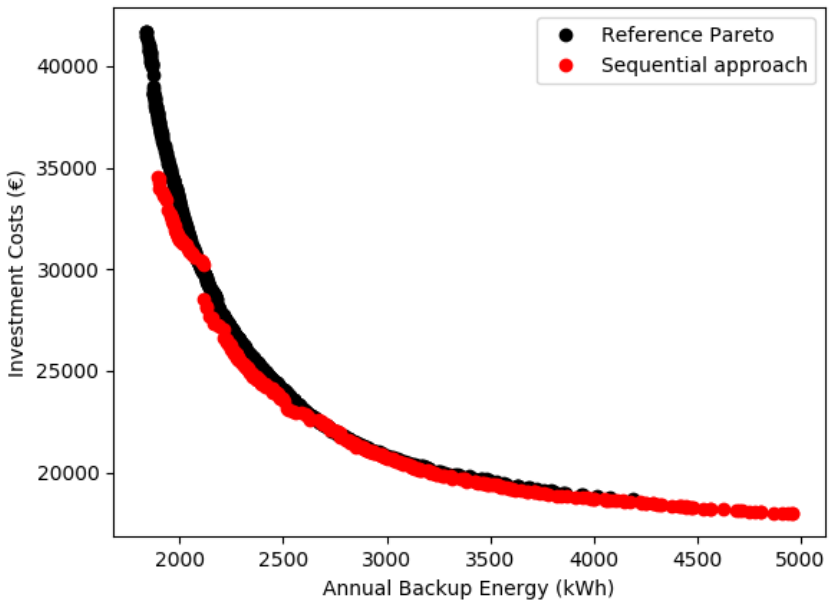
Appendix H. Considering only three Pareto front individuals



Daily (left) and integrated (right) backup energy of the three individuals selected from the Pareto front generated in the first round of OptiTypSS

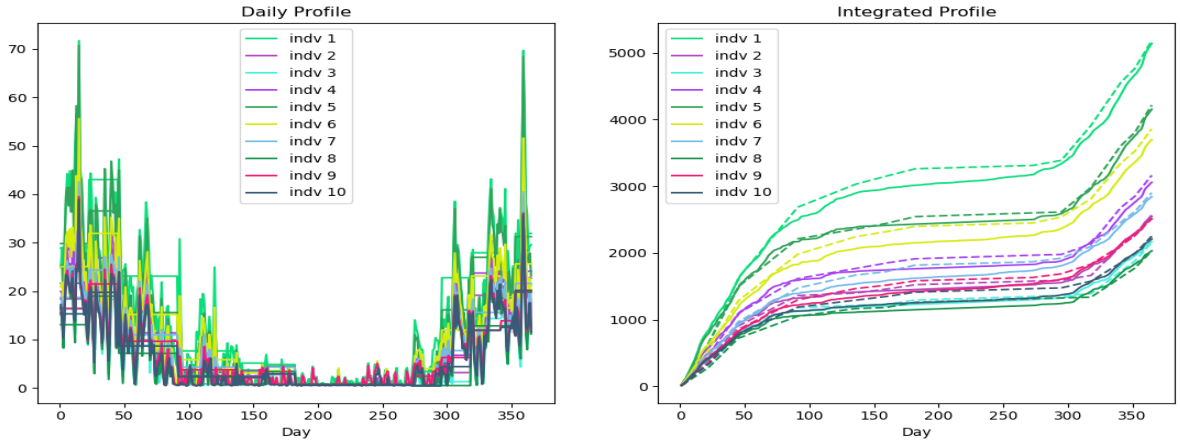
Indv	Annual sum (kWh)	Predicted sum (kWh)	Relative error (%)
1	2445	2351	3.8
2	2037	1974	3
3	3428	3378	1.4

Data of the selected individuals



Pareto Front obtained by considering only the three selected individuals without the initial ones

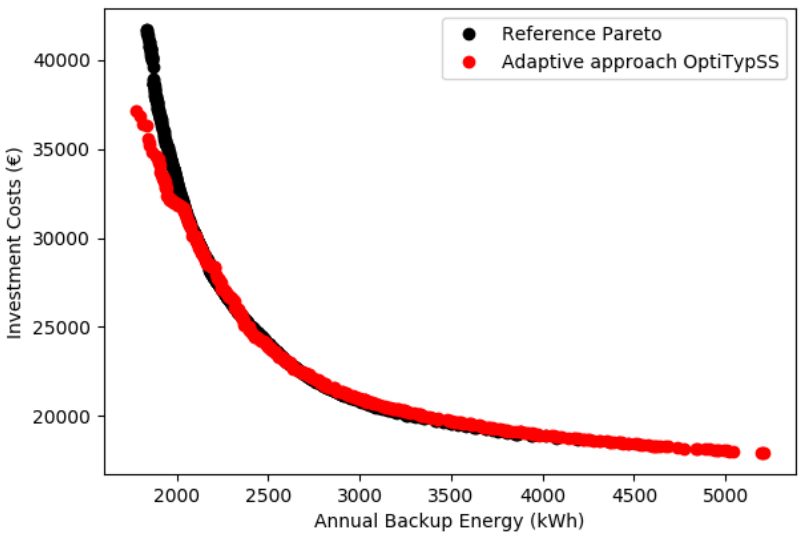
Appendix I. Considering 10 individuals in OptiTypSS



Daily (left) and integrated (right) backup energy of all ten individuals (initial and selected from the Pareto front generated in the first round of OptiTypSS)

Indv	Annual sum (kWh)	Predicted sum (kWh)	Relative error (%)
1	1979	1900	4
2	2063	2037	1.2
3	2196	2208	1
4	2575	2547	1
5	3513	3652	3.9

Data of the selected individuals



Pareto Front obtained by considering all ten individuals