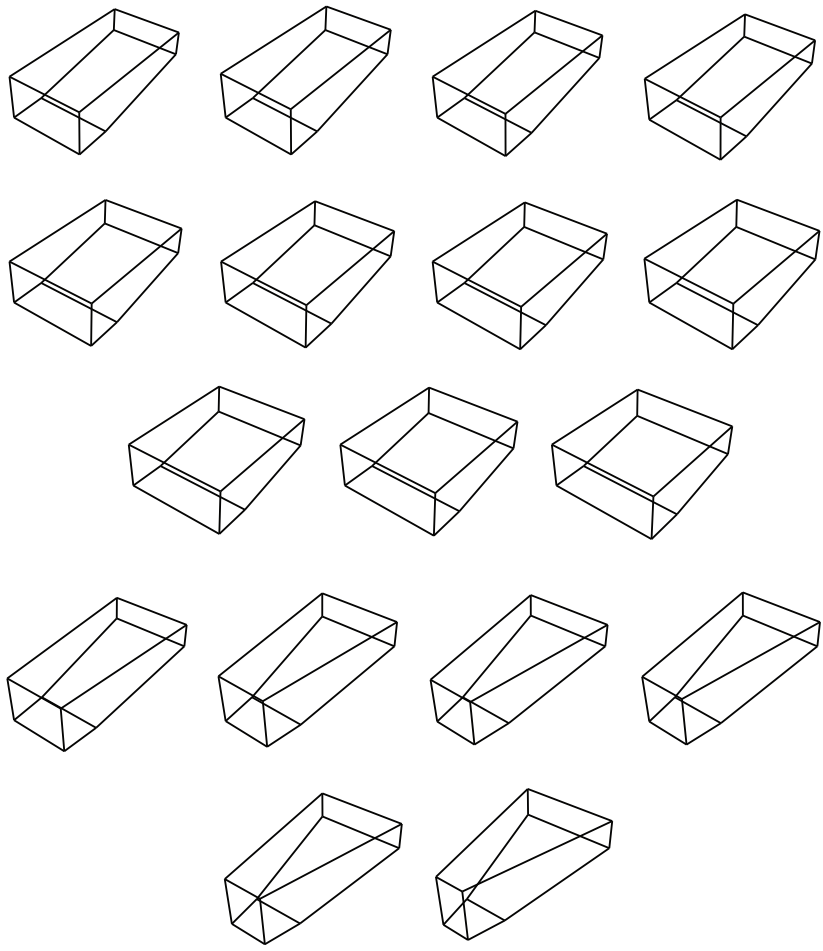


Computational Search in Architectural Design

By Tomás Méndez Echenagucia

Computational Search in Architectural Design

By Tomás Méndez Echenagucia

.

Tomás Méndez Echenagucia

Computational Search in Architectural Design

Tutors: Mario Sassone · Pierre-Alain Croset · Arianna Astolfi

A thesis submitted for the degree of "Doctor of Philosophy"
XXVI cycle
2011 · 2012 · 2013
Ph.D. in "Architecture and Building Design"
Politecnico di Torino

.

Ph.D. in “Architecture and Building Design”
XXVI cycle
2011 · 2012 · 2013
Politecnico di Torino
Viale Mattioli 39
I-10125, Torino, Italy
Ph.D. candidate: Tomás Méndez Echenagucia
Registration number: 179121
Tutors: Mario Sassone, Pierre-Alain Croset and Arianna Astolfi

.

A Perucho.

Contents

Acknoledgements

Back in 2006 I had a first conversation with Maarten Jansen, Mario Sassone and Arianna Astolfi about what i wanted to do for my Master's thesis. In the following months they proceeded to turn a naively ambitious idea into serious research, they taught me to understand design challenges as serious research opportunities, to elevate a project to a learning experience. This PhD research is an inevitable consequence of that meeting and their generosity.

I first want to thank my tutors Pierre-Alain Croset, Mario Sassone and Arianna Astolfi, their constant work and guidance in their respective fields, their generosity and perseverance made them ideal mentors and companions in this research.

True multi-disciplinary work is impossible without constant collaboration between specialists, this research would not have been possible without the constant work of Louena Shtrepi, Arthur van der Harten, Alfonso Capozzoli and Ylenia Cascone. I will always be grateful for their contribution.

I received a great deal of help from my Master thesis students Marco Palma, Maddalena Sarotto, Sabrina Canale, Silvia Pastorino, Chiara Bertolutti and Denise Barbaroux.

My PhD colleagues from various departments were a great source of support and discussion, i would like to single out Dario Parigi, Andrea Rosada, Silvia Cammarano, Lorenza Bianco, Iasef Rian, Andrea Dutto, Paolo Tecchio, Shaghayegh Rajabzadeh, Giuseppina Puglisi, Antonio Spinelli, Ilaria Ariolfo, Matteo Malandrino and Enrico Boffa.

Professional work is most often not only the inspiration for this kind of architectural research, but quite often it is the source of great knowledge. Working in interesting projects gave me the curiosity and thirst for research that took me through this process. Priceless advice, discussions and practical knowledge was imparted to me by working with Maarten Jansen, Vanja Frlan, Mawari Nuñez, Daniel Otero, Alejandro Méndez, Kristian Ceballos,

Frans Swarte, Alina Delgadillo, Michel Cova, Marco Amosso, Domenico Ghirotto, Daniel Beckman, Motoo Komoda and Bob Mahoney.

Conversations with may people were also invaluable for me during these years, sometimes they would involve the research and practical questions, other times just a simple outlook on design, science or life. E-mails, talks or phone calls with the following people helped me a lot: Alesia Griginis, Alberto Pugnale, Elisa Cattaneo, Daniel Bosia, Steve Baer, Andreas Kloeckner, Isabella Rombi, Ricardo Avella, Ariadna Weisshar, Daniel Alvares, Gustavo Méndez, Juan Pablo Méndez, Argenis Lugo, Maria Eugenia Sosa, Domingo Acosta, Alfredo Cilento, Luis Rosales, Beatriz Hernandez, Winfried Lachenmayr, Eckard Kahle, Thomas Scelo, Tappio Lokki, Sakkari Tervo, Yann Yurkoviz, Martin Vercammen and Kirk Martini.

I also have a deep appreciation of the financial support given to me by the Politecnico di Torino. In these days if austerity i never took for granted the significance of this contribution.

The constant support of my parents though the years is the reason why I am here.

Introduction

The computer entered into the vast majority of architectural design studios by the mid 90's, but for many years it was used only as a replacement to the drafting table. Popular Computer Aided Design (CAD) applications, along with the plotter, were mostly a computational version of the same tools architects had been using for centuries. The aid they provided the designer was mostly in the drawing area, not too much in the design or construction fields. Improvements in this first phase of computation in the profession relate to speed and reliability, they do not represent significant functional additions to the architect's toolbox.

Coming from the mechanical industry, computational drafting tools have their origins in the post-war research projects of american universities, most notably the MIT. As is the case with many other appropriated technologies, their diffusion had to wait many years.

After the mid 90s, architects were pushed as many other professions, in the mainstream of the informatics revolution: software became more and more sophisticated and powerful, and the complexity of digital tools was very high.

Computer graphics advancements, as well as 3D modeling techniques, produced a series of computer applications destined to create 3D renderings. These applications were mostly intended for use in the cinema and entertainment fields, but they found their way into the architectural office soon enough. Architects were eager to communicate their ideas to their clients in a more direct and intuitive way. The images generated can surely achieve photo-realistic results in a way that was not possible with previous mediums, but again here, now new design or construction functionality was introduced. Applications that were destined for the entertainment industry however, were endowed with other characteristics that interested architects. Representation was a part of their use, but the modeling techniques used in this industry were more than mere geometrical objects, the models con-

tained hierarchical and invariant relationships. These features were helping the architects create and modify their 3D models in better ways.

Information began to find its way into the models as well. Architects began to use the digital model to generate and store information. Perhaps the most heard about use of the computer in the architectural office nowadays is the use of Building Information Modeling or BIM. BIM can be summed up as the combination of data bases with 3D CAD models, to form geometrical objects that contain all of the information generated during the design, as well as the necessary to construct a building. BIM applications are designed to contain graphical information of the design object in many scales, to parametrically modify its geometry, and to exchange information with other involved professionals in a very efficient way.

Software packages that contain building performance simulation also began to appear. Simulation software range from Finite Element calculations of structures to Computational Fluid Dynamic simulations of natural ventilation of indoor environments. These group of applications is used mostly by specialist consultants to the designers and not by the designers themselves, but it is also a significant addition to the architecture and construction field. The fast and precise modeling of physical phenomena concerning the technical performance of the building being designed is a good thing to have.

Software in the architecture field was a standardized product. Commercial software houses introduced a big amount of functionality in their products with the intention of capturing the attention of the market, but their tools were the same for all architects, all projects and all processes.

In the first decade of the XXI century the architectural community began to take consciousness of the deep impact that the digital revolution had, and architects started to investigate its implications. The "Architecture non-standard" exhibition held in 2004 at the centre Georges Pompidou in Paris, is the manifest of such renewed attention to the relationship between architects and their instruments and techniques. A group of architects started to gain interest in the creation of their own instruments and to become software developers. They began to respond more to their own interests and not let themselves be pushed into any particular tool. The creative efforts of architects found its way into the creation of digital tools with the intention of generating design solutions that were not possible before.

This new generation of designers customized commercial software for their own creative processes. Applications are no longer the same for all architects and all projects, custom processes required custom tools. Customization did not come about for representational problems, custom applications were made for the creative process.

The reliable storage, exchange and manipulation of precise building information is surely useful during the design process. But do these abilities help designers generate better design ideas? Simulation softwares provide performance data as an output to a given building geometry, materials and configuration. But does this information on its own help architects find a good way forward? An argument can be made to say that CAD, BIM and simulation softwares have so far been used mostly for the later stages of design. Little use has ben given to these tools in the conceptual design stage and this is because they are not intended for such use.

The early or conceptual stage of the design process is not so much about information storage, exchange or manipulation, and much more about information gathering. Early design is not helped by the ability to represent in great detail building components and specifications, it is much better assisted by exploration and evaluation.

Search Algorithms represent an opportunity for designers and specialists alike to generate architectural solutions that maximize performance values such as structural or energy efficiency. Building shapes and special features can be explored and high-performing features can be signaled out. By studying high-performing shapes, designers can learn about the relationships between shapes and many performance related disciplines. *Parametric Modeling* in combination with building simulation software and search algorithms allow designers to gather specific information on not just a particular solution, but on entire sets or families of geometric possibilities.

The objective of this PhD research is to investigate how best to use computational search processes in the early phase of design. Search algorithms are implemented in combination with parametric models of different kinds of geometry, with the purpose of studying various building performances.

The early phase of design and the nature of the design problem are studied to get an idea of the kind of search process that would best accompany architects during this phase. Multi-disciplinarity and contrasting objectives are singled out as fundamental required characteristics of such a process. This leads to the proposal of multi-disciplinary studies into architectural shapes, both in the realm of complex curved geometry and in more traditional orthogonal forms.

The search process itself is studied in its capacity to generate solutions in a multi-objective setting. The process of selecting and formulating search problems, the parametrization of geometric families for study as well as the selection criteria for outstanding solutions are all topics of discussion. A particular type of algorithm called Genetic Algorithms is implemented and studied in length. Search processes are proposed for 3 architectural design

fields: structural, acoustic and energy design.

This PhD thesis is divided in **three** parts. The first part is entitled "The concept of Search in Architectural Design" and it is presents the conceptual framework for this research. The second part is called "Search Algorithms for Architectural design" and it presents the mathematical and computational theory involved in the research. The third part is called "Applications of computational Search" and as the name suggests, it presents all of the applications of the concepts explained before.

The **first** part of the thesis takes a look at computational search methods and tools, search algorithms and physical simulation models, all from the point of view of architectural design. There is particular attention centered in looking at historical moments in architecture in which computational or mathematical methods have come into contact with the discipline of architecture. Much has been said about the novelty of digital tools in our field, but there needs to be more information about what is not novel in these methods. This is important if one is to study these methods and try to establish their true impact in architecture, and to try and understand how best to implement them in practice. Many characteristics of these tools have been present in architecture for a long time, from the way we represent geometry, to the very way we explore design possibilities.

The **second** Part looks into the mathematical and computational aspects of this research. The search methods and algorithms used in this PhD thesis are described in detail, discussions on their advantages and known issues are presented as well. The first chapters of this part will look into search algorithms, a general classification, followed by the study of genetic algorithms in more detail.

Multi-objective search will first be discussed in a theoretical way in chapter 7. The concept of contrast is presented together with one of its mathematical representations. We will then look into the algorithms that address multi-objective search, in particular we will study in detail NSGA-II which is the genetic algorithm employed in this research.

The final chapter of this part will look into Parametric Modeling, an example of the parametrization of complex surface geometry is used to explain in detail how to create the model and how to define the possibilities contained in it.

The **third** part of this thesis is devoted to the application of search algorithms in three different disciplines, as well as presenting multi-disciplinary applications. There are four groups of chapters in this part.

The first group presents search algorithms for structural design. This group is made up of chapters 10 and 11. Chapter 10 presents case studies in shell structures, concrete and masonry shells shapes are explored in order to obtain high performance solutions. Chapter 10 is dedicated to load bearing masonry wall structures. This study is mostly concerned with wall openings and thicknesses.

The second group of chapters is reserved for acoustical design of concert spaces. This group is made up of chapters 12, 13, 14 and 15. Chapter 12 is a general introduction to the problem designing spaces intended for the enjoyment of music. A discussion on concert hall design in combination with an introduction to the traditional methods of measurement of acoustical quality. Chapter 13 explains the use of established acoustical parameters for the study of concert auditoria for the search case studies involving them and traditional concert hall types. Chapter 14 presents a study on the use of NURBS geometry in the acoustic simulation of complex shapes. This study was made in preparation to the application presented in chapter 15, this last chapter presents a more innovative acoustical simulation method that is concerned with the early sound in these rooms. In this last chapter, complex free-form shapes are the subject of interest.

The third group of chapters are assigned to the study of energy efficiency in masonry buildings. This group is made up of chapters 16, 17 and 18. Chapter 16 is an introduction into energy efficiency in buildings, the envelope and overall shape of the building, as well as the climates studied in this PhD thesis. Chapter 17 presents a case study on the proportions and orientation of the building, and chapter 18 presents studies on the building envelope.

The fourth group presents the multi-disciplinary search processes developed for this thesis. It is made up of only one chapter describing all of the case studies in multiple disciplines. Two combinations of disciplines are studied, acoustics plus structures and energy plus structures.

Part I

Introduction to Search in Architectural Design

1

Computational search in the early design phase

1.1 The Early Design Phase

It is an intrinsic characteristic of any design process to have phases that have increasing levels of detail. With the evolution of the process, the project needs to be described in more detail and with the appropriate medium. In a traditional architectural design process this is typically represented by a higher scale in drawings, from 1/100, to 1/50, to 1/20 and so on. A higher level of detail means that more and more decisions about the end product need to have been taken. Decisions are usually taken after considering the information available about the consequences, advantages and disadvantages of a particular set of solutions. This information can either come from past experience, studies into the solutions (simulations or expert consultations) or many other sources, but information gathering can be a time consuming and expensive process.

Design problems are almost always approached from a very wide angle in the beginning, the first glance at the universe of solutions has to be a panoramic view in order to consider, explore the highest number of solutions as possible. Designs in this initial phases are usually represented with quick sketches and very little detail is present in these drawings. They go from an *exploratory* phase, to a *development* phase and then to a *definition* phase.

Most design processes are iterative in nature. Steps taken in the process seldom follow a strict sequence, from 1 to 2 to 3. Jumps forwards and backwards on the process are very common. Most importantly for this

research, are the jumps that are taken back to early design. Going back to the exploratory phase happens when designers decide that they need to look further into the universe of all possible alternatives and consider others to the one that they were developing before. These jumps back however, do not mean a complete starting over of the process, because the designer comes back to this point with new information, either information regarding failed attempts and discarded solutions, and/or further information on the definition of the problem itself.

While we can find diverse studies on the design process that seldom agree with each other

Bryan Lawson describes two different "maps"to the design process, the RIBA Architectural Practice and Management Handbook (1965) and the map by Tom Markus and Tom Marver. Lawson points out that from the point of view of the information produced or the output of these maps, they all show a pattern of increasing levels of detail. For example, Briefing, Sketch plans, Working drawings and Site operations in the RIBA map, and outline proposals, scheme design and detail design in the Markus/Marver map. Designers may chose to start with a general view or start by details (selecting materials for example) but the general distinction between the phases remains (Lawson 2006).

In his 1976 article Boyd Paulson was perhaps the first to describe the relationship between design stages, the level of influence in the design and the cumulative cost of the project. He published a diagram (see figure 1.1) of this relationship that depicts a couple of curves describing the inverse relationship, in which, the further along down the design process the lower the level of influence on the project and the higher the expenditures of the project are (Paulson 1976).

Thereafter came many versions of this graph, perhaps the most known version is the MacLeamy curve, but the main concept remains the same. This relationship is a fundamental issue on the importance of the early design stages, and the need for good information and good decisions in this phase.

The early design phase is the moment where the big decisions on the building shape, and many of its defining components, such as structure and distribution are decided.

Building design involves many different technical disciplines apart from architecture. From structural design, mechanical, electrical and hydraulic systems, acoustics, lighting to energy efficiency. Every discipline studies very specific aspects of the building, all looking a different physical phenomena, and evaluating performance values related to user comfort, efficiency and

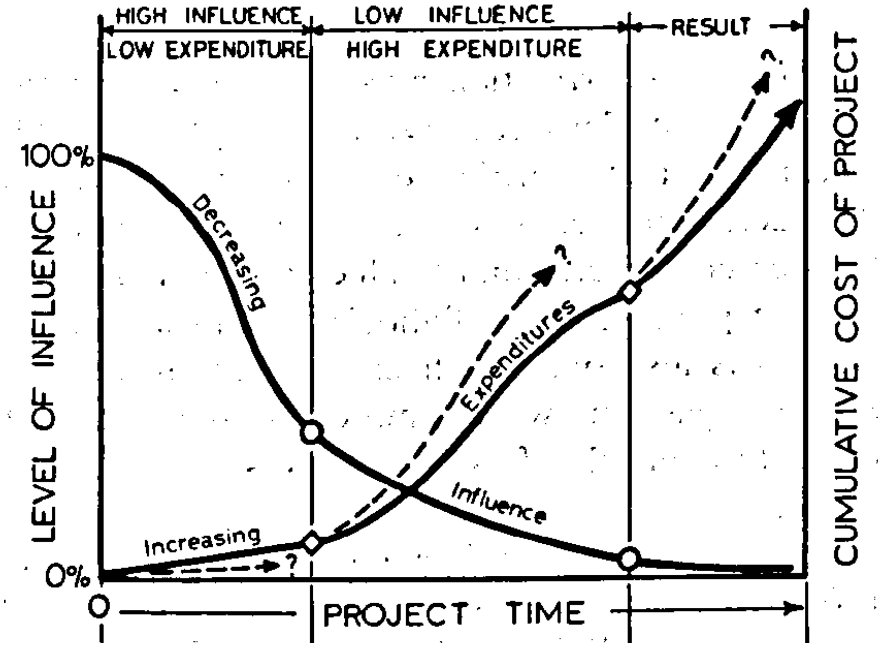


Figure 1.1: Boyd Paulson's curve (Paulson 1976)

quality of the building. All of these disciplines look at specific parts of the building, but most of these parts have functions that involve many disciplines. In fact, "Very rarely does any part of a designed thing serve only one purpose"(Lawson 2006). The façade is studied by lighting, thermal, ventilation and acoustic experts, all looking to improve the design from their point of view. Every component of the building is involved in a multi-disciplinary design process. This is even more evident if we talk about the overall form of the building, its orientation, size and shape.

The early phase of design is the moment when the most disciplines *ought* to be involved, when multi-disciplinary information is most needed.

In such a multi-disciplinary framework it is common to have situations in which the best results in one discipline are achieved by designs that do not correspond to the best result in another one. This issue is clearly described by Christopher Alexander in the introduction to his 1966 book. He describes

an example of a design problem showing contrasting objectives*. He talks about the choice of materials for a household appliance:

> "Time and motion studies show that the fewer kinds of materials there are, the more efficient factory assembly is - and therefore demand a certain simplicity in the variety of materials used. This need for simplicity conflicts with the fact that the form will function better if we choose the best material for each separate purpose separately. But then, on the other hand, functional diversity of materials makes for expensive and complicated joints between components, which is liable to make maintenance less easy. Further still, all three issues, simplicity, performance and jointing, are at odds with our to minimize the cost of materials. For if we choose the cheapest material for each separate task, we shall not necessarily have simplicity, nor optimum performance, nor materials which can be cleanly jointed."

(Alexander 1966)

He accompanies that statement with the diagram shown in figure 1.2. The diagram denotes the four objectives described in the text: performance, simplicity, jointing and economy. The plus or minis signs on the lines connecting the nodes or objectives, are there to signal if the relationship between this objectives is a positive or a negative one. A negative relationship signifies a conflict or contrast of interest between these objectives. In this simple example, alexander only depicts the relationship between simplicity and joining as a non contrasting objective.

These conflicts are very common and problematic in building design. This is especially true in the early design stages when so many different variables and disciplines are involved. When such a large part of the design is still to be decided, it is clear that conflicts have not yet been confronted and solved. This is an intrinsic aspect of this stage of design, and it must be given proper attention if we are to address design tools for the early phases of architectural design.

The resolution of these conflicts can be harder or easier depending on the case. As we will see later in the experimental part of this research, some contrasts are still manageable in the sense that a good trade-off between

*A mathematical description of contrasting objectives by means of Pareto Fronts is explained in section 7.6

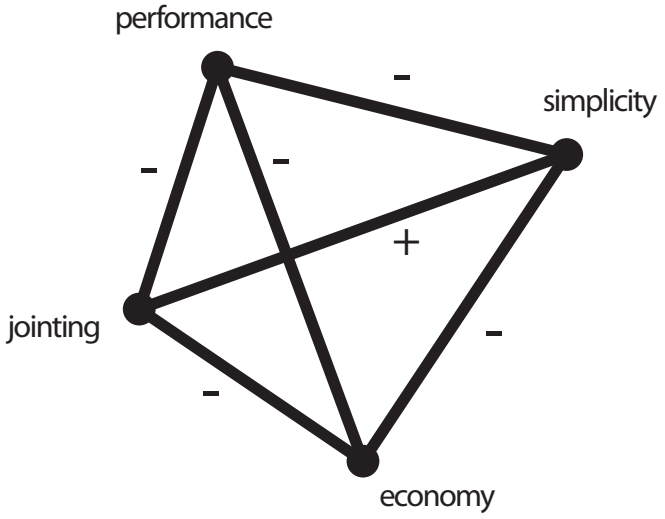


Figure 1.2: Christopher Alexander - Contrasting Objectives Diagram (Alexander 1966)

goals can be found. But other conflicts are harder to negotiate, and in these cases a final decision can only be taken considering more information.

We will refer to these conflicts in design goals as *contrasting objectives*.

If we take another look at Boyd Paulson's curve (figure 1.1) we can see that the potential for design improvement early on is huge, and that in the later stages of design changes are small and limited in scope.

Traditionally, bigger efforts are left for the final stages of the design process, when the technical issues of the project are usually taken into account. Optimization procedures are usually used by a specialist in the field of the particular issue (e.g. a structural efficiency, cost, construction time). However, in the final stages of design the main decisions in most different fields have already been made. When this is the case, it is often too late to make any significant changes in the design and the optimization process serves only as a final definition aid. In the final stages of design the field of possible solutions is very small, when compared to the possibilities in the early design stage. In the early stages of design, the design space is much larger

and so is the amount of information considered in this stage. When we refer to such a final definition process, the correct term is optimization.

The rational and systematic exploration of the space of feasible solutions during the early design stage considering multiple disciplines and contrasting objectives is what we define as a *search* process.

1.2 An Introduction to Search

Hutchison et al. give us the key points to define Search:

- A goal, an objective for the search.
- An Uncertainty about goal location. There can be no search if the goal is directly or easily attainable. Normally, as the search process evolves the this uncertainty tends to diminish.
- The adaptive varying of one's position.
- A stoping rule.

(Hutchinson et al. 2012)

Taking this four points into account we can sketch a very brief definition of Search[†]. Search is the operation necessary to achieve a certain goal when we do not know how to achieve it, by adapting and considering several positions or views about the goal until we either achieve the goal or we stop for another reason.

> "Search - the behavior of seeking resources or goals under conditions of uncertainty- is a common and crucial behavior for most organisms. It requires individuals to achieve and adaptive trade-off between exploration for new resources distributed in space or time and exploitation of those resources once they are found"

(Todd et al. 2012*a*)

The resources we are talking about in search can be varied, but in the present thesis we will be referring to information, either external or internal

[†]A computational and mathematical approach to search is discussed in chapter 5.

information. Most cooomonly this is building performance information (e.g. acoustical quality of a room, energy efficiency of a building envelope). As such the search space is the universe of possible solutions to a given problem, we search for these solutions either from external sources or from our memory of past problems and solutions (internal sources)[‡].

Exploration and Exploitation . An important component in the above definition is the trade-off between exploration and exploitation:

> "Finding a resource typically involves at least two components: an *exploration phase* that investigates possible locations as to where the resource might be located and an *exploitation phase* that involves resource acquisition. Often, the exploration and exploitation phases are not mutually exclusive, as animals often sample and exploit during exploration and continue exploring while exploiting.
>
> Because exploration typically takes away time from exploitation, modulation between the two can be represented as an optimal control problem in which organisms attempt to minimize the time spent exploring for resources but still acquire sufficient information to maximize the resource exploitation. . . More exploration can lead to finding better resources but less time available for exploiting those resources. This trade-off between exploration and exploitation is common to both eternal and internal search problems. "

(Hills & Dukas 2012)

Lawson perhaps characterizes the same exploration vs. exploitation relationship while talking about analysis and synthesis:

> "Analysis as the exploration of relationships, looking in the information available and the classification of objectives. Analysis is the ordering and structuring of the problem. Synthesis on the other hand is characterized by and attempt to move forward and create a response to the problem - the generation of solutions."

[‡]see chapter 3 for a discussion on internal and external search.

(Lawson 2006)

Computation became a very common search method for a large variety of goals, search spaces and disciplines. The power and speed of computation makes for a great tool in any search process, but especially those involving information search, as those that concern the present thesis. The architectural design process is a much more complicated process, one that involves many other tasks apart from information gathering, never the less, a very important aspect of designing is gathering the right information at the right time. This can significantly increase the quality of the end product of the design process. Computational search for the architectural design process is the subject of this thesis, we will look at the specificity of architectural design regarding search, information and the usefulness of this information.

1.3 The "Wicked" problem

In the 1950's, and particularly with a series of conferences in the 60s the field of Design research, or Design methodology was born. A series of books were published by the fields "founding fathers", such as Asimow's "Introduction to Design" in 1962, Alexander's Notes on the synthesis of Form of 1964 and Jone's "Design Methods" of 1970. These seminal works would later be called by Horst Rittel a "first generation" of design methods that had been a necessary but simplistic start to the field. The second generation started with his work on defining what a design or planning problem is, and how it differs from other more scientific problems. In their 1973 paper "Dilemmas in General Theory of Planning", Rittel and Weber in 1973 give us a clear idea of the nature of the design and planning problem:

> " The kinds of problems that planners deal with–societal problems–are inherently different from the problems that scientists and perhaps some classes of engineers deal with. Planning problems are inherently wicked...
>
> ...The problems that scientists and engineers have usually focused upon are mostly "tame" or "benign" ones. As an example, consider a problem of mathematics, such as solving an equation; or the task of an organic chemist in analyzing the structure of some unknown compound; or that of the chess player attempting to accomplish checkmate in five moves. For each the mission is clear. It is clear, in turn, whether or not the problems have

been solved. Wicked problems, in contrast, have neither of these clarifying traits".

(Rittel & Webber 1973)

Rittel and Weber go on to characterize the "wicked" problem:

There is no definitive formulation of a wicked problem. While a "tame" problem can have an unequivocal formulation describing it fully, "wicked" problems are formulated differently by different people, depending upon their understanding of the problem.

Wicked problems have no stopping rule. A mathematical problem can be easily said to be solved, but "wicked" can be worked on or improved almost indefinitely. If we think back to the previous proposition (that "wicked" problems have no definitive formulation) how can we then say when the problem has been solved. Formulation and solution of a problem go hand in hand. Rittel explains that in most cases designers and planners stop working on "wicked" problems not because the problem was considered to be solved in a satisfactory way, but because of other reasons external to the problem. Most commonly designers stop working when they run out of time or money.

Solutions to wicked problems are not true-or-false, but good-or-bad. Scientific or "tame" problems are either solved correctly or incorrectly, theorems are either proven or disproven unequivocally. Solutions to Design or Planning problems can be said to be better or worse from one another, they can be studied and rated, but they are not true or false.

Every solution to a wicked problem is a one-shot operation because there is no opportunity to learn by trial-and-error, every attempt counts significantly. Solutions generally cannot be undone without major investments of time and resources.

Every wicked problem is essentially unique. While "wicked" problems can have similarities and share solution approaches, they are always specific characteristics (for example social, cultural or environmental context, budgets, etc.) that make them all different from one another.

Architectural Design however cannot be reduced to a series of problem solving tasks. As we have seen by the work of Rittel, design problems have intrinsic characteristics that separate them from tame problems. Omer Akin the goes to describe how these problem solving states apply to the ill-defined or wicked problem:

"Types of representations and transformations possible in

well-defined problems are known a priori. Redefinition of these ground rules is not necessary and not allowed. In contrast to this, in design, discovery of new rules is desirable, even though a large set of conventions is available as part of the culture of design. Creative design solutions are often linked to the redefinition of conventional interpretations of design, and creativity is a ubiquitous goal for the designer.

Goal states of design problems are usually inadequately specified at the onset. There are no explicit evaluation functions that can be applied to a state that will result in the unequivocal identification of it as a solution state. Each designer applies his or her own specialized tests to determine wether or not a design is acceptable."

(Akin 1986)

When goal states are not known a priori, fully automated processes are not possible. Automated processes require knowledge on all possible outcomes and directions the automation could or should take. Pre-defined procedures are outlined for all possible scenarios in an automated process. So when we say that in a design process we have not only unexpected results but unexpected goal states, it is clear that design processes cannot be automated. Search processes and other design related tasks can be well formulated and defined for automation. Interaction before, during and after search processes is also an important part of a well defines search process.

1.4 Automation and Design Control

So far we have outlined the concept of an automated computational search process for the early design phase of architectural design. Like with any other process involving automation, the issue of control needs to be considered. Who is in control of the process before, during and after automated procedures is a question to be taken into consideration.

Automation in our daily lives is seen as any process that is done without direct human involvement. This is especially true now that computers are a part of almost any everyday object like a telephone or a car. An automatic car shifts between gears without the driver shifting the gear box with his hands, an automatic cat feeder serves food to a cat without its owner doing anything. But in reality all of these automated procedures were determined

by humans, the car builders established a rule that decides when the car shifts its gears, and the feeding machine is programed by the cat owner to feed the animal three times a day or so. Humans created these processes to do things for them, they defined what the automation does, when it does it and when it does not.

Some automated procedures even leave some decisions up to the system itself. Domotic systems for example decide weather or not to turn light on or off in a space, depending on the presence of users and availability of sunlight. They also heat or cool a space depending on its temperature, as well as various energy saving conditions that the end-user is normally not even aware of. But even in these cases the automated process is carried out only within the confines of what the systems designer defined for all situations.

Automated processes do not presume the absence of rules or definition, quite the opposite. Automated processes follow strict order and procedures defined by the processes author. Computer programs do not define rules or procedures in any scenario, programers do. We can therefore safely say, that any sort of automated process takes place during design, it takes place within the confines of what the designer deemed necessary or desirable.

Search for the early phase of design, as intended in this research, can be characterized as automated processes where the designer is completely involved in the definition of the rules and procedures involved, and where authoriality of the design object is not in question. Designers are also involved in the definition of the representation of the design object during the automated procedure.

1.5 Automation and Representation

Architects do not design by working on the building itself: they create and use different methods of representation and notation. Architectural design processes require representation methods, most commonly based on geometry. Geometry is used to represent spaces and building components. Geometry is expressed through different techniques (e.g. plans, sections, orthogonal projections or 3D models) that allow designers to visualize, edit and communicate their work. Geometrical representation is enriched by shared notations and graphical conventions, perhaps most notably for the purposes of communicating to builders all of the necessary details of the work the must create, the construction documents.

The geometrical representation of space, suitable to being directly trans-

lated into a numerical and algebraic description, is then the starting point for the automated search of design solutions by means of computation for architectural design. A complete description of the design object must be provided for its study. More specifically to search processes, a representation method capable of describing not just one design object but many is required. In fact, an implicit representation must be provided for *all* of the design possibilities that are intended for study.

The representation an entire set of design solutions implies the recognition of common characteristics to all of instances, the features that make each solution part of the desired set related to all other solutions. These common features can be called the *invariants* of the solution set. All solutions in a set are to be similar but not identical and they must share some features and they must be different in some others. Solutions will therefore be described by their common or invariant features and their variable features. We will refer to the numerical quantification of these variable features as design*parameters*.

A parallelepiped for example, is defined by having 6 faces, 12 edges and 8 vertices, and by having straight angles between all adjacent edges and faces. This topology is what makes a parallelepiped, no other characteristics are required for it to be a parallelepiped, it will remain so no matter what is its volume, length or height. The topology of the parallelepiped is its invariant, it defines it as such. The definition of its exact shape is finalized by the use of a combination of dimensions such as length and height or volume and either length or height. These dimensions are the parallelepiped's parameters.

A sphere is defined as a surface that is at any point equidistant to its center, it is not defined by any specific radius or volume, in the same way that all tetrahedron are conformed by 4 triangular faces. These definitions of platonic volumes are well known and serve as clear examples of invariants and parameters, but any topological relationship for any geometrical shape can be represented in terms of invariants and parameters. Dimensions of elements of any topology are common examples, but even topology can be parametrized.

We will refer to this kind of representation defined by invariants and parameters as *parametric modeling*.

Parameters are typically confined to domains of variation, they are only allowed to take values inside a used defined domain. When we use parametric models for search processes, parameters and domains determine the extents of the search space. The number of parameters will determine the dimensionality of the search space, while the domain's shape and range will determine its extension. This is especially important in establishing control

in automated design processes. The procedure mainly controlled trough the definition of invariants and parameters, as well as assessing the parameter domains.

Designers use the concept of *scale* in their design process: larger scales are used to determine overall shapes during the early design phase, while smaller scales focus on details on more advanced design stages. The selection of parameter domains is also subject to the same criteria, domains can be large when general building shapes are studied during a search process, and they can also be small when a more detailed study is necessary during an optimization process.

In automated search, the use of scale implies the concept of *resolution*. Parameter values are chosen inside a user defined domain, but this domain is not necessarily continuous. Domains are usually subdivided into discrete intervals, thus reducing the number of possible values and solutions belonging to the search space. A higher resolution implies smaller intervals in each domain, and thus a higher number of possible solutions, while a lower resolution implies larger intervals and lower number of possibilities. The concept of resolution can be used in a similar way as the concept of scale with the purpose of setting the level of detail in a search process.

1.6 Limitations and opportunities of a computational approach

1.6.1 Computational Performance Simulation

The mathematical description of physical phenomena was developed throughout the centuries, but its large scale application and diffusion was only made possible with the invention of the computer. In the field of mechanics, the well known Finite Element Method (FEM) dates back to the work of Ritz in 1902, but we had to wait until the 1960's to see its dramatic increase in application. Models have been developed to determine the thermal exchanges in the building envelope, the wind resistance of a skyscraper, sun radiation in building facades and roofs or entire city sections, and even the air flow patterns of a sterile operation room. All of these models give designers various performance measures that describe the physical environment they are designing. They all vary in complexity, accuracy, model uncertainty and time consumption.

Performance simulation is a very powerful tool, but *on their own* they are not of too much use to designers, especially in the early design phase.

In most cases, the definition of the building needed to run a simulation is too elaborate for the concept stages. Many decisions have to already have been taken in order to provide the software with sufficient information to operate. This means that the information required to simulate building performance is perceived as too labor intensive for designers to produce in this stage. Commonly they move on from early stages without such information, and only use simulation later on. The output that simulation software provides is not necessarily of any help, especially when the designer cannot afford to consult a specialist for the multiple design solutions that are usually considered in early design. Simulations results are often incapable of helping the designer choose the right path, or make more informed decisions. This scenario often means that in projects with a high technical requirements, for economic and time constraints, the designer is forced to consider very few alternatives, to shorten the early design phase, and to try and make the best of it in later stages of design. In order to introduce computational search in the early design stages, a different approach to performance simulation and data handling is required.

In this PhD research the use of structural, acoustic and energy simulation is discussed and implemented in several case studied. The Finite Element Method (FEM) is used to accurately predict the strains, forces, moments and displacements in building structures of various shapes, materials and loading cases[§]. The acoustic field inside a room is modeled[¶] in an approximate way trough one of the most diffused methods, the so called the raytracing method. Given a sound source, the method can predict the acoustic response in any given point of the room, considering sound reflection, absorption and scattering. Figure 1.3 shows an example of an acoustic study employing raytracing simulation of four different rooms. Energy requirements for heating, cooling and lighting for entire buildings cal also be estimated by the use of dynamic thermal simulation and lighting raytracing software packages[‖].

Simulation accuracy is very much dependent on the accuracy of the data given to the model. In physical problems described by relatively simple models the effect of input data errors on the output can be controlled. However, when the problem involves different physical models, interacting with

[§]A more detailed description of the use FEM software in this research is given in Chapters 10 and 11.

[¶]A more detailed description of the use Acoustic Simulation software in this research is given in Chapters **??** and 15.

[‖]A more detailed description of the use Energy Requirements software in this research is given in Chapter 16.

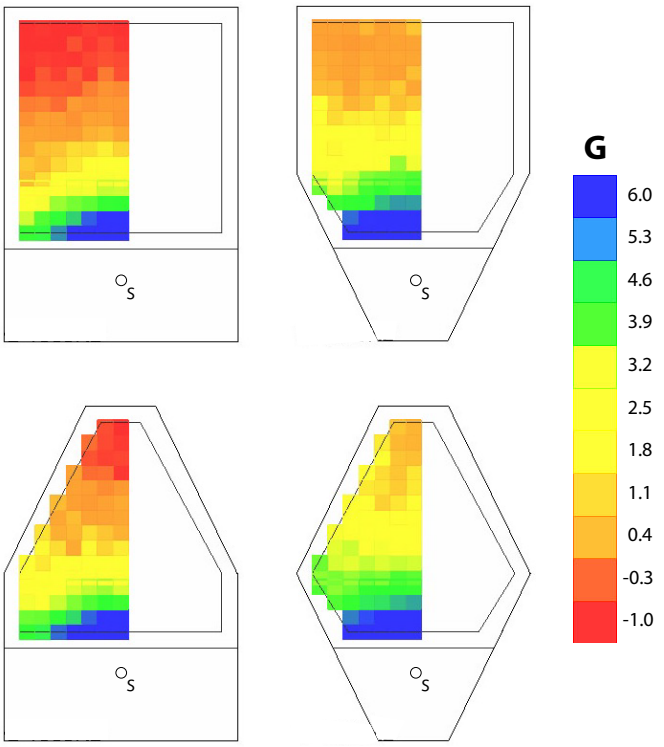


Figure 1.3: Acoustic Simulation Example. Distribution of the Sound Strength Parameter G in four different rooms.

one another in a multi-physics approach, or when non-physical aspects has to be considered (e.g. economic, social, subjective preference), the problem becomes very complex, and the influence of input error on output can increase dramatically. If the data used in the simulation is not correct the results will be completely random or misleading at best. If the phenomena that the designer is trying to investigate are not properly modeled by the simulator employed, the results are equally useless. Designers really need to understand the physical phenomena that are being simulated, the variables involved and how to interpret the results.

1.6.2 Software Customization

Commercial software houses provide an increasing amount of functionality specifically developed for Architectural design. Computer Aided Design (CAD) software has been present even in small architectural practices since the mid 90's. Perhaps the most talked about software tool in the recent years in the architecture and construction community is Building Information Modeling (BIM). BIM can be defined as the use of 3D CAD models in combination with data bases containing drawings, costs, and various characteristics of the building components described in the 3D model. BIM is however a tool that is more used in, and perhaps its better suited for, the design development phase. All of these tools are very successful and are widely used in design studios of all sizes.

Software houses are big companies that are in competition for the architectural software market. They make choices for the development and commercialization of software that are in their best interest, and their interests do not necessarily align with the architect's interests. Software houses are an external entity to the design studio, and their products cannot follow the requirements of each architect or each project.

Large Architectural design firms employ in house software development teams, in charge of creating or customizing software for the firm's projects. Example of these teams include the Specialist Modeling Group at Foster and Partners, the R&D team at Aedas or the Digital Technology Group at Herzog & de Meuron. Some of these teams have pushed the standard for design software all across the industry by collaborating with commercial software houses. However, this has mostly been a reality in very large firms working in large and complex projects with big budgets, but as the commercial software houses are beginning to see the need for software customization and user development, they are starting to create programing environments that are intended to help the architect do just that.

A significant role in this movement has been played by single users-customizers and online communities of that create and exchange customized applications, plugins and scripts that are in turn edited and used by other members of the community. Some of these applications are developed following an Open Source approach. These individuals creatively employed existing computational tools and geometric functions to create customized applications that went far beyond what the software houses were offering. These companies in turn realized the potential of customization and started to include more and more customization capabilities in their products. A significant example of this can be seen in the McNeel's Rhinoceros, the incorporation of the "rhino scripting" environment, and eventually the grasshopper graphical programing environment.

There is an important relationship between the work done by software houses and functionality. The more work software developers do, the less specific or custom functionality is left for the architect. The more work is done by the user-customizer, the more custom functionality he will have. Figure 1.4 shows a diagram of the developer user relationship discussed.

This principle applies not only to architectural software. Considering an example form a different industry, a smartphone application that tells you what the weather is going to be tomorrow just by talking to the phone and asking it verbally, in this case the user needs to make very little effort to get the application to function properly and achieve the desired result. The developer on the other hand needs to make a very big effort. He needs to develop voice recognition software, access weather services and display the output in a meaningful way. We can see this type of user-developer relationship in the center triangle in figure 1.4, we can see that the amount of functionality that the user gets access to is very limited, he can only get to know the weather forecast.

The relationship outlined by the third triangle in figure 1.4 can be seen in the case of the customization of the Catia software by a group of architects in Frank Gehry's office for the Guggenheim Bilbao museum. They took the NURBS functions present in the commercial software and developed tools on top of them to adequately represent all of the complex components of the buildings titanium cladding, among other elements.

When the user becomes a software developer himself he can determine the exact functionality he requires, and expand it as needed. Architects in this category usually use a great deal of available applications and customize them for their own use, often on a project by project basis. This requires not only some programming skills on the part of the architect, but also a good amount of time and effort. When combined with the efforts made by

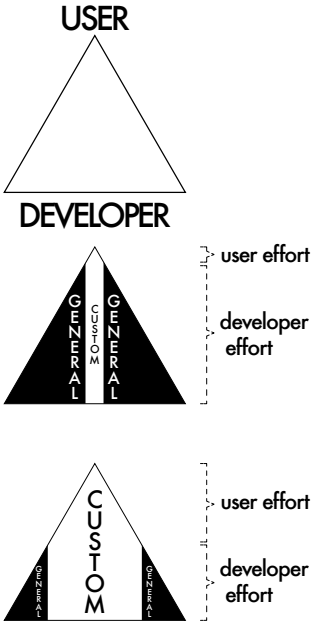


Figure 1.4: User - Developer and Custom use - General use software relationship

software developers, the architects efforts are rewarded by the availability of larger and project specific functionality thats was otherwise unavailable.

1.6.3 Software encapsulating knowledge

Design instruments such as drawing aids and geometrical calculations have always been a part of architectural design practices. They were traditionally close to the architectural profession, being created and employed by architects. The relatively recent arrival of software instruments such as CAD represented a separation between architects and their design instruments. Instruments started to come from other industries, and software tools were made in such a way as to not allow architects to make them their own, they became external to the discipline.

Andrew Witt talks about the inherent (and sometimes unintended) con-

sequences of software development for architectural design from an epistemological point of view. He starts from the distinction between design knowledge and instrumental knowledge in a form that could be applied to the traditional work of the architect.

> "*Design knowledge* is an intrinsic understanding by the architect of formal organization principles such as the relationship of parts to whole, and interrelationships of program constraints, spatial organization, ranges of material effects, and use of geometric methods. It may include disparate and heterogeneous organizational schemes and diagrams. These general principles may be redeployed in various contexts, and need not be tied to particular working methods or automatic tools. In this sense, geometric knowledge is a particular kind of design knowledge: although it may be deductive and procedural, it is not automatic and its application requires a synthetic understanding of design constraints. *Design knowledge* is the most enduring epistemic content of architecture as a discipline, sometimes even hastily equated with architectural knowledge itself.
>
> *Instrumental knowledge* is a more narrow understanding of the procedures to successfully operate a certain type of technology, which would include ability to operate a software, program, script, process, tool, instrument, or machine to intended effect. This is in contrast to the way the term "instrumental knowledge" is used in the epistemology of science, for example: as a description of theories of predictive reliability (and thus instrumentality). Instead, in our sense instrumental knowledge is in fact an intentional knowledge of instrument operation. Instrumental knowledge also enables the creation of systems of interrelated technologies intended to facilitate the aims of design. More generally, instrumental knowledge can include the ability to abstract the inverse constraints of these machines onto design with the aim of pre-rationalizing the design itself, as in the case of drawing machines, fabrication machines, or construction machines. This instrumental knowledge is powerful because it makes procedures encapsulated by the technology in question easily accessible, communicable, repeatable, hackable, and transformable."

(Witt 2010)

In his article "A Machine Epistemology in Architecture. Encapsulated Knowledge and the Instrumentation of Design" he draws a parallel between the geometrical representation machines of the XIX century, such as ellipsographs, and our current use of computational tools for the design, representation and production of buildings. In so doing, he argues that design knowledge has always been dependent of instrumental knowledge, despite the architect's relative disdain and ignorance in the latter:

> "The pervasive use of digital technology in the conception and execution of buildings dramatically increases our reliance on representational and operational systems of which we have incomplete understanding but that we nevertheless trust implicitly... The machine, particularly the computer, calls into question the self-understanding of architecture and its self-imposed alienation from technical processes. There is a strong tendency, arguably beginning with Alberti, to dichotomize design knowledge and instrumental knowledge, and to relegate technical or mechanical expertise to the domain of specialists or operators. Perhaps this can be explained by a mistrust of the architects need to rely on mechanical, electrical, computational, or conceptual operations of which the architect cannot have complete understanding. This trust in machines, however, far from being an innocent conceit, represents an implicit belief in the possibility that collective memory and design knowledge can be instrumentally encapsulated in machines. It represents not a barrier to advancement of architectural knowledge but a great opportunity."

(Witt 2010)

Witt's enthusiasm for the advancement of design knowledge through encapsulated instrumental knowledge could reach the design community in a stronger way by emphasizing the fact that when the architects are the impulse and creators of instrumental knowledge, the relationship between design and instrument is stronger. Architects should not completely conceit defeat in the generation of instrumental knowledge, but be at its forefront. This remark can be understood in the context of the user development relationship outlined above and in figure 1.4. The more the architect takes part in the creation of the software instruments he uses, the more will he profit from its functionality, and the better are the chances for him to make advancements in design knowledge.

Most contemporary software instruments used by architects come from distant fields. From computer graphics and animation to civil and aeronautical engineering, many distant disciplines have contributed to the architect's computational toolbox. This would explain the separation between design and instrumental knowledge Witt is describing. But the customization of software is an important instrument being embraced by the architecture community, and represents an opportunity to regain instrumental knowledge in the hopes of further advance our discipline.

In this process of instrument creation and encapsulation by architects, the dichotomy described above could tend to disappear, design and instrumental knowledge become almost indistinguishable in architectural practice. We can certainly say that design knowledge can be encapsulated in software or machines, facilitating their use. But we can also say that the use of encapsulated knowledge and design software tools such as those proposed in this PhD can further generate design knowledge. Information obtained during search processes, if carefully studied, represents an advancement on design knowledge.

1.7 Interactivity and Search

> "Take an optimization model. Here the inputs needed include the definition of the solution space, the system of constraints, and the performance measure as a function of the planning and contextual variables. But setting up and constraining the solution space and constructing the measure of performance is the wicked part of the problem. Very likely it is more essential than the remaining steps of searching for a solution which is optimal relative to the measure of performance and the constraint system".

(Rittel & Webber 1973)

In this 1973 quote, Horst Rittel warns us on the risks of using optimization (or search methods) for design problems. The wicked part of design and planning problems is very often not solvable by means of computation. Many aspects of design problems can be addressed by performance measures, but many cannot. Designers are therefore required to to ask the right question in such search processes, to formulate the problem correctly. When formulation of search processes is done successfully, computation can pro-

duce information, performance data and even propose favorable solutions to the formulated problem. The input from the designer throughout the automated process is fundamental.

The most important moment of interaction between the designer and the search process is of course the formulation of the problem: the definition of the parametric model (invariants and parameters), the selection of the performance criteria and the simulation method. It is in this moment when the designer has the most control over the search process and consequently when he can make the biggest mistakes.

There are other moments when the designer can interact with the search process. Christian Derix talks about this issue and the importance of interaction in his 2010 paper "Mediating spatial phenomena through computational heuristics":

> "Wicked problems like layout or urban design require the experience of designers to negotiate the many explicit and implicit aspects that can be represented through computation. Particularly, when design aspects are not discursive and the amount of data is large, the key organizing principle of designers and design teams are their learned heuristics, not performance indicators and data sets. While computation shouldnt imitate analogue heuristics, it can express its own search mechanism via visualization of processing steps. If a designer can interfere with computational heuristics and observe the search struggle, the opportunity identification between designers analogue and computational heuristics are given. This enables the validation for wicked problems when no explicit goals are set."

(Derix 2010)

The information produced by and during the search process can be viewed by designers and involve them directly, by interacting and modifying the process itself. There seems to be a limitation in the participation of the designer with regular search methods, that push their use to later design phases. On the other hand, interaction, the ability to provide input in real time, introduces a real participation that is essential for the early design phase. The ability to influence the result by means of interaction, is evidence of design, not optimization. Search processes are design oriented, not just performance oriented. Not all requirements or design ambitions are introduced as performance criteria, this is left to the discretion of the

designer. This does not mean that those design requirements and intentions are not a part of the search process, by means of interaction the designer is actively introducing them in the process. The search process would then have explicit and implicit design goals. From a performance point of view, the final result may therefore not be an "global optimal" solution, but it will be what the designer wants, a better informed design.

The opportunities of interaction in search algorithms are different depending on the algorithm itself. But the general idea is that the very formulation of the algorithm (or parts of it) are modified during the execution, responding to the designers evolving intentions. As the search process evolves, and the knowledge acquired by the decision maker is increased, implicit or explicit design goals can vary and the search process can be steered on a new direction. The most obvious way of doing this is by changing the parametric model itself, either by adding/subtracting variables, or by changing their domains. Changing parameters or domains, without modifying the design goal of the search process has no negative effect on the comparison of the resulting performance values. While the parametric model that generated the solutions changed, the performance criteria remained untouched, hence we can still compare solutions before and after the interactive modification.

Another way of doing this is by changing the design goal itself. It can be slightly modified to better suit the problem, or to try and manipulate the solutions being generated. This operation is perhaps counter intuitive when compared to the one described above. It has the disadvantage of producing results that are not comparable to each other. Results with different design goals cannot be compared on terms of their performance values.

The data produced during these interactions can be stored in a "search tree". A sort of Search history of the decisions made during the process, the interactions, modifications and performance values. Figure 1.5 shows a diagram of a search tree, in which the designer after 4 search iterations was presented with solutions a, b and c. The designer selected solution a (signaled by the letter A in caps) and after a few more iterations solutions d and e were considered, solution E was selected, and so on. After 13 iterations the designer ended the search process. As the diagram suggests, the data pertaining to all solutions and iterations is saved in the computer, enabling the designer to restart the process at any point, or to consult the data later on.

One of the main concepts that allow architects to engage the whole scope of the project in the early design phases is the concept of scale. Designers often use larger scales in these phases so that they can concentrate on the big picture, and make faster decisions on the vast majority of the building's

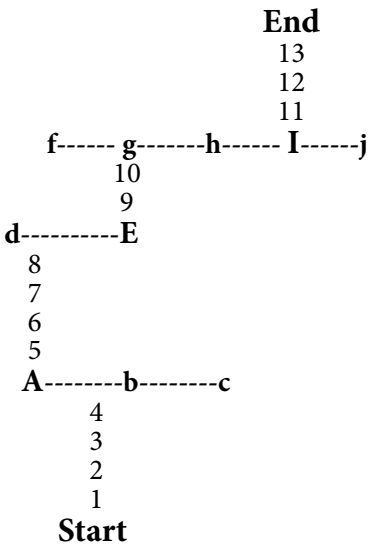


Figure 1.5: Search Tree diagram

form. Slow interactions with simulation models and search algorithms, mean that the designers need to leave these tools for later design. Calculation times are important when it comes to real-time interaction.

As previously introduced, the concepts of scale and resolution can be applied to Performance based Search in order to reduce calculation times by limiting the size of the search space and facilitate interaction. Like architects often do, search methods can first work with low resolutions in their parameter domains. This means discretizing the search space in a coarse way, effectively reducing the number of calculations needed to provide the designer with an broad idea of where the search process is leading him. This means that the designer would be utilizing broad strokes in the beginning and progressively improving the resolution as he focuses on specific areas, performances, or geometries. This can be an effective way to engage the designer and improve interaction.

An alternative way of interacting with the process is one that takes place

after the calculation is done, a sort of data mining of the search output. This mode of interaction is applicable when large search spaces are involved, resulting in large data sets. Interaction in this case cannot modify future results since all calculation is done, but it can be the starting point of a new search process involving new information, a new iteration with more information that can help better formulate a new process.

The Early Design Phase is about immediate restitution, real time interaction, the way a pencil gives an immediate result. Computational search is not immediate, depending on a variety of issues, a search process can take minutes, hours, even days. Interactions during the design search imply that the designer is constantly involved in observing and modifying the process. Thus the process should be producing feedback constantly, in real time if possible. This is not always possible.

2

Algorithms and Parameters in Architectural representation

Architects create and use different representation methods during the design process, they conceive, edit and study their buildings through some form of representation. Representation is also used to transmit design information to builders. Different media have been used by architects throughout history: The spoken word, the written word, the handmade drawing, the printed drawing, the scale model, the digital file. Some media are inherently more inductive towards one representation method over another.

Architectural representation has been addressed throughout history by architects in treatises and manifestos, It has been the subject of study in this discipline for along time. Andrew Witt considers it design knowledge and not instrumental knowledge, which is not surprising, since representation has always been considered a part of the architect's fundamental techniques.

The type of architectural representation described in this PhD thesis is hardly new, parametric models, as well as their iterative nature and relationship with the design process can be traced back to previous centuries. An understanding of the roots of these methods of representation seems to be an important step in relating existing design practices with automated processes such as search.

2.1 The roots of Parametric representation

Architectural scaled Models have been present since the very beginning of western architecture. There is evidence to suggest that wax and wooden

models were used by greek architects. there is some debate as to the exact purpose of these models, if they were for exhibition, survey, building or design purposes. It is also unclear the role they had in roman times, but it seems clear that for Vitruvius they played no part in the ideation of the building.(Scolari 2012, Carpo 2011)

The purpose of making models in architectural design is perhaps clearer in Alberti's treatise *De re aedificatoria*. Massimo Scolari describes Alberti's ideas for the use of Models:

> "His model is an instrument of experimentation and reflection with which to ascertain the buildings's structural stability, its orientation, the layout of the main walls, and the adequacy of its roofing. It is used to try out the most likely solutions to each single problem and to make a precise calculation of the costs of the work... the model should be "nudos et semplices", crafted from simple materials so that it is the architect's true conception that emerges, rather than the skill of the model maker."

(Scolari 2012)

The words "Experimentation and reflection" denote the fact that, for Alberti, models were very much a part of the design process and were not involved in the construction process. The model was not entrusted with a notational aspect as other means of representation. In Alberti's writings, orthogonal Drawings were the preferred method of representation when it came to construction. The iterative nature of design and design representation is also present on renaissance model making, Massimo Scolari quotes Filippo Baldinucci from his "Vocabolario toscano dell'arte del disegno" from 1681:

> "The first and most important task in the making of the work is the model, since it is by means of trying out his ideas and *altering* them that the Artificer arrives the most beautiful, most perfect solution. "

The fact that both buildings and scaled models are tridimensional physical objects has always been a great advantage for the model, in the sense that the information is conveyed more directly, a tridimensional object is represented by a tridimensional model. In contrast with orthogonal drawings, models present themselves as direct scaled copies of the building. The

only abstraction left to the viewer is the scale and size of the object in relation to human scale.

It was not until very recently that scaled models were given notational importance, and this only happened because of fairly recent technology. A process called 3D capture was used in the design of the Guggenheim Museum in Bilbao, in which a 3D model was "digitized" by means of computer sensors originally developed for medical purposes. The exact dimensions of the model are taken into the computer where further design and detailing is done, before using the computer to make a new physical scaled model for Gehry's inspection (Marshall 2001).

From scaled models, parametric models inherit their iterative design nature, their use as experimentation and analysis tools, and their tridimensionality.

Modern descriptive geometry defines orthogonal drawings as being projections from a point situated at infinity, meaning that parallel lines do not cross each other as they do in perspectival drawings (or rather that they meet at infinity), and therefore lines retain their dimensions throughout the drawing. It is for this reason that Leon Battista Alberti in his treatise assigns them the role of the notational documents to be used by the builders. He makes a clear distinction between representations suitable for design and reserves for construction only that can be measures with precession (plans and elevations).

In Alberti's conception of the architectural design process, the architect is not to be involved in the construction process. Therefore the information necessary to build his design must be very well described and detailed enough for the builders to proceed in his absence. As the distance between designer and builder grows, the need for accurate notational tools also grows. Architecture becomes more and more allographic. This is why notational orthogonal drawings become highly important in Alberti's idea of Architecture.

> "Alberti's design process relies on a system of notation whereby all aspects of a building must be scripted by one author and unambiguously understood by all builders. Its principal notational means reside in the scaled a measured drawings of plans, elevations, and side views defined in the second book of *De re aedificatoria*."

(Carpo 2011)

Alberti sets the basis for the modern design and construction process and even with the advent of computational CAD, CAM and BIM technologies, we still require orthogonal plans and sections for most construction processes, especially in the bureaucratic and legislative review stages.

It is widely known and documented that the dimensions of structural and ornamental features of classical architecture are strictly related to each other by means of a system or proportions. Bernard Chache explains the presence of parametric relationships in these proportions in Vitruvius' *De Architectura*:

> "Let us turn again to the oldest treatise on architecture that has come down to us. Its author, who adorns himself with the title of "architect", spent the greater part of his career designing machines of war. The components of these machines were assembled according to parametric relationships. The most important of these relationships one far more complex than any simple fraction served to determine a module that was dependent upon the weight of the stone that was to be catapulted. Invented for the purpose of calculating this proportion was an apparatus (the Greek word for it was armonia), constructed of wooden slats and a cable, a device which, while not a computer in the contemporary sense, nonetheless facilitated the execution of a large number of computing operations. Such contraptions must appear familiar to contemporary architects who design components that can be varied in dependency upon parametric relationships. Is it inconceivable to construct a trajectory of tradition between todays parametric design techniques and the oldest surviving architectural treatise?
>
> To be sure, one should guard against excessively hasty comparisons. The contexts of antiquity cannot be equated without further ado with the circumstances of our own times. Still, it would be an error to consider such historical contexts in strict isolation from one another, since that would eliminate at the outset all questions regarding the survival of related problematics."

(Cache 2009)

Chache illustrates the presence of parametricism using Vitruvius' writings on war machines, not on buildings. The reason for this might be that parametric relationships in these war machines served a design process, the construction of the machine was directly related to the parameter (the weight of the stone to be catapulted). This ties function and design very tightly through a parametric process. But surely parametric relationships in antiquity was not confined to the design of war machines. Mario Carpo describes parametric processes involved in the architectural order:

> "The classical columnar system, first described by Vitruvius and later known as "the rule of the orders", is based on precisely determined norms and standards. Every component has a recognizable form and a name; composition - the assembly of the parts of the system - follows rules similar to those applied by ancient rhetoric by literary discourse. The precise quantification of particular dimensions and distances was an essential part of this process. Most normative measurements were defined as proportions: the traditional units of measurements were derived from parts of the building itself - typically, but not exclusively, from the diameter of a column."

(Carpo 2003)

Although Alberti was a strong believer in orthogonal drawings as the notational tool to instruct builders, he knew that for the diffusion and reproduction of his treatise he needed a different method of representation to describe his versions of the classical orders.

> "Before the invention of print, manual copies of drawings were famously untrustworthy, and as a result, images were seldom used, or altogether avoided, whenever precise copies were required. In such cases, non visual media (alphabetical or alphanumerical) were deemed safer. For centuries in the classical tradition (from antiquity to the middle ages to the early Renaissance), most architectural descriptions were verbal, not visual."

(Carpo 2011)

In book III of *De Architectura* Alberti writes about Vitruvius' description for the Attic Base, while in book VII of *De re aedificatoria* he writes his own instructions for the same base. In his paper "Drawing with Numbers"

Mario Carpo explains some of these descriptions in detail and how they evolved during the course of history. Alberti's instructions for calculating the proportions of the Doric base are translated and explained as follows:

> "First, take the diameter of the column at the base,and divide it into two equal parts. This gives the total height of the base. Then, take this segment and subdivide it into three equal parts; the lower third is the plinth. The take what is left, divide it into four equal parts; the upper quarter is the upper torus. Then take what is left and divide it into two equal parts; the lower half is the lower torus. Then take what is left and divide it into seven parts, and the upper and lower seventh are the two fillets. What is left is the scotia, sandwiched the two fillets and tori. Thus the sequence is completed."

(Carpo 2003)

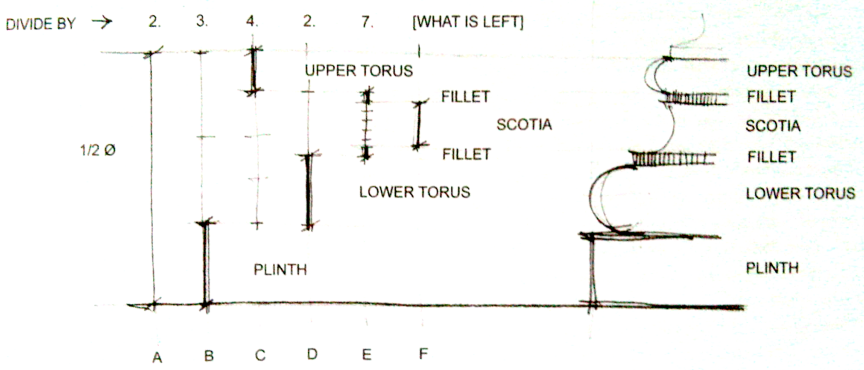


Figure 2.1: Mario Carpo's diagram of Alberti's instruction for determining the proportions of his Doric Base in the seventh book of *De re aedificatoria* (Carpo 2003).

These verbal descriptions or instructions that Carpo is talking about can be best defined by the term *Algorithm*. Figure 2.1 shows Carpo's diagram of the procedure and the resulting proportions. They are a sequence of mathematical operations by which we unequivocally obtain all measurements in the base, starting form one given dimension, in this case, the diameter of

the column at the base. The relationships between parts of the base are invariants, they do not change as the diameter of the column does. The diameter is the only number that is not obtained by the algorithm. If we think of this number as a variable number or designer defined *parameter*, this process could be called parametric modeling. It is the same exact operation used by contemporary designers with different geometries, parameters and design goals in mind.

Before Alberti, and in some cases after, this kind of algorithmic representation was used also as notation in the building site. The operation can be made in a geometric way with the use of a ruler and a compass. The algorithms described by the architect were reproduced live in the building site to obtain the desired geometry.

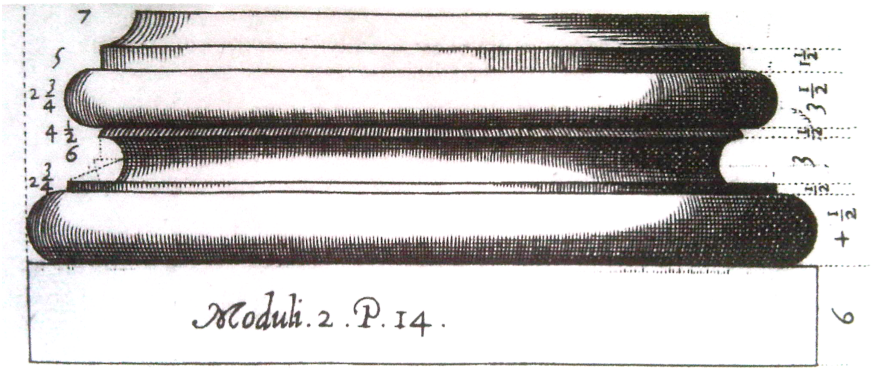


Figure 2.2: Attic base from Giacomo Barozzi da Vignola "Regola delle cinque ordini" (Rome, ca. 1562-63) (Carpo 2003).

Carpo goes on to write how after Alberti, and more importantly, after the invention of the printing press that could reliably reproduce images, verbal or textual descriptions of geometry started to change. Images started to appear, and proportions started to be expressed geometrically, then in fractions, and finally in numbers. Figure 2.2 shows Vignola's Attic Base. For the first time, proportions were pre-calculated (leaving only a few fractions) by the author (Vignola) and the reader simply used the numbers as dimensions, much like we use today in contemporary orthogonal drawings. The numbers in the image represent the number of "moduli" belonging to each measurement.

2.2 Contemporary Parametric representation

Computer Aided Design has its roots in the 1950's, but it only started being a part of the average architecture studio in the mid 90's. And even then, most contemporary architecture studios make use of this technology in a very traditional way. Digital Orthogonal drawings and 3d Models are used in a way that is not very different from what Alberti described, the only significant improvement the new media brought is that it allows the architect to interact with the building and redraw it with higher speed and precision and with lower costs. The arrival of CAD technology represented an improvement in speed ad reliability.

Non Rational B-Splines or NURBS were an important addition to the architect's representational toolbox. Complex Curves and curved surfaces have always been present in architecture, but this innovation on the mathematical description of this geometry brought with it precession, repeatability and ease of use the likes of which had never been seen before. This mathematical application gave architects the notational capability of describing very complex shapes.

While the arrival of these technologies might not have had a strong importance at first, they would pave the way for contemporary parametric modeling. CAD software is fundamental in the revival of algorithms as a method of architectural representation, and most importantly, their use as contemporary design tools. The power of computation is now beginning to be used for design purposes, not just representation.

Algorithmic representations in classical and renaissance treatises are, from a mathematical point of view, very similar to contemporary parametric models. The algorithms used by Vitruvius and Alberti were perhaps mostly used as geometric representation, is it arguably unlikely that they were used as design methods. Changes in the classical orders (hence on the algorithms) were not made by architects from one building to another, much less in a single building. So we can say the algorithms were not used as parametric models with design purposes but only as geometrical representation, their use was strictly notational *.

Perhaps the most significant difference between pre-computation and present use of algorithms is their use of CAD software as an immediate

*There are some geometrical an proportional constructs that were used to make geometry and structural elements parametrically to the room size (see (Tomasoni 2008) pp. 37-40 for examples on vault scaffolding construction in the XIX century), but these examples are hardly comparable to the use that is being seen today. Also Bernard Cache writes about the use of parametric models in Vitruvius' war machines.

graphical restitution so that the designer can quickly interact with the resulting geometry. This interactive aspect is a fundamental part of contemporary parametric modeling, because by interacting and modifying the model in real time, the designer maximizes the number of design iterations and hopefully improves the quality of the building in some way. While not perfect[†], parametric modeling is currently the fastest and most flexible geometrical representation method at hand, especially when compared to traditional drawings and scaled models.

This flexibility is not only exploited from a design efficiency point of view, different instantiations or "versions" of the model are often used in the same building or object:

> "A series is a framework of parameters designed by the architect, within which a variety of design versions may be realized. Each of these design versions is unique and yet also part of the series. The parts assembling each of the series' designs are no longer necessarily mass-produced but could rather be mass-customized... Versioning is at the core of the digital form itself; its signature and its authenticity derive from the parameterized repetition which give computer-generated design its characteristic combination of tightly disciplined structure and formal variability. Its not just the new calculus-powered curvaceousness, which is characteristic of a digitally informed age; it is also a disciplined groundwork of order that underpins the whole operation – the rhythm of a powerful Turing Machine that drives the versioning at the heart of the digital aesthetic."

(Rocker 2008)

Discipline and order are indeed required in the definition of a parametric model. To determine what is invariant, what needs to be constant throughout the entire series, is to determine what is fundamental and important in the design process.

The architectural practice today is characterized by its relentless speed. Design and construction must both be done at an ever growing speed that reduces the designer's ability to carefully consider all of the options available to him, and the consequences of his choices, often at the expense of the

[†]See (Davis 2013), in chapter 2 of his PhD thesis he explains the seldom discussed challenges of creating completely flexible parametric models with the current tools available to architects.

resulting buildings quality. A very important aspect of computational design tools to provide the designer with a geometrical representation method that is fast and flexible enough to accommodate this fast and demanding building environment. Higher flexibility and speed allow designers to consider a higher number of alternatives, edit their designs with much less effort (Davis 2013). Flexibility and speed are thus a very important advancement in representation, not just a slight improvement, it is the response to todays fast design and construction cycles.

Parametric modeling has also represented a bridge in the gap between design and construction. This representation method allows architects to describe a large amount of different objects in a precise and fast way:

> "If all that is built is built from notations, and if the drawings (or models) must contain all of the necessary data for an object to be built identically to its design, it follows that in most cases what can be built is determined by what can be drawn and measured in drawings. And as the notational system that encodes and carries data in architectural design is mainly geometric, it also follows that the potency of some geometrical tools determines the universe of forms that may or may not be built at any given point in time... This notational bottleneck was the inevitable companion of all allographic architecture from its very start... By bridging the gap between design and production, this mode of digital making also reduces the limits that previously applied under the notational regimes of descriptive and pre-descriptive geometries, and this may well mean the end of the notational bottleneck"

(Carpo 2011)

The notational bottleneck Carpo is referring to is at the heart of contemporary computational design and construction tools, and parametric representation of geometry is a key aspect. Moreover, we are increasingly reducing the difference between design representation and building notation, until eventually no bottleneck will exist.

In his book the "Alphabet and the Algorithm" Mario Carpo talks about algorithms that were used by designers to describe building components, and that those same algorithms were repeated by workers on the building site. This statement does not apply in todays machine production. Computer controlled mills or 3D printers do not follow the same codes that designers

use to generate their models. However, parametric models are an ideal environment for the representation of construction components and their assembly, regardless of their means of production, be they mass produced or customized.

The term mass-customization is generally applied to automated construction processes that are capable of producing components all different from each other with no additional cost. Mass-customization is comes as a response to mass-production in which the costs of molds and machinery was amortized by the production of many identical pieces. But mass-customized components only make sense when we have the means to design all of these different pieces in mass as well. The description of components interns of invariants and parameters allows designers to maintain rigorous control over all of the different pieces without manually drawing each one.

The algorithms that define parametric models are not described in verbal or written form as they were in Alberti's time, they are described in computer programming environments. There are multiple programing environments used in the design community today. They vary in their potential functionality, user interface and computation times. Initially parametric models were always done by incorporating customized pieces of software or "scripts" inside CAD environments. These were written by the user himself in various programing languages, sometimes adapted by the CAD programs to simplify their use. Examples of this are "Rhinoscripts" in Rhinoceros© and Maya Embedded Language (MEL) scripts in Maya©. Recently, commercial CAD software have been expanded to include graphical programing environments that enable users with little or no programing skills to create parametric models. Examples of this category in the architectural design community Generative Components© from Bentley Systems©, Grasshopper© for Rhinoceros© and very recently Dynamo© for Autodesk Revit©.

Designers can define a whole family of geometrical objects, to study in a search process, inside a single parametric model. However defining a parametric model is a process that needs to be made carefully from a design point of view[‡]. Designers must be aware of all of the possibilities of that particular parametrization, all of the geometry that is included and all of the geometry that is not included. This careful study of the possible outcomes of the model is fundamental when eliminating or considering possibilities.

[‡]An example of the consequences of different parametrizations of a given geometry, but from a search point of view is given in section 11.2.

3

Typology and Search

In the second half of the past century a theoretical discussion on types and typology began to take shape and became an important part of architectural theory ever since. It is worthwhile to revisit some of these texts with a new point of view, thinking about digital tools and parametrical modeling. Giulio Carlo Argan in his 1963 essay "On the typology of Architecture" citing Quatremère de Quincy gives us a clear idea of the type and the Model:

> "Quatremère de Quincy gives a precise definition of an architectural "type" in his historical dictionary. the word "type", he says, does not present so much an image of something to be copied or imitated exactly as the idea of an element which should itself serve as a rule for the model:
>
> > ... the model understood as a part of the practical execution of art is an object which should be imitated from what it is, the "type" on the other hand is something in relation to which different people may conceive works of art having no obvious resemblance to each other. All is exact and defined in the model; in the "type" everything is more or less vague. The imitation of "types" therefore has nothing about it which defies the operation of sentiment and intelligence."

(Argan 1963)

We can draw parallels between the type and the parametric model, as both being containers of a series of models or versions that have common

characteristics but are all indeed different from each other. Each type has invariant characteristics that all of the instances contained in it share, in the same way as parametric models have invariants. All objects of a given type are different, they contain differentiating features in the same way parametric models do. In the case of parametric models these features are called parameters. Quatremère de Quincy also alludes to the use of types in relation to the "conception" of new works of art. Types are not just meant for abstract theoretical conceptions, but they are a part of the creative process as well.

This parallelism between types and parametric models has been mentioned by architects and critics in the past*, but what is important for this PhD research is how this parallelism can help us understand the role that parametric modeling and search algorithms can have in architectural design. We will try to trace a link between typology and design thinking.

Very important to our discussion on Search and Typology is the way Argan describes how a type is created or "formed". Argan illustrates this process in the following way:

> "The notion of the vagueness or generality of the "type" which cannot therefore directly affect the design of buildings or their formal quality, also explains its generation, the way in which a "type" is formed. It is never formulated a priori but always deduced from a series of instances. So the "type" of a circular temple is never identifiable with this or that circular temple (even if one definite building, in this case the Pantheon, may have had and continues to have a particular importance) but is always the result of the confrontation and fusion of a series of buildings having between them an obvious formal and functional analogy. In other words, when a "type" is determined in the practice or theory of architecture, it already has an existence as an answer to a complex of ideological, religious, or practical demands which arise in a given historical condition of whatever culture."

(Argan 1963)

Types are *deduced* not formulated a priori, we create them while looking into past buildings, the information we use to trace types resides in past

*see for example the conversation between Antoine Picon, Mario Carpo, Ingeborg Rocker and Michael Meredith at the end of their lecture entitled The Eclipse of Beauty: Parametric Beauty (Rocker et al. 2011).

experience. Argan talks about "formal and functional" analogies between these buildings, alluding that types do not only refer to formal characteristics, but they also are formed for functional or performance reasons. The final phrase in the previous paragraph is also key. By saying that when a type is determined, there is already an answer to complex demand, Argan is telling us that types become types because what we see in them is important information. We do not create types randomly, we create them to guide us in future experiences, to learn from the past. Commonly, the invariant characteristic that defines the type is embedded with "answers", with a virtue of some sort, in some cases this can be measured by some performance evaluation. This virtue might even be the reason why the type was formed in the first place. Successful building features are repeated, and soon enough types are formed.

In his essay "Typology and Design Method" Alan Colquhoun also alludes to the knowledge present in architectural types, and the designers ability to adapt it to the present:

> "In mentioning typology, Maldonado is suggesting something quite new and something that has been rejected again and again by modern theorists. He is suggesting that the area of pure intuition must be based on a knowledge of past solutions applied to related problems, and that creation is a process of adapting forms derived either from past needs or from past aesthetic ideologies to the needs of the present".

(Colquhoun 1969)

Alan Colquhoun is talking about intuition, the design knowledge present in the architect, and how it must be based on something else, past knowledge and its adaptation to present problems. Both Argan and Colquhoun present us with a dual outcome from the study of types: (i) an abstract knowledge of the types, their definition and relevance in architecture theory, and (ii) a more practical or operative use of the types, as containers of architectural forms and function. Aldo Rossi wrote about typology in many and varied forms, assigning types all kinds of values and ideals. In this small passage he talks about the function of types:

> "In all of these definitions it seems that the function of types is that of warning us in advance of what will be the future experience; in other words they enable us to anticipate the course

of the design process."[†]

(Rossi 1975)

Again in this passage we see the idea of anticipation, of advice from the past that is useful on foreseeing something in our project, the implication that the study of types has a function in the design process.

3.1 Typology and performance based Search

The two most important characteristics of types that we can take form the previous discussion is their outcome into (i) design knowledge and (ii) practical operative use. An analogy between types and parametric models was outlined above, as they both define the forms in invariant and variable elements. A more interesting analogy can be traced between the study of types and Search processes as proposed in this PhD thesis. Types are not only defined by their forms, but as Argan writes, by their formal and functional aspects. Parametric models on to themselves contain only geometry, but search processes give us performance based information on a large series of solutions. Functional descriptions of entire sets of solutions are the outcome of the search process, and from this outcome we can derive both (i) design knowledge on the set and (ii) practical operative information pertaining to a current and specific architectural project. From this point of view we can see a clear parallel between types and search processes.

Performance evaluations on their own, with no exploration involved, give us only knowledge on one particular aspect of a design solution, they do not give us much design knowledge or paractical information. Optimization processes, as opposed to search processes, give us very practical information on very detailed, specific and limited aspects of a present project, more generalizable design knowledge is not given.

The knowledge present in architectural types is quite varied in its nature. If we think back at Andrew Witt's studies on Design and Instrumental knowledge[‡], we can surely say that Types allude to both design and instrumental knowledge. Some Types allude to spatial values, some to technical

[†]Translated from: "In tutte queste definizioni sembra che (esprimendo i concetti in forma sintetica) la funzione dei tipi sia quella di avvertirci in anticipo di quale sarà l'esperienza futura; in altri termini essa ci mette in grado di anticipare il corso della progettazione."(Rossi 1975)

[‡]see section 1.6.3 in page 26 of this thesis, and his article "A Machine Epistemology in Architecture. Encapsulated Knowledge and the Instrumentation of Design"(Witt 2010)

performance values such as structure or circulation distribution. Argan gives us his classification of types:

> "Although an infinite number of classes and sub-classes of "types" may be formulated, formal architectural typologies will always fall into three main categories; the first concerned with a complete configuration of buildings, the second with major structural elements and the third with decorative elements."

(Argan 1963)

Other authors have completely different classifications of types, and they certainly vary in the different disciplines associated with architecture. In his masters thesis, Myron Goldsmith talks about the effects of scale in structures, most relevant to our discussion he talks about the limitations of each Type of structure, in particular he makes the example of railroad bridge structures (Goldsmith 1953).

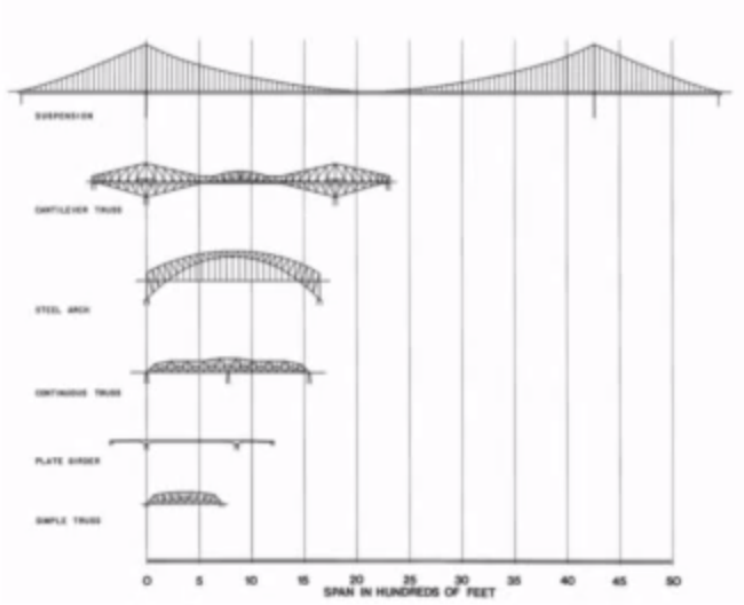


Figure 3.1: Myron Goldsmith: Bridge Structure Types (Goldsmith 1953)

Figure 3.1 shows us a diagram of several bridge Types and how they compare to each other in terms of their span. The diagram of course only shows us one bridge for each type, one suspension bridge, one steel arch bridge, one simple truss bride, etc. But we know that for each type there are a great number of existing bridges, many instantiations of the type. Yet all bridges of the same type share roughly the same performance metric, the same limit in span. More accurately, we can say that each type has one particular instance that has the longest span, a single solution can represent the maximum performance of each type.

As it was discussed above, types are discerned after a good number of instances of the type have been built. The common invariant feature among these instances is very often the cause of the repetition of the type, its virtue. The study of types can therefore be a good starting point in the definition of parametric models for the purposes of performance based search processes. Of we expect to find high performing solutions, for either explicit or implicit goals of our search process, then the invariant feature in our parametric models has to be well thought out. We can stand to learn from types as to how to formulate search processes.

Architects often explore types in all three of Argan's categories (and many others [§] very early in the design process, one of the very purposes of the conceptual design phase is the decision of a large scale geometrical shapes and for example the principal structural elements. Performance based search processes should not be different. The discovery of strong performing building features by means of search processes is analogous to the process of the creation of a type. In other words, if we can "anticipate" what the design process will be from the study of types, if we can translate past experiences into design knowledge, the same is true of performance based search processes. Knowledge on the behavior of a large set of solutions can be achieved by means of search.

While traditional types are formed by looking back at a series of buildings, they are deduced, performance based search, by means of simulations done at the moment of design can help us discover new and high performing building features. Design knowledge can be generated not only by deduction, but by performance evaluations of large sets.

The study of typology can help us better formulate questions for our search processes. If we consider Rittle's warning about dealing with "wicked" problems trough automation,[¶] we can say that we can better formulate

[§] An example considering Concert hall types is discussed in the section 3.2.

[¶] See Rittel's quote on page 29.

search processes and ask the right questions if we take a look at types, their performances, variables, constrains and most importantly the commonalities between types. Innovative and never before seen high performing solutions however are less likely to come from the study of types. Search processes based on performance simulations can give us information on solutions that do not yet exist. Therefore, search processes are capable of generating knowledge on future solutions, regardless on the question asked, be it an old problem or a new one.

Search processes give us performance values for entire sets of possibilities, allowing us to group high performing solutions and study their features. When it is the case that high performing solutions share common formal characteristics we can begin to learn why some solutions are better than others, we can extract important knowledge that is useful both in present and future problems, both design knowledge and practical operative information. We can then generate parametric models containing those high performing features as invariants, meaning that at most of the instances contained in the model are also high performing. Arguably, if these instances were to be built and studied, they could eventually become architectural types.

3.2 The Origin of a new Type: The case of the Berlin Philharmonie

Architectural Types are not created but rather deduced from studying a series of past buildings. Following this logic we will look at the "Vineyard" type of concert hall. We can see a series of built examples all over the world, and recognize in them many common characteristics. The vineyard concert hall however did not exist before the design and construction of the Berlin Philharmonie (figure 3.2) in 1963. We can therefore say that architect Hans Scharoun and acoustician Lothar Cremer designed a concert hall that would later become a Type.

Before the design of the Berlin Philharmonie most concert halls belonged to either the "shoebox", Fan shaped or Hexagonal type. But after Berlin, a large number of concert halls were built following many of the characteristics of this important hall. This makes for an interesting case in the study of types, since it is very recent and we know a lot about the architectural and acoustical design of this room, as well as how the room type has since been applied in many other concert spaces. We know what remains constant in

these rooms, and what makes these rooms architecturally and acoustically interesting.

Figure 3.2: The Berlin Philharmonie - photo credit: Alfredo Sánchez Romero.

"Music in the center" was the main requirement Architect Hans Scharoun made to Acoustic consultant Lothar Cremer. he wanted the audience to completely surround the orchestra, giving them the chance to sit behind the musicians and face the conductor, or sit beside them and look at their performance from up close. The architect argued that traditional rooms, where the orchestra performs at one end and the audience sits strictly in front of them, have a limitation when it comes to orchestra and audience communication (Beranek 2004). Under Scharoun's model, the audience (all 2215 of them) would sit no farther than 30 meters from the stage. This spatial relationship between the orchestra and the audience that the architect desired presented Cremer with important challenges:

> "The original concept of Scharoun was to have a completely circular hall with a shape close to an amphitheater where the orchestra director would be standing exactly at the centre of the

> circle, under a dome shaped ceiling an acoustically very dangerous concept as this geometry is prone to serious acoustic focusing. The principle behind Scharouns concept was to position the orchestra as close as possible to the centre and thus create the most "democratic" hall. To respect the fundamental rules of acoustics, Cremer suggested a ceiling with a tent shape rather than a dome and to break up the symmetry of the hall by using convex curves. He replaced the concave curves, which tend to focus sound, with convex curves, which diffuse sound. The idea of a central orchestra was kept. Also, the fact that the audience is located behind and to the sides of the stage, combined with the absence of a balcony has resulted in a room width that is much bigger than that of shoebox halls, and clearly wider than what is acoustically acceptable without having to introduce compensating elements. The latter elements, consisting of large wall sections, or partial walls creating "vineyard terraces", helped to reduce the apparent width of the hall and create acoustic reflections, leading to the concept of the vineyard concert hall."

(Kahle-Acoustics & Altia-Acoustique 2006)

The Berlin Philharmonie concert hall was a big success both with musicians and audiences. They both appreciated the new found intimate contact between each other, and the acoustics of the room were also praised. Since the opening of the Berlin Philharmonie and other vineyard halls, several studies have been done on their acoustics and been compared with other types[‖]. Several strengths and weaknesses have been revealed and the importance of its features been detailed. The type has been of course refined and developed in different ways.

The four most important concert hall types are the Shoebox halls, the Fan-Shaped hall, the Hexagonal hall and the Vineyard hall. Historically and still today the most popular of them has been the Shoebox, but from the construction of the Berlin Philharmonie on, the hexagon and fan-shaped halls have been built much less frequently and the vineyard hall has seen a big development worldwide (Meyer 2013). Other recent examples of concert halls built under this type are the Danish Radio Concert Hall by Jean Nouvel (Figure 3.3), the Disney concert Hall by Frank Gehry or the Elbe Philharmonie by Herzog & de Meuron.

[‖] see for example (Hidaka et al. 2008)

Figure 3.3: The Danish Radio Concert Hall - photo credit: Frans Swarte.

Of course not all novel buildings translate into new Types. Not all innovative building features are high performing and appreciated by architects and the public. It is therefore interesting to investigate why did the Berlin Philharmonie became a type. Why did architects and acousticians repeat its distinguishing features? Why didn't they attempt other ways to improve the intimacy without the use of the vineyards terraces?

The reason for the repetition of the vineyard concept was the combination of the effective architectural idea with the high performance acoustical solution. Scharoun's idea "Music in the center" was proven very successful in the architectural realm and the experience of the concert was enhanced by it, but without Cremer's terraces and reflecting walls, this concept would not have been acoustically satisfying. The terraces and their walls achieved the desired early reflections that would be otherwise missing, and this is why *both* the stage in the center and the vineyards were repeated, and not just one or the other.

In the first part of the twentieth century fan-shaped halls were very popular among modern architects, and many of them were built. Quite a few

Figure 3.4: The Aula Magna of the Unversidad Central de Venezuela by Carlos Raúl Villanueva.

examples can be mentioned, from Le Corbusier's unbuilt Palace of the Soviets, to Alvar Aalto's Auditorium of the Helsinki University of Technology, to Carlos Raul Villanueva's Aula Magna in Caracas. All of these rooms have angled walls opening up away from the stage, a concave curved back wall, and the seats are arranged in concentric circles of increasing radii, much like the greek amphitheater. This arrangement of the seats guarantees that seats in the same row are all at the same distance from the stage, making it an efficient seating arrangement and enhancing the intimacy of the room. Architects repeated this type because they liked the arrangement and they associated its shape with the greek acoustical quality. While the greek amphitheater was an outdoor environment, these spaces were enclosed, and the room shape was not optimal. The big distances of most listeners in the center of the room to the nearest wall and the opening angle of these walls, causes a big problem in receiving early sound reflections. The concave back wall also causes problems.

The fan shaped rooms were very successful from an architectural point of

view, they were repeated all over the world. But they are being increasingly phased out for acoustical reasons.

The main lesson to learn from the Berlin Philharmonie is that Types are created for a reason, they are there because of some performance, spatial or economical reason that makes them desirable, repeatable and/or interesting to designers and users in general. The vineyard was repeated for both architectural and acoustical reasons. This is why when we look back into a particular type we can distinguish spatial and architectural characteristics as well as many technical advantages or problems related to the type. This is why Aldo Rossi confers types the ability to put us in the condition of foreseeing the design process ahead of us.

There is much to learn from the study of types, and this is especially true when using parametric models in combination with automated search methods. The questions asked when types are created should be a guiding example when we formulate search goals. The differences and variations the geometry of the single type should also be of example when we parametrize geometry for a performance search algorithm.

3.3 Cognition and search: Clustered search spaces

Search is a very broad term that describes many activities in the human (and animal) world. Human cognitive search mechanisms can be divided into two important categories, external search and internal search. External search is related to external goals, such as food and water, but also to external information. Internal search refers to Memory search, search for information we have obtained in the past and hopefully stored in our brains.

The interest in this distinction lies in the fact that when we design we mainly look for information that is external to us, we employ external search mechanisms, but we also rely on memory to access information we already found and might help us solve the problem. For example, when we use typological knowledge present in our memory, we employ an internal search mechanism. Parametric search processes are more related with external search mechanisms since it implies new goals that cannot be associated with memory.

From the beginning of the human species, we have always used our brains to search. The first and perhaps most primitive search problems for humans (as for all other animals) is the search for food or water, it is an external

search process. Food and water resources are not homogeneously distributed in the human habitat, food sources are distributed in patches of land heterogeneously present. For reasons related to landscape, climate and soil conditions, food and water resources vary in quality and quantity in different regions, but not in a random way. This primordial search problem had an heterogeneous, patched, clustered and non-random search space. The search space for this primitive problem seems to have shaped our brains and cognitive abilities thereafter (Hills & Dukas 2012).

Studies on human cognition have found a deep relationship between the search space in problems such as finding food or water (a spatial search process), and our search mechanisms. We seem to have adopted a spatial-like search mechanism for all kinds of search, internal, external, spatial or non-spatial.

Like in the search for food or water, external search spaces are engaged by humans in a local search first, global search second pattern. Humans tend to search for resources first in a local space, typically a space where they recently had success in finding resources. If they are unsuccessful in finding resources (or information) in that local environment, then they move on to a global search with a wider search space, and seldom return to local spaces where they have previously been unsuccessful. This local search first and global search second pattern is present both in external and internal search.

> "Human subjects tend to focus their visual attention on the vicinity of a recently detected target but switch their attention to other spatial locations if no target is found at this area within a short giving-up time. This behavior, which is reminiscent of area-restricted search is called *inhibition of return*"

(Hills & Dukas 2012)

We look for information first where we think we are most likely to find it, and if unsuccessful we expand the scope of the search into a more global space. This mechanism is ver similar in internal search. Internal memory search is directly related to the way we store information in our brains:

> "Studies in written language - presumably reflecting the internal structure of cognitive information - find evidence for a strongly clustered environment. With nodes representing words and links representing relations within words, these language

> networks often reveal a small-word structure, indicating that words are much more likely to appear together in small clusters of related items that one would expect by chance... Moreover, this structure of language and free association networks is well correlated with the order in which children learn about language. This indicates that the patchy internal structure of memory may be tightly linked with the patchy external structure of information."

(Hills & Dukas 2012)

The following example explained by Hills et al. may help to clarify these ideas about memory search:

> "Research on sequential solutions in problem-solving tasks also demonstrates that people show local preservation in internal search environments. For example, people tend to produce solutions that are more clustered together (i.e. similar) that one would expect by random generation; for example, in math search tasks and anagram search tasks. In one case Hills et al. had participants search within scrambled sets of letters for multiple words. Participants would see a letter set, like BLNTAO, and they could find "BOAT","BOLT", etc. An analysis of the string similarity (e.g. bigram similarity comparing the number of shared letter pairs: "BO", "OA", etc.) between subsequent solutions determined that participants tended to produce solutions that were most similar to their last solution. This was true even though previous solutions were not visible. Results indicate that participants were searching locally around previous solutions, before transitioning to a global search strategy."

(Hills & Dukas 2012)

Humans tend to store items in they memory in a clustered environments similar to the one where they got the information in the first place. Clusters in memory however are not organized according to repeatable categories. This is one important difference between internal and external search mechanisms:

> "An item in memory can belong to different representations simultaneously: the word "cat" can belong to the category "pets"

> as well as to the category of "predators". The representation need not be based solely on semantic similarity but also, for instance, on phonological similarity ("cat" and "bat"). Thus could potentially belong to more that one patch."

(Hutchinson et al. 2012)

Clusters in memory may categorize information in ways that are not consciously decided by the person, and therefore internal memory search is less conscious. This might not mean that is less efficient, but external search processes can surely be more planned and systematic.

It seems that we are naturally pre-disposed to classifying information into clusters, so we should not be surprised at the presence of types and typology in architecture theory and practice. Types become clusters of information about the built environment, and these clusters are ready for us to investigate building potential for future projects.

The search space in performance driven search for architectural design is often unknown and unexplored. We do not know beforehand how the search space is populated with high performing solutions. We might be looking in a space that does not have what we are looking for, or we might we looking in a space where all solutions are high performing, either way we have little information from which make decisions. Also, we cannot be sure if we are dealing with homogenous or a patched environments, are the best solutions all similar and close to each other in the search space? or are they very different and far from each other? We need to employ tools that are able to give us relevant information regardless of the type of environment we have. It seems logical to use local first and global second patterns when we search for design information. We can use parametric models as clusters of information about design goals, especially when innovative forms and new design goals are in question. These clusters of information are there to be searched computationally, much as types have been studied in the past by the deduction and analysis of the existing built environment.

4

Automation and Authoriality

When automated processes are involved, design control and authoriality are often a point of discussion. A traditional point of view on automated processes would stipulate that the authorship of the resulting object is not clear, that automation somehow robs the designer of his design intent.

Automated processes are surely being more and more employed in todays design process, but they are hardly new. By looking into older examples we can discuss authoriality and design control with a larger context. This issue by going into detail in different scenarios of past and present use of automation.

Planned randomness

A recent trend in architecture, particularly in the design of facade and ornamental elements, is to look for the appearance of randomness, the use of casual and apparently chaotic geometry. In contrast with the standardized element and the repetition of forms, this recent trend intends to give the opposite impression, one of uniqueness, randomness and non repetition. In most cases this is achieved by the use algorithms, and then built with the help of digital fabrication processes. Voronoi diagrams, fractal elements and other algorithms are very present in contemporary architecture, examples are the Beijing water cube by PTW architects, the serpentine gallery pavilion by Toyo Ito and the Grand egyptian museum by Heneghan Peng (under construction). In all of these cases there is no actual randomness behind the generation of the form. The form was generated by an algorithm that was carefully planned by the design team. In this case there are no issues of

control, just the appearance of randomness. The end result is the subject of the algorithm, it is always under the control and supervision of the designer.

One example that merits mentioning is the stonework for the Therme Vals building by Peter Zumthor depicted in figure 4.1. In this case the architect wanted to avoid the look of a standardized stone wall, while still maintaining strict control over the geometry. The stones are laid in rows, all stones in the same row have the same height, but not all rows have the same height. In addition, Zumthor specifies another rule to this "natural" pattern, the combined height 2, 3 or sometimes 4 rows has to always add up to 15 centimeters, creating a repeating modulus. This allows him to use these rows as steps in the staircases without interrupting the pattern or breaking a row. Apart from these rules or constraints, the stonework is casual, and the length of each stone is different. We can describe this method as a set of constraints, a constrained* bit random selection of stones. The end result is a very controlled, yet not regular looking pattern.

In all of these examples, architects never relinquish complete control over their designs. In all of the above cases there are some features that are established by the architects as invariants, constraints are always presents. When the architect establishes constraints he is making a strong statement that determines, to a good degree, the final outcome. He establishes what cannot be changed, what makes the design his own.

It is true that in some of these examples there are features of the outcome that are not a direct result of the architects initial intent, he did not draw them with his pencil like in a more traditional process. But when the architect makes the conscious choice to let an automated process generate these random or external features, he is also making an authorial decision. Often he is looking for a particular result that he deems better achieved by means of automation or randomness that by means of his own hand.

The Accident

Perhaps the most representative work in this category lies outside the realm of Architecture, Jackson Pollock's dripping and splashing painting techniques depended on accidental and uncontrollable events to apply the paint to the canvas. Pollock, when talking about his technique, claims to have complete control and denies the use of the accident but then he also talks about the unconscious and the painting having life of its own. But it is clear that he used the splatter, and dripping of the paint as a means of expression.

*Axel Kilian devoted his PhD thesis to the study of constrained algorithms in architectural design (Kilian 2000).

Figure 4.1: Detail of the Stonework in the Therme Vals hotel building by Peter Zumthor - photo credit: Marco Palma.

This technique allowed him to achieve results that were not possible without it, hence this technique is crucial to the artist's intent, not just a means of exploration.

Even in such cases when the accident is deliberately chosen, and there is almost a complete absence of invariants, the artist is making a choice that is reflected in the resulting artwork. Authoriality is therefore not in question.

Since architecture is an allographic discipline, the accident is only employed by the architect in the design process itself, not in the realization of the finished product. The accident is used as an exploration tool during the design process, the accident has to then be interpreted and developed by the architect. In this process the architect takes full control of the design, the accident is an exploratory start. The use of scraped pieces of paper or hand-drawn random scribbles is comparable to Pollock's drippings in the sense that some accidental elements are involved in combination with the architect's intention.

Accidents can also present in the architectural construction process, especially when handmade components or natural materials are involved. A good example of this is seen in masonry buildings. Brick colors for example are all quite similar but never exactly the same. Brickwork patterns are usually pre-defined by architects, but the arrangement of the different tones of brick color are not. This is left up to chance or the will of the artisan laying the brick, yet the authorship of the building is never put into question, the artisans involved in the construction are almost never known or recognized. The same can be said about other construction processes involving artisans and their crafts.

Form-Finding

In 1675 Robert Hooke wrote an anagram containing the phrase "As hangs the flexible line, so but inverted will stand the rigid arch". This statement simply declares that the shape taken by a chain loaded by its own weight (hanging in tension) is the same shape of the arch that will carry the loads in compression (Heyman 1995). Gravity is giving the chain a particular and very specific shape, in a way we can call this process a computational process, the computation being done by nature.

This process became known as a form-finding process. Form-finding implies the computation of a particular shape that satisfies some natural phenomenon following an initial user defined configuration. A simple hanging chain model for example, gives different shapes depending on the length of the chain and the position of the two support points. Not all form-finding

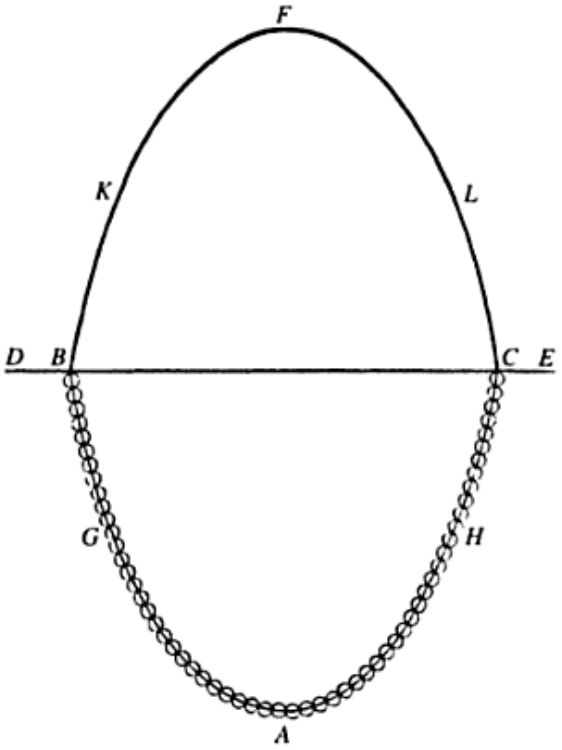


Figure 4.2: Robert Hooke's Hanging Chain

processes are done via natural processes, recently, computer simulations of those processes have been developed and employed for architectural design.

The most prominent example of a form-finding process is Antoni Gaudi's hanging models for the Colonia Guell in which he found the compression only shapes[†] to the vault structure by hanging chains and weights and reversing the result. Frei Otto used another form-finding process for the purposes of computing minimal surfaces between rigid curved elements (see Figure 4.3). These shapes were then used to design Tensile Structures, for example Otto's olympic stadium for the 1972 Munich Olympics.

[†]A good example of form finding processes for compression only complex surfaces can be seen in the work of Philippe Block (Block 2009)

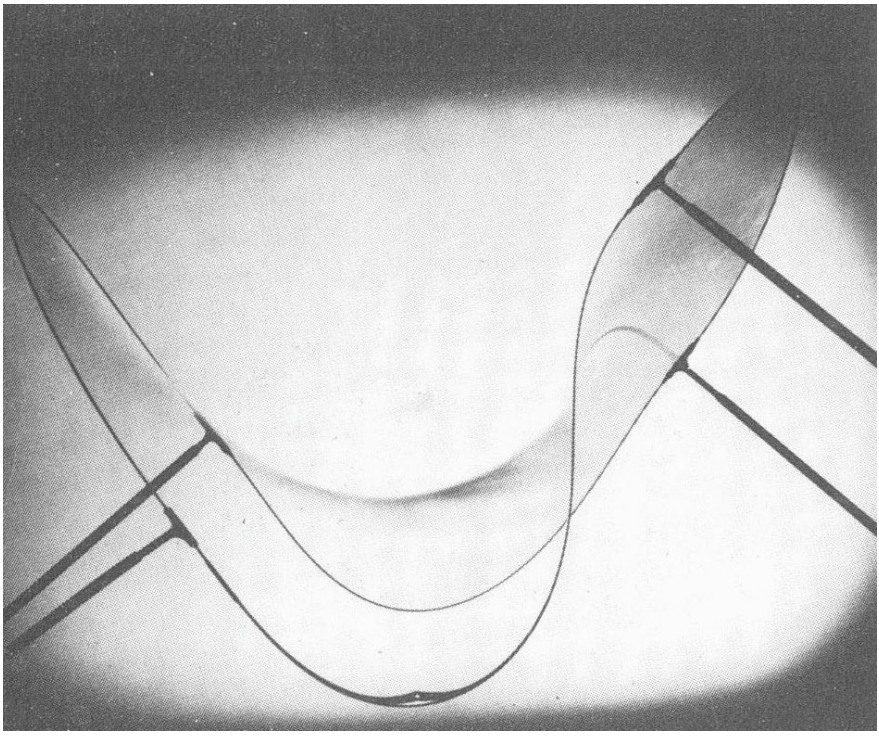


Figure 4.3: Frei Otto - Minimal Surface Studies by use of soap films.

In all of these processes there are instances where the designer leaves aside control over the design. He does not directly trace a shape with his hand, instead he leaves it up to the process to compute the shape, knowing that the end result will have an ulterior benefit that would not be possible to achieve otherwise.

In form-finding processes the interaction between the designer and the model is usually a fundamental part. Most of these processes generate a solution in a very short amount of time, and present it for the designer to consider, discard or modify. The resulting form is usually an input for the designer to make a modification in the model parameters (e.g. weights on the hanging model, shape of the metal elements in the soap film study, or numerical parameters in a digital form-finding tool). So of course the

designer regains control over the result immediately after the process.

Computational Search

Computational Search Algorithms like the ones described in this PhD also have elements of control to be discussed. In section 1.7 a discussion is presented on how designers can interact with search algorithms, but in between those interactions, it is the algorithm that is doing most of the work, and depending on the algorithm, there are plenty of random events involved. For example Genetic Algorithms contain various operations that involve the use of stochastic sampling, most notably the selection of the first generation of solutions is made completely and purposely at random[‡]. But randomness in some operators does not mean that the overall purpose of the algorithm is lost, on the contrary. Randomness is employed to increase the exploratory power of the algorithm, and to counteract the exploitative power of the genetic operators.

Architects in the past have dealt with operations in which the resulting design is not a direct result of their hand, many examples of this are shown in this chapter. But this does not mean that the end result is outside of their design intent.

It is easy to see that, with the increased use of computational tools, this kind of event is being used more and more, and the part of design that is left outside of the designers control is always bigger and bigger. However, at the end of the random event, the designer *always* has the choice to take or leave the result. The designer is obliged to decide how the rest of the design process will evolve, if it will be defined by the information found during search or not. If the designer decides not to take the results, he can then re formulate and regenerate new solutions, or simply undertake an entirely different process. If on the other hand, the designer decides to take the results and use them in his final project proposal, he takes the automated results as his own, assumes the responsibility of the choices made for him by the automation, accident or random event. He resumes control of the process.

[‡]see the complete description of the genetic algorithm in chapter 7.

Part II

Search Algorithms for Architectural design

5

Search problems and algorithms

Computer scientists over the years have developed an immense body of work in the field of search, creating a large number of algorithms and studying their efficiency. Selecting the right algorithm is a task that is very much related to the search problem it needs to solve. Computer scientists have also developed a big number of search problems and studied their complexity in terms of their how efficiently algorithms can solve them. A good example is the Traveling Salesman Problem (TSP). This problem involves a salesman that needs to visit n number of cities in his sales itinerary, and this cities are not all equidistant, travel between cities takes different amounts of time and have different cost (see figure 5.1). The problem requires the algorithm to search for sales routes that minimize the travel time and/or costs while taking him to through all of the cities in his itinerary and bring him back home.

This problem is not very related to the architectural search problems that are described in this thesis. If we are to select an appropriate algorithm for our problems, a good understanding of search problems in general and of the search problems we face is required.

A good definition of a search problem from a computational point of view is provided in(Schooler et al. 2012). Schooler et al. simply state that a search problem is represented by three elements (S, f, W) where S represents the search space of the problem, $f : S \to R$ is a function that assigns objective values to the solutions contained in S, and W is a set of constraints. This is a very general definition of a search problem, that applies to many situations. It certainly applies to the search problems discussed in this PhD thesis:

- Search spaces S are defined by the user by setting up a parametric

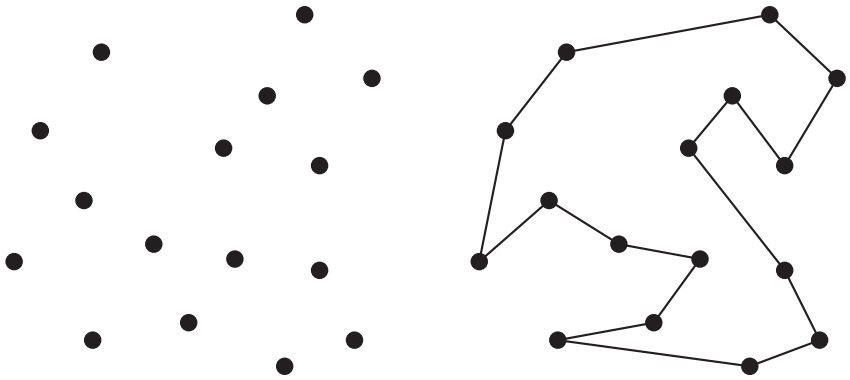


Figure 5.1: Diagram of he traveling salesman problem and one solution - Image from (Edelkamp & Schrödl 2012).

models. The number of parameters in the model define the dimensionality of S. The extension of S is defined by the parameter domains.

- Objective values are assigned by performance evaluation via computational simulation. Functions f are therefore defined by the performances we chose to use during the search process.

- Constraints are rare in our search problems, most of the limits to the search problem are set with parameter domains and invariants. Including in infeasible solutions (or black areas in the search space) is done by means of constrains W. This is not present in this thesis, but it is not excluded as a possibility in search methods.

Clearly this definition applies to our problems, and it gives us a starting point, but a more specific definition of search problems that we face is required.

Search algorithms in this PhD research are used to explore parametric models for high-performing solutions to multiple and contrasting performance functions. Parametric models may contain a large number of solutions, all identified by combinations of parameters.

A more detailed characterization of the problems in this thesis can be done by studying the search spaces involved. The search space S in architectural search problems is defined in the creation of the parametric model, it is confined by parameter domains. Perhaps the most important characteristic of our search space is that it is continuous. The number of possible values that a parameter can take is infinite, as is the number of subdivisions inside a given domain. If for example we have one parameter that defines sphere radii, and it has a domain between 1 and 10 meters, there are infinite spheres in this model.

Continuous search spaces are problematic for search problems, the number of possible solutions to search is infinite, as opposed to other problems like the traveling salesman. The number of possible routes in the TSP is directly related to the number of cities in his itinerary. The more cities there are, the more possible routes the salesman can take. But however high the number of cities the number of routes is always finite. Only if infinite cities are present does this problem have infinite possible solutions.

All search algorithms inevitably need to transform continuous search spaces into discrete ones, in other words it is impossible for algorithms to consider infinite solutions, only a finite number is studied. But some algorithms are better equipped to select deal with continuous search spaces and infinite possibilities. Even if we discretize the search space, thus dramatically reducing the number of solutions, if a high number of parameters are present, the number of possible solutions are can still be very high. This makes exhaustive search processes only feasible in very simple problems with a very coarse discretization of the search space.

Another important way of characterizing search problems is by means of the so called objective space*. The objective space is the counterpart to the search space, it represents not the solutions or their parameters, but instead it shows their objective values, the result of the f function. If we take for example a problem involving a single f function and a set of solutions contained in S, we can represent the objective space for such a problem with a curve. Figure 5.2 shows two objective spaces for single-objective minimization problems. We can see that solutions in S contain varying objective values for the same f function.

The problem presented on the left in figure 5.2 shows what is called an unimodal problem, meaning that the problem has a single minimum (optimal) value. The problem shown on the left is a multimodal problem, it contains multiple minima, meaning that the curve has various valleys. Some

*A detailed definition of objective space with examples is given in section 7.4.

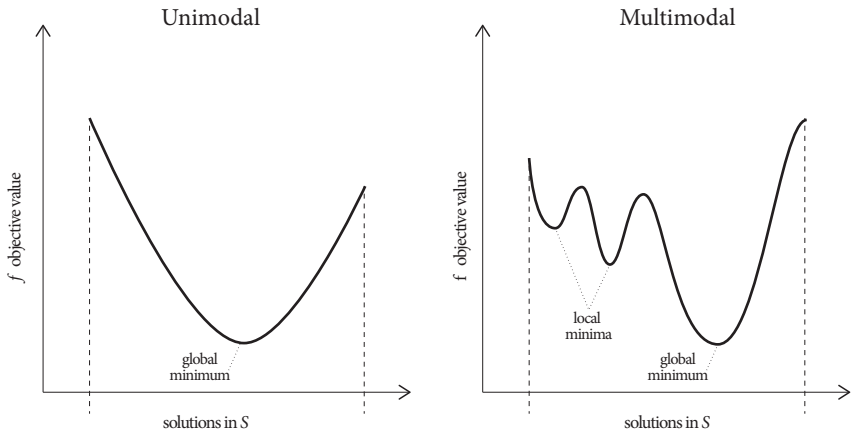


Figure 5.2: Objective Spaces showing Unimodal and Multimodal problems.

of the minima are only minimum when compared to the solutions adjacent to them. these are called local minima. The minimum value for the entire solution set is called global minimum. The distinction between unimodal and multimodal problems also apply to problems with a larger number of objective functions.

Multimodality represents a mayor challenge for search algorithms as they tend to confuse local minima with global minima, meaning that they do not converge into the solution the user has been looking for. Objective spaces might also be discontinuous or highly irregular.

The characteristics of the objective space are determined by the objective functions contained in the problem. In architectural search problems the objective spaces are impossible to know beforehand, we are blind to their shape and complexity when we perform a search process. More importantly, they are very different depending on the fact that we are studying structural shapes, acoustic quality in rooms or energy efficiency. All of this issues point to the fact that the search algorithms that we select for this PhD thesis needs to be able to deal with complex and unknown objective spaces, most likely multimodal. The algorithms need to be *robust*.

Another important characteristic of the search algorithms to be used in this PhD research can be signaled out by establishing what we are searching for, what distinguishes sought after solutions from the rest. As it has been

explained above, we are looking for high-performing solutions given our objective functions, calculated via performance evaluation software.

The identification of what the algorithm is looking for allows us to exclude a great number of search algorithms and focus on one category. The category of search algorithms that is best suited to find high-performing solutions out of an entire set is that of optimization algorithms.

In the previous chapters, a very important distinction was made between search and optimization. This distinction is not intrinsic to the inner workings of the algorithms, it regards the purpose and the moment in the design process it is carried out. This distinction does not refer to the mathematical definition of optimization, or different algorithm types. All search algorithms tend to minimize some objective function, but there is a group of algorithms that is more deliberate in this purpose, and that works in continuous search spaces. They are most commonly known as optimization algorithms. Optimization algorithms are search algorithms that minimize objective functions.

5.1 Algorithm classification

As it was declared above, we will focus our attention on optimization algorithms. As such, a classification of optimization algorithms is presented bellow. The most important way of classifying optimization algorithms is to divide them according to their optimization strategy. In this case we get deterministic and stochastic algorithms.

Deterministic algorithms are also referred to as classical because they represent the first efforts in optimization algorithms. Many deterministic algorithms also called gradient-based, because they look for optimal solutions by climbing or descending (for maximization or minimization problems respectively) in the direction of the highest gradient, the steepest hill. This means that they depend on a certain knowledge of the search problem. Algorithms in this category include Newton's method, pattern search and gradient descent. They can be defined by the following characteristics:

- Deterministic algorithms follow a strict formulation and produce the same result every time they are used on the same problem and from the same starting point.
- Deterministic algorithms consider one solution at a time, and they modify their position following the information obtained by studying only this solution.

- Deterministic algorithms' success is very much dependent on their starting point. In multimodal problems, they will arrive at the global minimum only if their starting point is in the basin of the global minimum and not of any local minima.

- Deterministic algorithms usually converge at a minimum (global or otherwise) at a very fast pace, especially when compared to stochastic algorithms.

- Exploration is not a big part of deterministic algorithms, they tend to converge into a minimum rather quickly without exploring other possibilities. This is why they have problems with multimodal problems.

Stochastic algorithms are characterized by the use of random or casual operations during their search process. This is done to improve exploration and guarantee that as big a part of the search space as possible is considered. An important part of their success and widespread use is the fact that they require no information on the problem to be solved, they are blind to the shape or complexity of the objective space. Algorithms in this category include Simulated Annealing, Particle Swarm Optimization, Evolutionary strategies and Genetic Algorithms. Most of these algorithms have their inspiration in natural phenomena or animal behavior. They can be defined by the following characteristics:

- Random events during the search process means that stochastic algorithms follow produce different results when they are used in the same problem.

- Stochastic algorithms consider one many at a time, and they modify their position following the information obtained by studying all of these solutions. This group of solutions is often called a population, and these algorithms are sometimes referred to as population based algorithms.

- Stochastic algorithms' have many starting points. For this reason, in multimodal problems they will arrive have a greater chance of finding the global optima when compared to deterministic algorithms.

- Stochastic algorithms usually converge at a minimum (global or otherwise) at a much slower pace, especially when compared to deterministic algorithms.

- Exploration is an important part of Stochastic algorithms, they have not only one position but many in order to better consider the entire search space.

5.2 Algorithm Selection

Search algorithms have been used by researchers and practitioners in the architecture and engineering fields for some time. The problems addressed by these architects and engineers are very varied in nature, disciplines and complexity, but an important trend towards the use of stochastic algorithms and genetic algorithms in particular is present in the literature. An investigation into optimization methods used in Building Performance Simulation (BPO) and Building Performance Optimization (BPO) was carried out by Attia et al. and published in their 2013 article "Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design":

> "Through the 2000s, the development of mathematical and algorithmic techniques and the advancement of BPS tools gave way to BPO tools that could solve multi-objective optimization problems of a design. Mechanical and structural engineers working on complex buildings have been among the early adopters of BPO techniques, but architects and other engineers now start using these techniques as well. Today, there is a strong trend towards population-based search algorithms such as evolutionary algorithms[†] and particle swarms. These algorithms have been proven to be very successful in optimizing one or many performance criteria while handling search constraints for large design problems."

(Attia et al. 2013)

This trend is also visible in other architecture related disciplines such as structure and acoustics. Some of this examples are to be considered as search processes in the way they are presented in this PhD thesis, and others are more in line with the traditional optimization process. But in all cases the use of stochastic population based algorithms is significant.

[†]Evolutionary algorithms is a term that is used to group all stochastic algorithms that are based on natural evolution. These are mainly Genetic algorithms and evolutionary strategies.

This PhD thesis does not present a comprehensive study of many search algorithms, nor a comparison of them as to which is best suited for some problems, or which one is more efficient. This PhD thesis focuses mainly on the use of genetic algorithms. Exhaustive comparisons of different algorithms would only serve the purposes of improving efficiency of the search process, and that is no small matter. Future efforts in comparing search algorithms as to their adaptation to architectural search could prove to be very useful in reducing calculation times or usability. But it is unlikely that the use of other algorithms (Apart from genetic algorithms) would provide additional architectural knowledge, or shed further light into the subjects discussed in this research. This thesis focuses on the inclusion of search algorithms and performance information in the early design phase, therefore algorithm comparisons lie outside of it's scope.

The selection of genetic algorithms for this thesis is not casual. They have been employed by many researchers and practitioners in the architecture and construction field.

GA's have been around longer than some of the other algorithms described above, and for this reason, they have been used and studied more in depth that other. Their efficiency has been discussed and testes in many occasions, and their use has been suggested by researchers in related fields. Furthermore, many implementations of multi-objective GA's are in existence, and their performance has been evaluated and proven successful (Zitzler et al. 2000, Deb 2001).

6

Genetic Algorithms

6.1 Intruduction

Genetic Algorithms (GAs) were first proposed by John Holland in the mid 1970's in the University of Michigan, hist most important publication being "Adaptation in Natural and Artificial Systems" (Holland 1975). They have been successfully employed in varied fields of study, more important for this work, they have been widely employed in the architecture and construction field, see for example (Bogar et al. 2013, Méndez Echenagucia, Pugnale & Sassone 2013, Miles et al. 2001, Turrin et al. 2011).

> "Genetic Algorithms are search algorithms based on the mechanics of natural selection and natural genetics. They combine survival of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. In every generation, a new set of artificial creatures (strings) is created using bits and pieces of the fittest of the old; an occasional new part is tried for good measure. While randomized, genetic algorithms are no simple random walk. They efficiently exploit historical information to speculate on new search points with expected improved performance."

(Goldberg 1989)

GAs follow a darwinian model of search in which the concept of fitness is defined by the user and it can be any such function that can be expressed

numerically. In nature, survival of the fittest has, through the process of evolution, generated animal species that are very well equipped to life in their particular habitat. We can look at evolution as a search process intended to find the attributes that can best guarantee an animal's survival and reproduction under particular environmental conditions. In this analogy, survival in a specific environment is the problem set, the animal population is the set of proposed solutions, and natural selection and reproduction is the search process that generates the best solutions. While in nature its the environment to determine what problems the search process needs to solve, in computational genetic search its the user who determines what the population of solutions must achieve. This is one of the reasons why this technique has been employed in such varied fields.

As Goldberg said in the definition above, new solutions are created in GAs by the use of bits and pieces taken from the best of the previously considered solutions. These bits and pieces are taken from the problem variables. Like in all search algorithms, in GAs solutions to the given problem are characterized by a set of variables, and these variables are typically represented by numbers. These numbers are then coded, there are many ways of coding the variables, including real numbers coding, but perhaps the most common code used is binary code. Following the biological analogies used to describe genetic algorithms, this code then becomes the *chromosome* or *genome* of the individuals in the population. Like in nature, the offspring or new generation of individuals are made up from the chromosomes of their parents*. The success of the GA is related to the selection of the right parents, and the correct combination of their chromosomes.

Genetic algorithms differ from more traditional search methods in a few key points:

- GAs are population based, search is made in several points at each iteration.
- GAs normally work from a coded version of the parameters or variables, not often from the parameters themselves.
- GAs use objective or fitness functions to drive the search process, they do not use any other auxiliary or additional information.
- GAs use probabilistic transition or movement rules, they do not use deterministic calculations to determine the next step, such as gradient based search.

*A more detailed explanation of this operation is given in section 6.3 bellow.

(Goldberg 1989)

Genetic algorithms perform very well in multi modal problems. While more traditional search algorithms are vulnerable to getting stuck in local minima, GAs can surpass this issue by relying on the fact that they are population based and that their transition rules are probabilistic and not deterministic. Individuals often get stuck in local minima, but this does not mean that the GA is stuck, because there are other individuals searching the solution space. Also, the stochastic elements present in the GA can steer the search process to new directions, often out from local minima and towards global minima.

The population approach to search also facilitates the use of GAs in multi dimensional problems (problems with multiple design variables). The fact that GAs perform in Multi-dimensional and multimodal problems is associated with what is called search *robustness* (Goldberg 1989).

6.2 Exploration vs. Exploitation

In any search method, either computational or human search, there are two distinct processes at work: Exploration and Exploitation[†]. Exploration is the process responsible for covering the entire search space, include the vast majority of solutions in the search, and in the case of population based search to guarantee the preservation of diversity in the solution population. Exploitation is the process responsible for signaling out the best performing solutions, to direct the search process towards promising areas and generally reduce the search space by focusing on the best solutions.

These two processes can be considered opposite processes in that one increases the search area while the other reduces it. One preserves diversity, the other tries to focus on a small number of solutions that most often are quite similar to each other. It is precisely the contrast between these two processes that gives GAs their robustness. Most search algorithms with a deterministic transition rule, especially gradient based search algorithms, have very little exploratory power and concentrate on exploitation. This is one of the reasons why they have do not perform well in multimodal problems.

Exploratory and Exploitation power of GAs is significantly determined by its operators, and the parameters needed to control these operators. This

[†]a cognitive approach to exploration and exploitation is discussed in section 1.2

is why it is very important to get to know the operators and select the right ones.

6.3 A Genetic Algorithm Run

Genetic algorithms are made up of a series of operations performed to and with the population of candidate solutions. Coding and Scaling for example are necessary operations in a coded GA, but the most characteristic operations in GAs are the Selection or Reproduction operator, the Crossover operator and the Mutation operator. Some authors single out the mutation operator as being optional and not fundamental to the functioning of the GA, but others contend that they significantly improve the GAs efficiency.

Let's briefly look at the pseudocode of a simple GA, we will be using a binary coded GA:

```
1: Generate a random and coded population
2: for i ← 1, number of Generations do
3:     Decode and Scale the population
4:     Calculate fitness values for the population
5:     Run Selection operation
6:     Run Crossover operation
7:     Run Mutation operation
8:     Perform Exit condition test
9:     if Exit condition test = True then
10:        Exit GA
11:    else if Exit condition test = False then
12:        Continue to the next Generation
13:    end if
14: end for
```

In this section we will go through the Genetic Algorithm operations listed in the above code, and describe in detail its operators. For this purpose we will use a simple problem we will call Test problem A:

$$Test\ Problem\ A: \begin{cases} Maximize & f(x) = \frac{x_2}{x_1+0.1}, \\ subject\ to & 0 \leq x_1 \leq 1, \\ & 3 \leq x_2 \leq 5. \end{cases} \tag{6.1}$$

where x_1 and x_2 are the two problem variables, we can think of them for example as geometrical variables in a form search process, for now let's just think of them as variable numerical values. The object of this problem

is to find x_1 and x_2 values that maximize $f(x)$. We can also see that x_1 is confined to values between 0 and 1 and x_2 between 3 and 5, these are the domains of our two variables.

We will also be using a few GA specific elements that we need to properly run the GA, we can think of them as GA inputs. We will list them here and they will be explained further down. Some of these we already mentioned in the problem description:

Population Size (N)	6	
Number of Variables	2	
Number of binary digits	8 for x_1	4 for x_2
Variable Domains	$x_1 \in [0, 1]$	$x_2 \in [3, 5]$
Mutation Probability (p_m)	0.2	

6.3.1 Initial Population

In order for us create the initial population we need to use the number of binary digits in each variable. Binary digits define how the variable domain will be discretized. In this example we chose 8 binary digits for x_1, this means we will divide the x_1 domain (0 to 1) into 256 equal parts. This is because, a binary code 8 digits long contains 256 numbers. We chose 4 binary digits for x_2, so we divide its domain into 16 equal parts. Using this example we can compute how the discretization of the problem results.

We first calculate the domain length for each variable:

$$D_{len} = |D_{max} - D_{min}| \tag{6.2}$$

where D_{max} is the highest member of the domain and D_{min} is the lowest. In our example D_{len} is 1 for x_1 and 2 for x_2. We can now calculate the length of the discretized element for each variable:

$$L = \frac{D_{len}}{n} \tag{6.3}$$

where L is the length of the discretized element and n is the number of divisions or the amount of numbers in the binary digits we selected. In our example we get $L = 0.004$ for x_1 and $L = 0.125$ for x_2. We discretized x_1 in a much smaller element than we did x_2, because the domain was smaller for x_1, and most importantly we chose a higher number of digits for its discretization.

Search Complexity It is important to notice from this example that the number of possible solutions the GA will be considering is determined in these two numbers, the number of variables and their discretization. The higher number of variables or the more we discretize them, the higher number of possible solutions. We can relate the complexity of the *search* problem we pose to the GA to the number of possible solutions. In our example we have 16×256 possible solutions. This complexity can also serve us as a way to judge how many iterations of the GA are necessary to obtain an acceptable result.

Returning to the initial population, we can see that we need a binary number of $8+4$ digits for each individual (the chromosome of each solution). We generate the population using 12 random values (either 0 or 1). This is repeated this for the number of individuals in our population, in our example 6. The resulting population is the following:

Individual												
1	0	0	1	1	0	0	0	1	1	0	1	1
2	1	1	1	0	1	1	0	0	0	0	0	0
3	1	1	1	1	0	0	1	1	0	1	1	0
4	1	1	1	1	1	1	1	1	1	1	1	1
5	0	0	0	0	1	1	1	1	1	1	1	1
6	0	0	0	0	0	0	0	0	0	0	0	0

With this initial population the main loop of the GA can start.

6.3.2 Decoding and Scaling

We now need to decode and scale the binary numbers into values that we can use to calculate fitnesses. The decoding is done using the number of binary digits for each variable. In our example, the first 8 digits belong to the first variable and the other 4 belong to the second one. So let's take the chromosome of individual 1 and decode it. The chromosome is:

$$0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 1$$

If we divide it according to our coding scheme we get:

$$00110001 \quad 1011$$

We now decode these values into integers:

$$00110001 = 140 \quad 1011 = 13$$

Next we need to scale these values into the variable domains. This is done with the following equation:

$$S = D_{min} + (D_{max} - D_{min}) \times \frac{1}{M_{bin}} \times d \tag{6.4}$$

where S is the scaled value, M_{bin} is the maximum value obtainable with the number of binary digits of the value in study and d is the decoded value for the variable (140 and 13 in our first individual). In our example M_{bin} is 255 for the first variable and 15 for the second. This results in scaled values of 0.54 for x_1 and 4.73 for x_2. If we repeat this operation for the entire population we get these results:

Individual	x_1	x_2
1	0.54	4.73
2	0.21	3.00
3	0.81	3.80
4	1.00	5.00
5	0.94	5.00
6	0.00	3.00

This is the decoded and scaled population. We can now proceed to the next step in the GA which is fitness calculation.

6.3.3 Fitness Calculation

This is probably the step that need less explanation in the whole GA process. We simply follow the formula detailed in the problem definition (equation 6.1 in page 82). Of course in our example the fitness function is a very simple mathematical formula but we can think of this also as a building performance simulation or any other fitness function we can use to describe our search process. Following our formula and using our decoded and scaled values, we get a fitness for all of the individuals in our population:

Individual	Fitness
1	7.30
2	9.60
3	4.17
4	5.54
5	4.80
6	30.00

If we study the fitness formula and the results, we can see that in order to maximize the fitness we need a low x_1 and a high x_2 value. We can also see that the fitness is mostly sensitive to x_1, thus we made a good choice in investing 8 binary digits (and the added search complexity) to this variable, since we will be able to study it more in depth than x_2. In short we can already deduce that the best individual for this problem is:

x_2	x_2	code
0	5	000000001111

We will keep this code in mind when we study the rest of the GA operators and how they improve the fitness of the population.

6.3.4 Selection or Reproduction Operator

The Selection operator has three main functions:

- Identify good solutions from the population.
- Multiply those good solutions.
- Delete bad solutions from the population.

(Deb 2001)

The selection operator has the objective of choosing which of the individuals in the population will be used for reproduction. This group is called the mating pool. It is a very simple operation but it has a significant impact in the success of the GA.

From an exploration and exploitation point of view, the selection operator is most related to exploitation, but depending on the particular selection operator, they can have an influence on both. A selection operator that employs some randomness on the selection scheme is related to exploration, while selection operators that only focus on fitness to select individuals for reproduction are more related to exploitation. Different authors have proposed different selection operators that can accomplish these three tasks. Blickle and Thiele report the following list of selection operators and compare their functionality:

- Tournament Selection
- Truncation Selection

- Linear Ranking Selection
- Exponential Ranking Selection
- Proportionate Selection

(Blickle & Thiele 1995)

The Tournament selection operator uses randomly selected couples from the population to compete for a place in the mating pool. All of the members in the population are selected for two competitions with different members of the population. The competition in the couple is settled by means of their fitness values. This means that the best individual in the population will win both of its contests, and thus he will be copied twice in the mating pool. This process guarantees that there is a higher chance of good performing individuals to be selected for reproduction, increasing the exploitative power of the GA. However, since the couples are selected randomly, there is a little exploratory aspect to this particular brand of selection operator. For example, if all of the best performing individuals are coupled together, some of them will eliminate each other, leaving places open in the mating pool for not so well performing individuals. In a more exploitative selection operator, these low performing individuals would be cut out, but leaving them insures a certain level of diversity in the population, increasing exploration.

In the Proportionate selection operator the mating pool is filled in proportion to the fitness values of the individuals. If the average fitness value of the population is f_{avg} and the fitness value for the i_{th} individual is f_i, then the i_{th} individual would be expected to have a f_i/f_{avg} number of copies in the mating pool. This method is comparable to a tricked roulette wheel.

The Truncation Selection on the other hand employs no randomness in its selection. It simply sorts the population according to fitness, and selects the first individuals in the list using a user defined fraction of the population, for example 1/2. In this case there is no exploration added to the GA, exploration is left to the remaining operators.

Some selection operators require user defined parameters (for example the fraction of the population used in the truncation selection). As with other kinds of operators, the correct selection of this parameters is important for a correct GA run.

Let's use the tournament selection operator for our example GA and compute the mating pool. First we need to randomly select couples to compete with each other, twice. For example:

Tournament	Individual A		Individual B
1	3	vs.	4
2	1	vs.	6
3	2	vs.	5
4	4	vs.	5
5	1	vs.	2
6	3	vs.	6

If we run these tournaments, selecting as the winer the individual with the highest fitness (highest for a maximization problem like our example, lowest for a minimization problem), we get the following results:

Tournament	Winner
1	4
2	6
3	2
4	4
5	2
6	6

As we expected, the best individual in the population (individual 6) won both of its contests and has 2 copies in the mating pool. This was also the case for individuals 4 and 2, so we can say that individuals 1, 3 and 5 were eliminated from the GA. It's interesting to note that even though individual 1 had a higher fitness value than individual 4, it was eliminated and 4 went on. This is a product of the stochastic nature of the tournament selection operator. The resulting mating pool is the following vector:

Mating Pool	4	6	2	4	2	6

6.3.5 Crossover Operator

The crossover operator is responsible for creating a new population from the individuals present in the mating pool, it is the reproduction of the individuals in the mating pool or parent individuals and the generation of the offspring individuals or new generation. This operation is done by using the problem variables, the characteristics of the individuals. As we have seen, these variables are normally coded in some way, most commonly in binary numbers.

As it was previously mentioned in the introduction of this chapter, solutions are described in terms of their chromosomes. The crossover operator

simply copies part of the genes of one parent and the other part from the second parent, thus generating the chromosome for the child. This operation can be summarized in this way. We assume two parents A and B with the following chromosomes:

Parent A △△△△△△△△△△△△
Parent B □□□□□□□□□□□□

Now we need to specify a crossover point, the point in until which one string of genes will be used, the rest of the string will be taken from the second parent. This number is usually selected at random. Let's say we select the 8_{th} gene to be the last one in parent A. This means that the crossover operation will proceed as follows

	1 2 3 4 5 6 7 8	9 10 11 12		
Parent A	△△△△△△△△	△△△△	⟶	△△△△△△△△□□□□
Parent B	□□□□□□□□	□□□□		□□□□□□□□△△△△

The crossover generated two completely new individuals that we could call AB and BA:

Offspring AB △△△△△△△△□□□□
Offspring BA □□□□□□□□△△△△

The crossover operation supports any kind of coding, using binary code, using a larger alphabet (or rather an alphabet with a higher cardinality) to formulate the strings [‡] or even the use of real coded variables, meaning that the variables are represented numerically as real numbers.

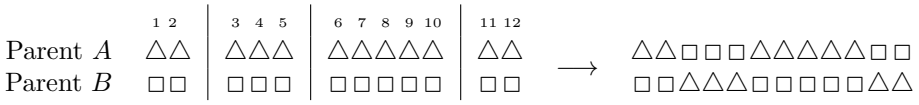
The above example employed the Simple Crossover operator. It is called the simple crossover operator because it employes only one crossover point. As we saw, this crossover point is normally selected at random, but the possible crossover points to select from need not be all gene positions. Some Crossover operators choose to only allow crossover points to be in between variables. If we for example have a coding scheme that attributes 2 genes for each variable, and there are 4 variables in the problem, then the chromosome of any individual would look like this:

	$variable1$	$variable2$	$variable3$	$variable4$
Individual	□ □	□ □	□ □	□ □

[‡]for a discussion on the effects of the cardinality of coding alphabets see (Deb 2001) pages 108-109.

In this example the only acceptable crossover point would be in between the variables, where the | lines are shown. This is especially useful in the case of real coded GAs. Otherwise there is the risk of allowing variables to go outside of their respective domains.

The other important type of crossover is the Multiple point crossover. As its name suggests it allows for the chromosome to be split at multiple positions, requiring multiple crossover points. Using the same two parents A and B we can show an example of this kind of crossover. In this case we will use 3 crossover points after the 2_{nd}, 5_{th} and 10_{th} genes:



Multiple point operators include a higher randomness to the GAs procedure, however, Goldberg (Goldberg 1989) explains that in the case of the Crossover, this increased randomness is not necessarily a good thing. He argues that increasing the braking points significantly improves the chances of breaking significant pieces of the chromosome called "schema". This makes the crossover more like a random shuffle of genes, and less like a planned creation of schemata.

Let us return to our example GA (Test Problem A). We will be using the simple crossover operator with variable crossing points. We will take the first two individuals in the mating pool and combine them to create two offspring individuals, then we will take the next two, and so on until we go trough the entire mating pool and we create a complete offspring population. We will begin with parents 1 and 2 from the mating pool (individuals 4 and 6):

Individual 4	11111111	1111	$\longrightarrow$	111111110000
Individual 6	00000000	0000		000000001111

The results of this crossover operation are quite interesting for our example. If we look at the second offspring in this operation, it is precisely the best possible individual we signaled out above (see page 86). Although there is some randomness involved in the operation, the fact that we ended up with this optimal individual is not completely by chance. It is the product of the genetic operator used until this point. The fact that the selection operator chose individuals that had 0s and 1s in the correct spaces to generate the correct variable values is not by chance. The selection operator chose them

because they had good fitness values, and then their combination yielded an even better result. On the other hand, if we look at the first result in this crossover, we notice that is exactly the opposite, it has highest x_1 value in the domain and lowest x_2 value in its domain. Hence we get the lowest possible fitness value. This is also a possible outcome in the GA. However this kind of bad result will get eliminated in the next generation and the population fitness will grow.

From an exploration vs. exploitation point of view, the crossover operator is involved in both processes. The recombination of genes from two individuals into a third one implies big changes in focus of the GA in the search space. The previous crossover example shows that the parents positions in the search space were replaced by completely new positions far way. This suggests that the exploration is at work in this operation. But as we just explained, this exploration is not entirely random, so we can see exploitation taking part in this operation as well.

If we finish the crossover for the rest of the population we obtain the following coded population:

offspring												
1	1	1	1	1	1	1	1	1	0	0	0	0
2	0	0	0	0	0	0	0	0	1	1	1	1
3	1	1	1	0	1	1	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	0	1	1	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0

Keep in mind that the crossover point is random, so not all crossover operations have the same result, even when the same individuals are involved.

6.3.6 Mutation Operator

The Mutation operator can be defined as a diversity preservation operator. It's role is mainly to increase the exploratory power of the GA by slightly altering a few genes in the chromosomes of the population. The number of genes that get altered is determined by the Mutation probability. This parameter is selected by the user as another GA input, and has different meanings in different mutation methods. the correct mutation operator for a GA must be selected depending on the coding scheme, we will first examine the simple mutation operator for binary coded GAs using offspring individual 3 from our example. For each gene in the individuals chromosome a random value $r \in [0, 1]$ is generated. The simple mutation operator

modifies a gene if this random value r is lower that the mutation probability p_m selected by the user. In our example the mutation probability p_m was set at 0.2. We will now run the simple mutation on the offspring individual 3:

Chromosome	1	1	1	0	1	1	0	0	0	0	0	0
Random Value r	.43	.56	.11	.95	.62	.22	.33	.80	.73	.59	.25	.44
Mutate	×	×	√	×	×	×	×	×	×	×	×	×
Result	1	1	0	0	1	1	0	0	0	0	0	0

We can see that of all of the genes in the chromosome, only one was modified because only one of the random values generated was lower than our selected mutation probability p_m (0.2). In the case of binary code, the modification of the gene is obvious, it its a 0 it gets turned into a 1 and vice versa.

Goldberg (Goldberg 1989) introduced a particular mutation operator that was computationally less expensive than the simple crossover, because it does not require the generation of a random number for each gene in the GA. This operator, called Mutation clock operator works by establishing a counter that reduces in number by each gene processed, and the next mutation occurs when the counter reaches 0. The length of the counter is different each time, its determined by a random value r, and it is inversely proportional to the mutation probability p_m. The higher p_m the shorter the counters will be. In this case r is only generated every time a new counter is needed, not at every gene.

When the coding scheme is not binary or even real coded, the modification of a gene is different. There are many different mutation operators for this coding schemes [§]. An example is the Non-Uniform Mutation in which the mutated gene is switched to a new value, but the value is selected by means of a probability curve. This curve is in the shape of a tent, and its highest at the previous gene value (the parent gene). In this way, the gene is likely to be modified by a value that is not that different from the parent gene.

Global and Local exploration There is a difference in exploration between the mutation and crossover operators. We saw earlier that the modifications involved in the crossover operation entailed big movements in the search space, we can call this *global* exploration. Mutation on the other hand tends to modify the chromosomes in small ways. The impact the mutation of

[§]for a review of mutation operators for real coded GAs see (Deb 2001) pages 122-124.

a single gene has on the individual depends on the point in which the change is made. For example, in binary coding, the first digits are reserved for high values, so changes in these digits have a higher impact than changes in the final digits. But even so, they would only make changes in one variable. While crossover operations may change quite radically, sometimes changing more than half of the genes. So we can say that the mutation operator is responsible for small movements in the search space or *local* exploration. The mixture of local and global exploration is also a reason for the GAs strength and robustness.

Another important aspect of the mutation operator is avoiding homogeneous populations. If we let the GA gain too much exploitative power, we run the risk of having a completely homogeneous population before the GA converges into a global optimal solution. The extreme case of this being a population made up of individuals who are identical. In this case, neither selection or crossover operators will help us regain diversity, and we will not see an increase in fitness, the GA is in fact stuck. This does not happen when we introduce mutation into the algorithm, and in particular, when a good mutation probability p_m value is selected. However, we must also be carful not to select a p_m value that is too high, because its might significantly reduce exploitation power. With a low p_m, schemata have very little probability of being affected by mutation (Goldberg 1989) so we can say that the effect of mutation is mostly beneficial if the correct p_m value is selected.

6.3.7 Elitism

Elitism can be though more as a concept than an actual operator, for it can be introduced into the GA in different ways. The point of elitism is to preserve a fraction of the best individuals in the population, in order to not let the maximum fitness value in each generation decrease. While the fitness tends to always increase generation by generation, in some kinds of problems, some stagnation can be found, and in this cases, genetic operators may even produce individuals with a lower fitness than their parents. Elitism is the process of keeping the parents in case they have better fitness that their children, thus keeping the fitness value constant in the next generation. This is a guarantee that a good solution is never lost in the GA operations unless a better solution is found. This is possible because GAs work with populations, we can afford to reserve a few "elite" spaces in the population to make sure fitness values don't decrease.

Elitism dramatically increases the efficiency and convergence speed of

GAs, especially un multi-objective problems (Deb 2001). A common way of introducing elitism into GAs is to directly copy the best α individuals directly into the next generations, while still having them participate into the selection process. Like all other GA input parameters, a good balance needs to be achieved in the selection of a α value. If we select an α value that is too high, we run the risk of loosing diversity in the population very quickly, and if we select a very low α we do not take advantage of having an elite population. Low α means higher exploration and high α means high exploitation. Values of $\alpha = 1$ and $\alpha = 0.1N$ (10% of the population) are common (Deb 2001).

6.3.8 End Conditions for GAs

Depending on the problem, there is no way of knowing when a GA has reached an optimal result, or if the GA can improve the fitness of the population by continuing to calculate. There are also no concrete rules of thumb for most of the GAs parameters such as population size or mutation probability. So deciding when to stop a GA is also a choice that is left to the judgment of the user. Never the less, at least one ending condition must be set, but multiple end conditions are also possible and in some cases desirable. There are three basic types of end conditions for GAs:

- Fitness related end conditions
- Number of calculations end conditions
- Time limit end conditions

When the problem is formulated in such a way that allows us to know a minimum acceptable fitness value, then it is very useful to have an end condition relating to fitness. For example the GA can end when the fitness value of the best individual in the population surpasses a user defined minimum fitness value. This avoids unnecessary further calculations and makes the GA more efficient. If the object of the search is not a single solution, the average of minimum fitness value in the entire population can be used in the fitness related end condition. Since there is no way of knowing whether the GA will ever reach the given minimum fitness, additional end conditions are usually given to complement this one (this is true of all fitness related end conditions). When there is no way of knowing a minimum acceptable fitness value, or such a value is not desired, other end conditions must be employed.

Another fitness related end condition is the maximum generations without improvement or maximum stagnant generations. This is useful when the user wants to achieve a certain result quickly and if the GA is not successful in a user defined number of generations, the user will reformulate the problem and restart the GA.

The maximum number of calculations can be estimated by setting a maximum number of generations plus the number of individuals in the population. It is clear that the more calculations the GA performs, the better its result will be (especially when there is an elitism component in the operators). It is then obvious that setting a number of calculations to perform is a valid end condition. This value can be set using the complexity of the problem as a guide, the more variables there are in the problem the more calculations we are going to need. But like it was said before, there is no exact way of knowing a perfect number of calculations.

When the problem at hand needs to be solved within a certain time window, then setting a maximum calculation time as an end condition is an obvious solution.

In our GA example, if we would have set a fitness related end condition the GA would have been over fairly quickly, but this is due to the simplicity of the problem we chose. If we would have set a maximum number of generations as an end condition, then the GA would not have been as quick. As we saw in the pseudocode, if the end condition would not have been met, then the loop would continue. In more complex problems than our example, the GA would continue to search both locally and globally for high performing solutions and storing valuable information in the process. This information can be accessed later on by the user as a sort of data mining of the GA progress, and in so doing, get a good understanding of the problem that was formulated.

7

Multi-Objective Search

7.1 Introduction

Now that we have seen the mechanics of the genetic algorithm (a single objective search method) we can look into the main issues of multi-objective search. In this chapter we will look into the theoretical aspects of multi-objective search. We will try to do so in a way that is independent from any specific search algorithm, and try to look more into the issues, possible outcomes and what to expect from a multi-objective search. The issues relating to specific multi-objective search algorithms will be dealt with in chapter 8.

7.2 Difference between single and multiple objectives

The obvious and most important difference between single objective, and multiple objective search is of course the number of objectives. But the implications of this difference do require some attention, especially when these multiple objectives are in contrast with one another. In order to describe this difference, let's make an example of a multi-objective problem with contrasting objectives. A typical case for is the contrast between the cost and the quality of an object. We can associate quality with the materials and manufacturing techniques of an object, but the better they are the more they tend to cost. In this chapter we will be using another mathematical test problem, in this case having two simple functions, described as follows:

$$Test\ Problem\ B: \begin{cases} Minimize & f_1(x) = x_1, \\ Minimize & f_2(x) = \frac{1+x_2}{x_1}, \\ subject\ to & 0.1 \leq x_1 \leq 1, \\ & 0 \leq x_2 \leq 1. \end{cases} \tag{7.1}$$

Test problem B states that we should minimize f_1 and f_2 which are both dependent of x_1 and x_2, and they both have a particular domain. A quick study of f_1 and f_2 reveals that the more we increase x_1 the higher the value for f_1 and the lower the value of f_2. This two functions are clearly contrasting each other. Moreover, we can say that the optimal value of f_1 is reached when $x_1 = 0.1$ and the optimal for f_2 is reached when $x_1 = 1$ and $x_2 = 1$. So if we had to minimize either f_1 or f_1 we would know what to do, but if we need to find minimal values for *both* f_1 and f_1, we do not know which x_1 and x_2 values to take. We cannot say that $x_1 = 1$ is optimal for both functions.

One way of solving this problem that is very common is to use a weighted sum of both functions. To a assign a weight factor for each function according to their relative importance. This kind of solution would in fact translate problem B into the following problem:

$$B\ (Weighted\ sum): \begin{cases} Minimize & f_w(x) = x_1 \cdot w_1 + (\frac{1+x_2}{x_1}) \cdot w_2, \\ subject\ to & 0.1 \leq x_1 \leq 1, \\ & 0 \leq x_2 \leq 1. \end{cases} \tag{7.2}$$

where w_1 and w_2 are the weight factors for each function. We could for example decide that f_1 and f_2 are equally important to us and so we give a value of 0.5 to both w_1 and w_2. In this way we are simply translating f_1 and f_2 into a single function f_w, translating the multiple objective problem into a single objective problem. We would then proceed to use a single objective search algorithm and find a single optimal solution to the problem. This approach is perfectly valid when we have information that can give us the correct weights for each function. However, in most real world problems this is not the case.

Let's discuss some of the real world objectives that are addressed in this research, like the multi-disciplinary problems discussed in chapter 19. Can we say that the structural efficiency in a building envelope is more important that its environmental quality and energy efficiency? Can we say this for all buildings? Can we say this for any building? And perhaps more importantly, can we say in which measure one is more important that the

other? Is using a 50-50 value a valid approach? The acoustic performance of a room compared to the structural capacity is also no easy matter to solve. Keeping in mind that the weighted sum approach only works if we assign correct and exact weights to all functions, in such problems using this approach would lead to arbitrary results that are of little interest.

7.3 The concept of Dominance

Going back to test problem B, we previously stated that we could not say which values of x_1 and x_2 are optimal for both f_1 and f_2. But there is something we can say about some values of x_1 and x_2 with respect to others, and it is the concept of domination. Let's start by comparing a set of solutions:

Solution	x_1	x_2	$f_1(x)$	$f_2(x)$	color
1	0.1	0	0.1	10	●
2	1	0	1	1	●
3	1	1	1	2	●

Solution 1 is the minimal(best) solution for f_1 and the maximum(worst) for f_2. Solution 2 is the exact opposite, the maximum for f_1 and the minimal for f_2. If we compare this two functions we cannot say that one is better that the other if we consider both functions, we can say that neither of these solutions dominates the other.

If we compare solutions 2 and 3 however, we can see they have equal results for f_1, but solution 2 outperforms solution 3 in f_2. In this case we can say that solution 2 *dominates* solution 3.

If we generalize this concept, we can say:

- In order for solution A to dominate solution B, solution A has to outperform, or equal B in all functions, as well as outperform B in at least one function.
- If solution A outperforms or equals solution B in all objective functions except in one in which solution B outperforms A, then A and B do not dominate each other.

To continue studying these relationships, we will introduce the search and objective spaces, as a way of visualizing these solutions in their variable domains and their function results.

7.4 Search Space and Objective Space

Multi-objective search problems are often analyzed by looking into two separate domains, called the search or decision variable space and the objective or function space.

The *search space* is a representation of the proposed solutions described by their variable values. Their variable values are mapped in such a way as to showcase their values and the relative distances and similarities between them. The search space has as many dimensions as the problem has variables, and this dimensions are confined to their respective variable domains.

The *objective space* on the other hand, is a representation of the solutions from the point of view of their objective function values. They are mapped out according to the values obtained in each objective function present in the multi-objective problem. Hence, the objective space has as many dimensions as the problem has objective functions. The confines of the objective space are determined by the objective functions, they depend on the values obtained by them.

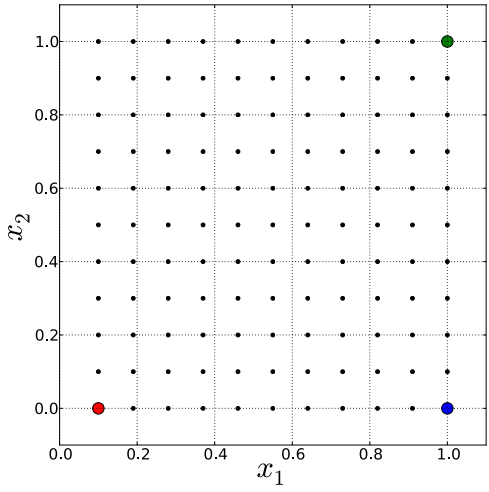
Figure 7.1a shows the search spaces for our test problem B and 7.1b shows the objective space for the same problem. The search space shows x_1 and x_2 values selected in a grid, this is done so that we can get a good idea of how the entire set of possible x_1 and x_2 values end up in the objective space. In both spaces solution 1 is represented with a red dot, solution 2 is in blue, and solution 3 in green. The rest of the values are represented with small black dots.

This way of representing the problem shows very clearly the relationship between the two functions. One very important feature to focus on is the contrast between the two functions *. This contrast can be seen in the objective space (figure 7.1b) by the fact that when solutions approach a minimal value for f_1 they loose optimality for f_2. Another important feature that can be seen in the objective space is the Pareto Front.

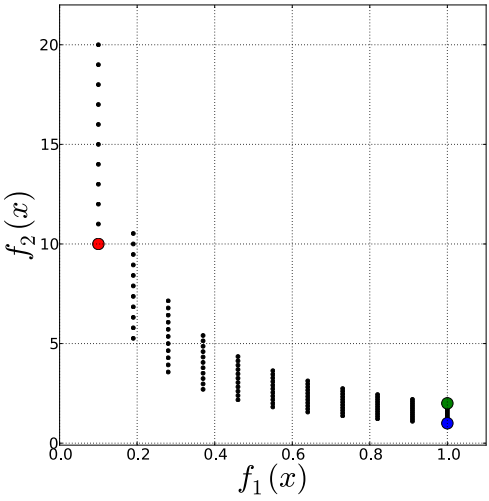
7.5 The Pareto Front

In the previous section we explained the concept of domination, we will now explain what is the Pareto Front, starting by the concept of a non-dominated individual. A non-dominated individual is one that is not dominated by any other individual in the population. A non-dominated individual typically

*A more detailed study of contrasting objectives is given in section 7.6.



(a) Search Space



(b) Objective Space

Figure 7.1: Search and Objective Space for Test problem B

dominates many of the other individuals in the population, and it is never dominated by others. A non-dominated individual may have many individuals which he does not dominate, but none that dominate him.

If we look back at individuals 1,2,3 in test problem B, we can see the following relationships:

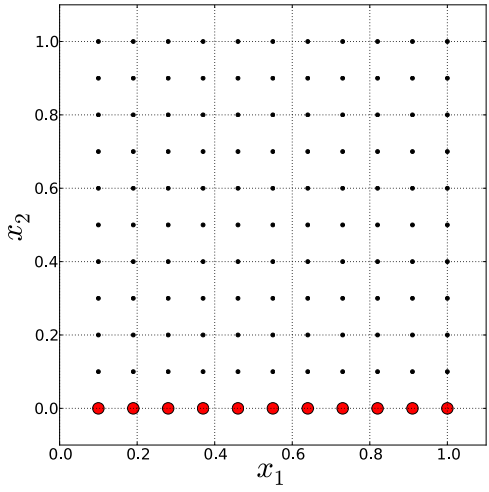
- Individuals 1 and 2 do not dominate each other
- Individual 2 dominates individual 3
- Individuals 1 and 3 do not dominate each other

Therefore we can say that only individuals 1 and 2 are non-dominated. Although individual 1 does not dominate individual 3 (as does individual 2) individual 1 is never dominated by any other solution, therefore it is non-dominated.

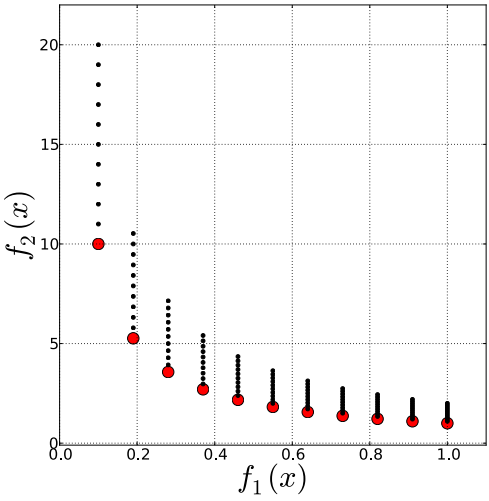
The *Pareto Front*, also called trade-off set or non-dominated set, is the set of all non-dominated solutions in a given group. They represent the set of solutions that cannot be said to be better from each other if we consider all objective functions in the problem. In our example, if we only considered solutions 1,2,3, then solutions 1 and 2 would comprise the Pareto front. But if the consider any other possible combination of x_1 and x_2 that we see in figure 7.1a, then the Pareto set would be made up of a few more individuals.

Figure 7.2 shows the entire Pareto front for test problem B, in the search and objective spaces. We can see that individuals 1 and 2 are still present in the Pareto Front and that individual 3 is not. Moreover we see a series of individuals in the objective space forming a curve that is convex towards the ideal point (in the case of a problem with two minimization functions this ideal point would have coordinates 0,0). We discretized the search space into 10 equal segments in x_1 (11 points) and 10 equal segments in x_2 (also 11 points) for a total of 121 points analyzed. If we had used an infinite number of points, the Pareto front would be a continuous curve representing all of the solutions that are non-dominated. However, since we only studied 121 solutions, we only obtained a part of the entire Pareto front. We can say that the individuals we found "lie" in the Pareto Front, and they are certainly part of it, but we cannot say that they comprise the entire non-dominated front.

The red dots in figure 7.2 represent the non-dominated set in the entire population, so we can say that the population could be divided into two groups, the non-dominated and the dominated. But if we wanted to divide the population (from a Pareto dominance point of view) into more



(a) Search Space



(b) Objective Space

Figure 7.2: Search and Objective Space for Test problem B with Pareto front in Red.

groups than just two groups, we can introduce more domination related sets. The Red dots can be considered as the first in a series of Pareto fronts. The second Pareto front would be made up of those individuals that are only dominated by individuals in the first front. If we eliminated the individuals of the first Pareto front from the population and recalculated the non-dominated set, we would in fact obtain the individuals of the second front. We can continue this subdivision into sequential fronts until all of the individuals of the population are part of one front, thus obtaining a series of sets that are "non-dominated sorted"[†].

Pareto fronts can have many different shapes, depending of course on the objective functions analyzed. As was previously mentioned, the objective space has as many dimensions as the problem it studies has objective functions, so this means we can only graphically represent objective spaces of up to 3 objective functions. But this does not mean that we cannot calculate of find Pareto fronts that have 4 dimensions or more, we simply need to represent them in a different way, numerically or through $2D$ or $3D$ "sections" in a multi dimensional space. This representational difficulty aside, the study of the shape of the Pareto front is an interesting way of studying the problem. Among the interesting features of such analysis, we can mention the discontinuity of a Pareto set, the degree of contrast (or lack there of) between objective functions and the relative dimension of the front with respect to the entire domain of feasible solutions.

Another important study is done in the position of the Pareto individuals in the search space. Figure 7.2a shows that the Pareto set in our test problem B lies entirely at the bottom of the search space and that it forms a continuous area of the space. But this is not necessarily the case, as we will see later on, the position of Pareto optimal individuals are often non continuous. This means that finding the entire set would be a multi-modal problem.

The extreme points in the Pareto set are interesting individuals, as they represent the optimal solutions for one of the objective functions. In our example, these extremes are individuals 1 and 2. We could expect the two extremes of a Pareto front to be in quite different positions in the search space, and this is commonly the case, but as we mentioned, there is not always continuity between search space and objective space. We could find individuals that are next to each other in the Pareto front, and that are very far apart in the search space. For this reason, it is considered an important

[†]for a more in depth analysis of the non-dominated sorting procedure see section 8.3.1 bellow.

element of a Pareto analysis to find a uniform distribution of solutions in the Pareto front, so that we can then get a good idea of where the entire Pareto set is in the search space. To get a good idea of the *diversity* of the solutions that cannot be said to dominate each other.

Pareto fronts can be studied independently of whether a problem has minimization or maximization functions or even mixed functions. The only difference in this case is the position of the ideal or optimal point in the objective space. As we saw above, if the problem is has both minimization functions, the optimal corner is the lower-left corner, max-max problems have it in the upper-right corner, and so on.

7.6 Contrasting Objectives

In this section we will study the contrast between objective functions from a theoretical point of view. To help us understand these issues we will be looking at the objective space of a series of problems, and most importantly the shape of the Pareto fronts in these cases.

We should perhaps start by talking about non contrasting objectives. Not all multi-objective problems have contrasting objectives, this of course depends on the objective functions we study. Sometimes we do not know exactly how these functions relate to one another before we analyze them. It is for this reason that is quite useful to study Pareto front shapes. Let's start by seeing what the objective space of a couple of non contrasting functions looks like. First we describe the problem, in this case we will call it Test Problem C:

$$Test\ Problem\ C \begin{cases} Maximize & f_1(x) = x_1 + x_2^2, \\ Maximize & f_2(x) = x_1 \cdot x_2, \\ subject\ to & 0 \leq x_1 \leq 5, \\ & 0 \leq x_2 \leq 1. \end{cases} \tag{7.3}$$

In this case we see a problem in which both functions need to be maximized. A quick view of the objective functions of problem C reveals that the more we increase the values for x_1 and x_2 we optimize both functions. So we have a set of non contrasting functions.

Figure 7.3 shows the objective space for problem C. We see that the Pareto front is comprised of a single solution (represented by the red dot). This is due to the fact that there is no contrast in the two functions. The combination of x_1 and x_2 that maximizes f_1 also maximizes x_2. Hence, this solution dominates all others in the population, and it is the only one in the

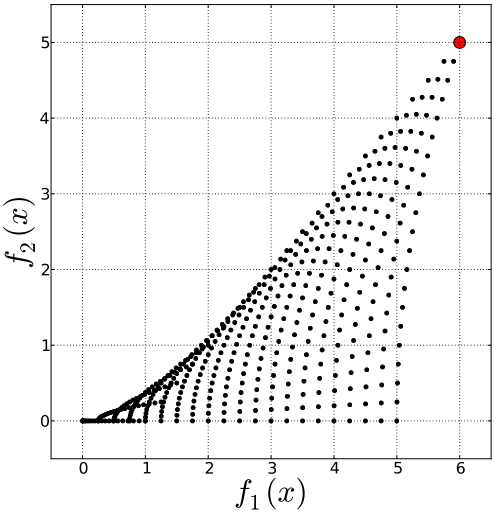


Figure 7.3: Objective Space for Test problem C.

front. The extreme case of a non contrasting problem would be one in which all functions result in equal values. If we draw the objective space for a two function problem in which $f_1 = f_2$ we would get individuals that all lie in a 45° line with a one solution Pareto front at the end of the line. In cases like this there is no need for a multi-objective approach to the problem, because we could just optimize one of the functions and find the same individual that we would find if we optimized the other. In such cases the multi-objective approach has no additional information to offer than the one provided by a single-objective search process.

Let's now move on to more contrasted problems. To do so we will study two more test problems, D and E. Test problem D is defined as follows:

$$Test\ Problem\ D \begin{cases} Maximize & f_1(x) = x_1 - (x_2)^a, \\ Maximize & f_2(x) = x_2 - (x_1)^a, \\ subject\ to & 0 \leq x_1 \leq 1, \\ & 0 \leq x_2 \leq 1. \end{cases} \tag{7.4}$$

where a is a constant that we will use to transform the level of contrast

of the problem. Looking at f_1 and f_2 for problem D we can already say that they are contrasting equations. Furthermore we can say that the end result of the search process in problem D depends greatly on the value given to a.

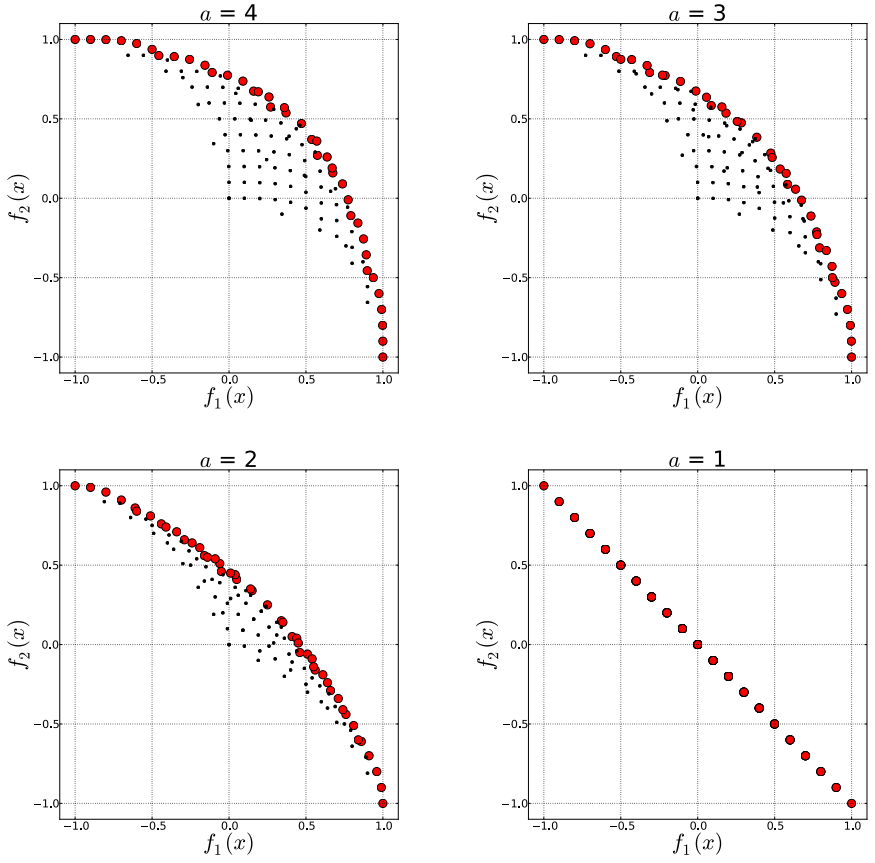


Figure 7.4: Objective Spaces for Test problem D with varying values for a.

Figure 7.4 shows the objective space for four versions of problem D, where $a = 4, a = 3, a = 2$ and $a = 1$. When $a = 1$ the Pareto front forms a 45° line that is perpendicular to the ideal point (in this case upper right corner). We can also see that as the value of a increases, the Pareto front becomes an

increasingly convex curve (convex towards the ideal point). A convex curved Pareto front denotes contrast in the objective functions, because increasing the value of f_1 necessarily means decreasing f_2. But we can study the rate in which f_2 decreases as we increase f_1, and use this rate to compare the contrast between these four Pareto fronts. In this sense we can say that the higher the rate of decrease, the higher the contrast. Of course the rate of decrease in these curves is variable, but we will look at the rate of decrease at the midpoint.

Looking at figure 7.4 we can see that in all cases the Pareto Front goes from a f_1 value of -1 (the worst value) to an optimal value of +1. The same is true for f_2. Let's start by looking at the rate in which f_2 decreases as we increase f_1 in the case of $a = 4$. When f_1 is at -1 we get $f_2 = 1$, but when we increase f_1 to 0, $f_2 = 0.75$. So from -1 to 0, we only lost 0.25, we have a rate of descent of 25%.

Now, let's make the same observation for $a = 2$. We can see that the point where $f_1 = 0$ corresponds to an $f_2 = 0.50$, A rate of descent of 50%. So we can clearly see that the rate in descent is higher in $a = 2$. Following the same logic, we can say that among these four curves, the highest rate is that of $a = 1$, in which, the rate of descent is 100%. We can conclude that convex Pareto fronts are a sign of contrasted problems, but also that they are more and more contrasted as they become less convex, and more linear.

But what happens when the Pareto front becomes concave? To study concave Pareto fronts we will use test problem E described as follows:

$$Test\ Problem\ E \begin{cases} Maximize & f_1(x) = \frac{x_1}{x_2}, \\ Maximize & f_2(x) = \frac{x_2}{x_1}, \\ subject\ to & 0.1 \leq x_1 \leq 1, \\ & 0.1 \leq x_2 \leq 1. \end{cases} \tag{7.5}$$

The resulting Pareto front is shown in figure 7.5. The rate of descent at the midpoint in this case is close to 198%, which is higher than the one we found in the linear Pareto front (100%). In the case of concave fronts, the more concave they are, the higher the rate and the higher the contrast.

To summarize this section, non contrasted objectives are identifiable by a Pareto front made up of a single solution. Contrasting objectives are shown by fronts with more than one solution. The degree of contrast of the functions can be determined by the shape of the front, convex is a low contrast, linear is a constant contrast, concave is a high contrast. Pareto fronts may be a combination of shapes, they may contain linear, convex and concave parts. They might also be discontinuous. But we can say that

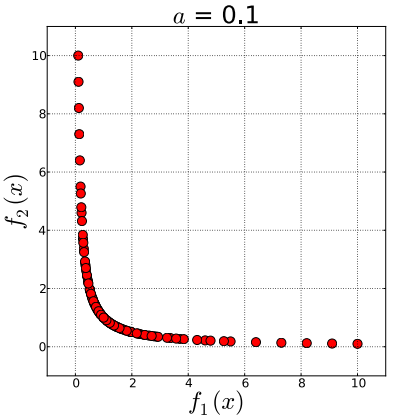


Figure 7.5: Objective Space for Test problem E.

the behavior that we have studied will be true for the parts of the front corresponding to the shapes we have just mentioned.

7.7 Final Selection Criteria

So far we have learned that in a multi-objective problem, there is no way of determining which is the single individual that can best satisfy *all* objective functions. But in real world multi-objective problems, we will have to select a single solution for the problem. If for example, the problem we face regards the structural vs. acoustical properties of a building, we cannot build all of the solutions in the Pareto front. The Pareto front represents a very useful tool in reducing the number of alternatives in the search space, to a limited number, but we still need to choose a single solution out of the front. This section tries to address this issue and talk about criteria for making a final selection.

Pareto fronts describe only the functions we introduce into the problem, so the first alternative to select a solution from the Pareto front is to use information that was not introduced in the problem. This is what we have described in the first part of this thesis as implicit search goals and what Deb calls higher-level information (Deb 2001). We can use functions that were not present in the original problem, or we can use information that is

possible to introduce in the problem.

Not all functions are introduced in multi-objective problems for a series of reasons. Sometimes the phenomenon these functions describe are not as relevant as the other functions, or sometimes the functions are computationally too expensive to be used in a search process. The convective thermal exchange in complex geometry for example is computationally very expensive to calculate, sometimes requiring more than a day. Ideally we would not need to run this calculation more that a few times.

There are other factors about the building that we usually do not introduce into search processes that can help us select from the Pareto front. The one that comes first in mind is an aesthetic judgement of the solutions. Search processes are quite useful in gathering information about the design object that can be measured, calculated or simulated. Typically, we use it to inform us on physical phenomena occurring in the building, its structural capabilities, acoustical or lighting quality of the spaces or thermal exchanges. Also economical and construction issues can be measured or simulated, but aesthetic issues are not. Search algorithms can mostly help us with the "tame" parts of the design process, and provide us information that can help us deal with the "wicked" parts[‡]. The importance of designer interaction in search processes was discussed in section 1.7, and the selection of a final solution is an important moment of interaction.

There is also the alternative of further refining the Pareto individuals with the use of Data Mining techniques. If we have Pareto sets made up of too many alternatives for a designer to consider, we can also use clustering and classification algorithms that can further reduce the number of options. We can look at the distance in the Search space (the variable differences) as a way of differencing individuals in the front. We know that a good multi-objective search process will produce a set of well distributed individuals in the Pareto front, so there is already a bit of data mining involved. But the solutions in the set may come from any part of the search space, and this information is not present in the Pareto front. A useful refinement of the Pareto set would be the separation of solutions by search space distance, to signal out solutions that are most different form a variable point of view. As is the case in the studies shown in this research, this variables represent geometrical features of the solutions. So we would be sorting solutions that have the shapes that differ the most.

In this frame of mind, diversity is an important issue. We want to keep diverse population from a variable (search space) point of view. Following

[‡]For an explanation of "tame" and "wicked" problems see section 1.3.

this logic it makes good sense to keep a record of all of the Pareto fronts found during the process. A search history that can help the designer interact and consider near-optimal solutions, that are perhaps very meaningful from a geometry point of view. We can think of this as a different kind of exploitation, a designer involved directly in the exploitation process can yield important results.

8

NSGA-II

8.1 Introduction

NSGA stands for Non-dominated Sorting Genetic Algorithm. After a first non elitist Multi-Objective GA, NSGA-II was developed by Kalyanmoy Deb and his students in 2000 as an elitist version of NSGA (Deb 2001). NSGA-II has been employed successfully in many architecture and construction related problems.*

Being a genetic algorithm, NSGA-II shares the same overall GA dynamic that was explained above. There is a main loop that iterates generation by generation, there is fitness evaluation (in this case we have multiple fitnesses) and there are selection, crossover and mutation operators. These operators however are specially designed to work in multi-objective problems. In addition to these modifications, NSGA-II has two special operators that will be described bellow.

As it was outlined in chapter 7, the final output of the multi-objective search process is not a single solution. NSGA-II was designed to obtain a set of solutions evenly distributed in the Pareto front, taking full advantage of the fact that GAs work with populations of solutions. This solutions will be evaluated for their dominance within the population, and they will also be evaluated for their distribution along the Pareto front.

An important characteristic of NSGA-II is that it works with a population that will change in size during the procedure. If we select an initial

*see for example (Attia et al. 2013), in this article there is a complete review of optimization tools used for building performance optimization. A part of the article is devoted to the use of NSGA-II in this field.

population of size N, during the course of the main loop, this population will be doubled to $2N$ and then taken back to N. In this way NSGA-II considers a *parent* population and a *current* population in the same loop, thus providing the possibility of elitism.

8.2 The NSGA-II procedure

Figure 8.1 is the flow chart for NSGA- II. We can see the general procedure is not very different ftom that of the simple GA. The main loop is enclosed in the segmented rectangle, the population size at changing points is signaled in red, and the operations that most differ from the GA are signaled in yellow.

While in the single-objective GA the fitness values were one for each individual in the population, in NSGA-II we have a matrix of fitness values. The matrix has a dimension that is equal to the number of individuals in the population times the number of fitness functions in the problem. The fitness calculation is basically done in a double loop, as follows:

1: **for** $i \leftarrow 1, numPopulation$ **do**
2: **for** $j \leftarrow 1, numFitnessFunctions$ **do**
3: $FitnessValues\ [i, j] = f_j(x)$
4: **end for**
5: **end for**

In this way the entire population is calculated for $f_1, f_2, f_3, ..., f_n$, and the fitness values are stored in the matrix $FitnessValues$.

We can also describe the NSGA-II algorithm with the following Pseudocode:

1: Generate a random and coded *Initial* population with n number of individuals
2: Decode and Scale *Initial* population
3: Calculate fitness values for *Initial* population
4: Copy resulting fitness values to *Parent* population vector
5: Generate a random and coded *Current* population with n number of individuals
6: **for** $i \leftarrow 1$, number of Generations **do**
7: Decode and Scale *Current* population
8: Calculate fitness values for *Current* population
9: Combine *Parent* and *Current* creating a vector of size $2n$

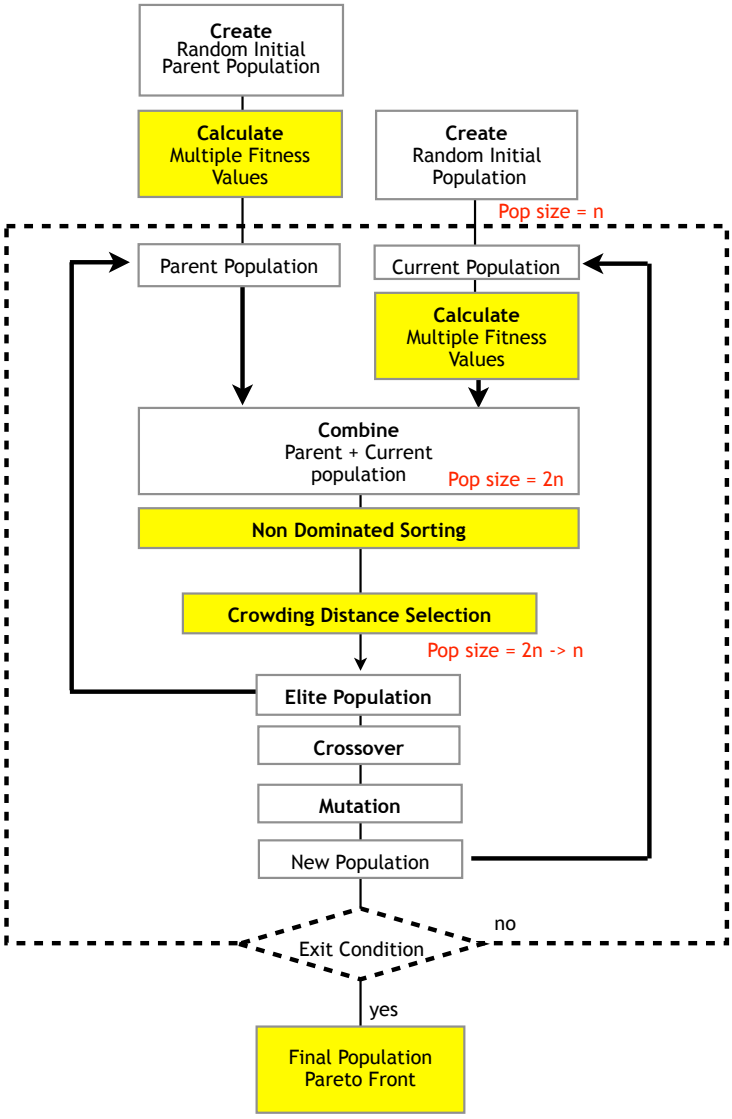


Figure 8.1: NSGA-II Flow chart

10: Run Non Dominated Sorting Algorithm for the combined population
11: Run Crowding distance Algorithm for the combined population
12: Replace *Parent* population with first n individuals from the combined population according to the NDS and CD results
13: Run Tournament Selection operation
14: Run Crossover operation
15: Run Mutation operation
16: Use the resulting population to replace *Current* population
17: Perform Exit condition test
18: **if** Exit condition test = `True` **then**
19: `Exit` NSGA-II
20: **else if** Exit condition test = `False` **then**
21: Continue to the next Generation
22: **end if**
23: **end for**

In steps 1 trough 4 of the above pseudocode we generate, decode, scale and calculate fitnesses for the initial population that will serve as the first *parent* population. Since we are only at the beginning of the Multi-Objective Genetic Algorithm (MOGA) we do not yet have a parent population so we need to create one randomly and have ready its fitness values before entering the main loop. Also necessary before entering the main loop is the generation of the random *current* population, seen in step 4.

Step 6 marks the beginning of the main loop, and the first thing we do in it is to decode and scale the randomly generated current population (step 7). We calculate fitness values for the current population, and we combine it with the parent population, thus forming the *combined* and changing the number of individuals in the population to $2N$. The combined population is further studied by the non-Dominated sorting and crowding distance operators. These special operators will be described in section 8.3. Their main function is to provide information necessary to use the selection operator. Since we have many fitness functions, selection cannot be done by simply selecting the individuals with the highest fitness (as was explained in chapter 7). Selection in NSGA-II is done by using the sequential Pareto fronts, the non-dominated individuals are preferred to the dominated ones. Further refinement is done in the crowding distance to signal out the best distributed individuals on these fronts.

During selection, we take the best half of these individuals, they will replace the parent population such as they are, thus preventing the elite individuals from being lost in further operations. We will use these same elite

individuals for crossover and mutation operators. During selection we return to a population size of N. We perform crossover and mutation operations to create the new population, the offspring population. We replace the current population with this new population and we conclude the operations in the loop.

8.3 Special Operators

All Multi-objective search algorithms (genetic or otherwise) need to incorporate special operators that allow them to work with multiple fitnesses. In comparison with the normal GA, NSGA-II has a series of modifications to its operators, most importantly its selection operator. In particular NSGA-II does not use the fitness values directly to select the best individuals and consequently the individuals who will be used for reproduction. As its name suggests, NSGA-II uses a Non-dominated Sorting (NDS) algorithm to asses the position of all solutions in the objective space, to sort them according to Pareto fronts. Additionally, NSGA-II uses a special *diversity preservation* algorithm called Crowding Distance (CD). It is the combination of the values obtained with the NDS and CD algorithms that NSGA-II selects its best individuals.

In this section we will go trough these two algorithms in detail in order to better understand the way NSGA-II works.

8.3.1 Non Dominated Sorting

The Non-dominated Sorting algorithm starts simply by comparing all of the individuals in the population. The comparison is done in all of the different fitness values present in the problem. Each fitness value can come from different types of problems (maximization or minimization).

When comparing individuals in the population, the algorithm has to produce two sets of information:

- The first one is a *Domination Count*. For each individual in the population, the algorithm has to count how many other individuals dominate it. This information is stored in a single vector that is as long as the population of individuals, called *DominationCount*.
- The second is a *Dominated set* of each individual in the population. This means that for each individual in the population, the algorithm must keep a list of all those other individuals which are dominated

by it. This information is kept in two vectors. The first vector called *DominatedSet* contains all of the dominated individuals, the dominating individuals that correspond to the dominated set are kept in the second vector called *DominatingIndividuals*.

The following example can help to clarify the method. Figure 8.2 represents the objective space of a two objective problem f_1 and f_2. On the X axis we see fitness values for f_1 and on the Y axis the values for f_2. Both functions are minimization functions. The numbers in the figure (1 to 9) represent individual solutions in our population, and we need to sort these individuals into non-dominated sets (a sequence of Pareto Fronts).

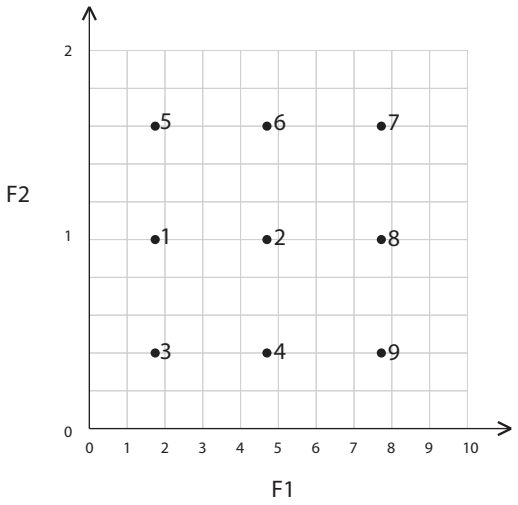


Figure 8.2: Example Objective space for f_1 and f_2

In this case the *DominationCount* vector would be the following:

Individual	1	2	3	4	5	6	7	8	9
DominationCount	1	3	0	1	3	5	8	5	2

The *DominatedSet* and the *DominatingIndividuals* vectors would be the following:

Dominated Set	1111122233333333444445568 99
Dominating Individuals	5672867812456789267896777 78

Once we have all of this information memorized, the NDS algorithm has to use these three vectors to define and memorize the Pareto fronts. We first define the way we will memorize the fronts. They will be copied in two vectors, one of them called *ParetoFrontIndividuals* and it contains all of the individuals of the populated, organized by Pareto fronts, the other one is called *ParetoFrontIndexes* and it contains the indexes in which a particular front ends. In this example:

ParetoFrontIndividuals	3	1	4	2	5	9
ParetoFrontIndexes	0	2	5			

The first front is comprised solely by the individual 3, because the index indicates that the first front ends at the 0 index, the second front goes from the 1 to the 2 index, so it comprises the 1 and the 4 individuals, and the third front goes from the 3 to the 5 index (the 2 the 5 and the 9 individuals). This result is shown in figure 8.3.

To create the Pareto fronts, the algorithm has to perform the following procedure explained in pseudocode:

```
1: ParetoFrontNumber = 0
2: while length(ParetoFrontIndividuals) < numPopulation do
3:     for i ← 1, numPopulation do
4:         if DominationCount = 0 then
5:             ParetoFrontIndividuals.append(i)
6:         end if
7:         ParetoFrontIndexes.append(length(ParetoFrontIndividuals)−
   1)
8:     end for
9:     ParetoFrontNumber = ParetoFrontNumber + 1
10:    for all k ∈ ParetoFrontNumber do
11:        for h ← 1, length(DominatedSet do
12:            if DominatingIndividuals[k] = ParetoFrontIndividual[h]
   then
13:                if DominationCount[ParetoFrontIndividuals[k]] => 0
   then
14:                    DominationCount = DominationCount − 1
```

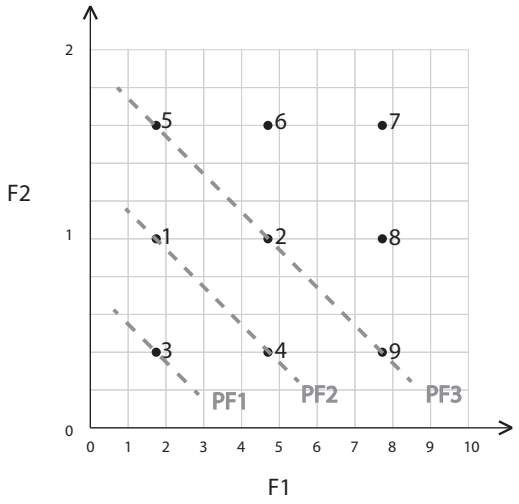


Figure 8.3: Example Objective space for f_1 and f_2 - Solved for the first 3 Pareto Fronts

15: **end if**
16: **end if**
17: **end for**
18: **end for**
19: **end while**

After this algorithm we have the whole population divided into sequential Pareto fronts. These fronts are then going to be used by the selection operator, along with the results of the crowding distance algorithm.

8.3.2 Crowding Distance

The crowding distance algorithm has the responsibility of determining which solutions are most dissimilar from each other from the point of view of all of their fitness values. This guarantees that the end result of NSGA-II will be a set of solutions well distributed along the entire Pareto front. In genetic terms, this operation is often described as a niching operation. It bears resemblance to the clustering operations common in data mining techniques, in that the Euclidean distance between individuals is used to determine their

similarity.

For a given set of solutions F, and a given set of fitness values m the crowding distance for all individuals $f_{i,m}$ in the set can be calculated by the use of the following algorithm:

1: **for** $m \leftarrow 1$, number of Fitness Functions **do**
2: $f_m^{max} = max_m(F)$ ▷ we find the maximum fitness m in F
3: $f_m^{max} = min_m(F)$ ▷ we find the minimum fitness m in F
4: create vector I_m so that $I_m = sort(f_m, >)$
5: $d_{I_{first}^m} = \infty$ ▷ we assign an infinite CD for the first individual in I_m
6: $d_{I_{last}^m} = \infty$ ▷ we assign an infinite CD for the last individual in I_m
7: $d_{I_j^m} = 0$ ▷ we first assign a CD of 0 to all individuals
8: **for** $j \leftarrow 1$, number of individuals in I_m **do**
9: calculate and assign $d_{I_j^m}$ using equation 8.1
10: **end for**
11: **end for**

$$d_{I_j^m} = d_{I_j^m} + \frac{f_m^{(I_{j+1}^m)} - f_m^{(I_{j-1}^m)}}{f_m^{max} - f_m^{min}} \tag{8.1}$$

As we can see in steps 2 and 3 the maximum and minimum values in each particular fitnesses are signaled out. They are very important in that they represent the maximum distance between solutions in that fitness function (in that dimension). This maximum distance serves for a sort of normalization operation that is done in equation 8.1 as we can see by their presence in the denominator. All other distances are compared after being normalized with the maximum distance.

Extreme solutions in each function are protected. The extreme cases serve to mark the end point of the Pareto front, hence they are given an infinite distance to allow them to pass the subsequent genetic selection operator. In that selection operator the individuals with the highest CD will be considered superior to those with lower CD. The total crowding distance of each individual is equal to the sum of all of its distances in all dimensions.

Once we have both the crowding distance and the Pareto fronts, we have all we need to employ our NSGA-II selection operator.

8.3.3 NSGA-II End Conditions

The exit conditions for NSGA-II, as for the simple GA, can be many combined and they can be varied in nature. End conditions relating to number

of calculations such as number of generations, and end conditions relating to time are valid alternatives. Fitness related end conditions on the other hand make much less sense when compared to the simple GA. In the simple GA we can establish that the algorithm should stop when the fitness function reaches a minimum acceptable value. This is not a good solution in the case of multiple and contrasting objectives, because we might reach that value without obtaining a complete and well populated Pareto front. A maximum number of stagnant generations can be used to stop a MOGA that is not evolving properly, however, stagnation is far less likely when we have many different functions. Also in many cases the MOGA can remain stagnant but still work to better populate the Pareto front.

After these considerations we can see that perhaps the most common and sensible end condition for a MOGA is a maximum number of generations or a maximum calculation time.

8.4 NSGA-II Python Implementation

NSGA-II was implemented into Python for this research. In this section we briefly describe the programing details of the implementation.

Since the main objective of this research is to use MOGAs to search for high-performing architectural shapes in many disciplines, the MOGA had to remain completely independent of the fitness functions employed. A function neutral NSGA-II implementation was thus created, following the following structure:

The Main file contains all of the user related inputs, it is the only file that needs editing to make different MOGA runs. Fitness functions are selected in this main file, as well as all of the genetic variables such as population size, binary string length and mutation probability. In this main file, an instantiation of the NSGA-II class is created, and trough it we pass the all of the above mentioned input data to the MOGA.

The NSGA-II class contains all of the necessary functions to run the MOGA, all of the operators are separated into functions, thus allowing us to replace a given operator with another almost effortlessly. The main loop of the MOGA is contained in the most important function in the class. This function is called Multioptimize, and it follows the pseudocode described in page 112. It is considered the most important function because it is the only function that is called directly in the Main file, and because it is the one calling all of the genetic operator functions.

The Multioptimize function shares information directly with the fitness

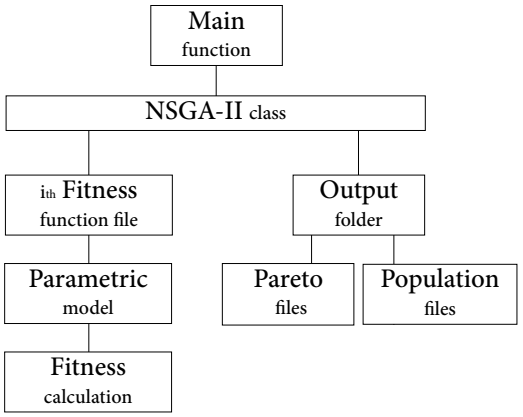


Figure 8.4: Diagram of the Data structure for the NSGA-II implementation.

functions, they are called by it, and they receive the decoded and scaled individuals to be studied. This fitness functions generally call many other functions, to calculate structural, acoustical and/or energy phenomena and attribute a fitness value accordingly. This functions will be studied in more depth in the chapters dedicated to each discipline. But perhaps more importantly, contained in the fitness function, is the parametric model. The individual solution variables in this research almost always represent geometric data to be fed to a parametric model. This model is often content specific, meaning that the model generated for structural analysis is quite different from the one used by energy simulations. They differ in their level of detail, type of computational geometry, their representation and commonly their discretization.

Fitness values are returned to the Multioptimize function in oder for it to pass it to the necessary genetic operator functions. At each generation of the MOGA, two output files are written and saved in their respective folders as shown in figure 8.4. One file contains all of the data necessary to describe the individuals in the combined population from the search space point of view. This is the population file. The other file expectedly contains the information to describe the objective space at each particular generation. It is called the Pareto file.

The particular implementation of NSGA-II used in this research uses an end condition that considers only the number of generations. When the

MOGA has completed a user defined number of generations the algorithm stops and generates another file containing the final results. This file is stored in the output folder.

8.5 Mathematical Benchmarks for NSGA-II

A series of simple mathematical functions were used to test the capabilities of the NSGA-II python implementation to find Pareto front individuals. By observing the resulting Pareto fronts and comparing them to those found in literature, we can asses the correctness of the implementation of NSGA-II.

8.5.1 Benchmark A

The first benchmark problem we presented to NSGA-II is found in "Multi-Objective Optimization using Evolutionary Algorithms"(Deb 2001) page 176. In this book the problem is called $MinEx$, it is used throughout the book to compare the performance of many different algorithms. $MinEx$ is described in equation 8.2:

$$MinEx \ : \begin{cases} Minimize & f_1(x) = x_1, \\ Minimize & f_2(x) = \frac{1+x_2}{x_1}, \\ subject\ to & 0.1 \leq x_1 \leq 1, \\ & 0 \leq x_2 \leq 5. \end{cases} \tag{8.2}$$

The implementation of NSGA-II was used to search for solutions for $MinEx$. The following GA input was given to NSGA-II:

Population Size (N)	100	
Number of Variables	2	
Number of binary digits	8 for x_1	8 for x_2
Variable Domains	$x_1 \in [0.1, 1]$	$x_2 \in [0, 5]$
Mutation Probability (p_m)	0.1	
End Condition	Number of Generations	50

Figure 8.5 shows the parameter and objective spaces for Benchmark A, $MinEx$, at the end of its run (after 50 generations).

We can see that the algorithm found a very good distribution of solutions along the Pareto front. The results compares fairly well with the results found by Deb in his book. We can interpret this as a sign that the Python

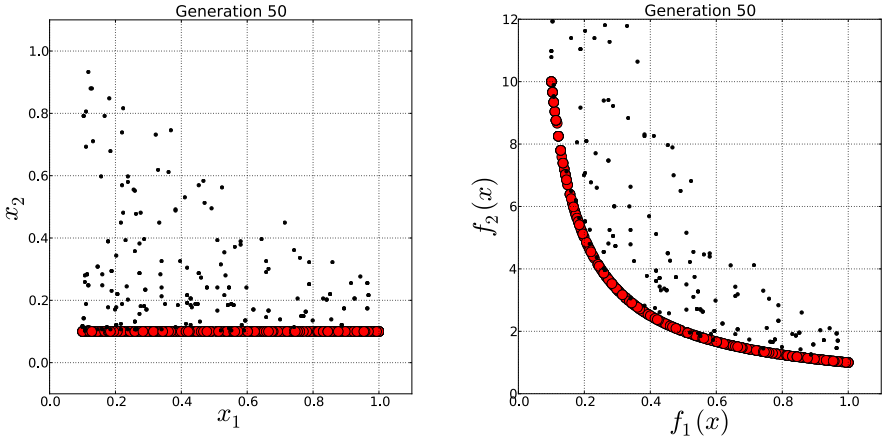


Figure 8.5: Parameter and Objective Spaces for Benchmark A.

implementation of NSGA-II done for this research was successful. Figure 8.6 shows the search and objective spaces found at generation 1, 5 and 10. We can see the different Pareto fronts found by NSGA-II in these generations and note how the evolutionary process improves the result at each iteration.

Generation 1 has a non-dominated set that is not yet close to the real Pareto front that we see in Generation 50. However, already at this early stage, we can see that x_2 values are all between 0 and 1, no solutions with x_2 between 1 and 5 are being considered because they give sub-optimal values. This is a very quick reduction of the search space. By generation 5 almost all $x_2 = 0.1$ (the correct value) but not perfectly. As a result, the Pareto curve is not yet perfectly drawn. We can also see that at generation 5 the distribution of individuals in the Pareto front is not very regular. When we reach generation 10 all $x_2 = 0.1$, and we have a fair distribution, but this distribution is further improved generation by generation. By the time we reach generation 50 the distribution has improved considerably.

8.5.2 Benchmark B

The second Benchmark we set for NSGA-II is described in (Zitzler et al. 2000). Zitzler et al. wrote a series of two function problems to test and compare the efficiency of multi-objective search algorithms. These problems

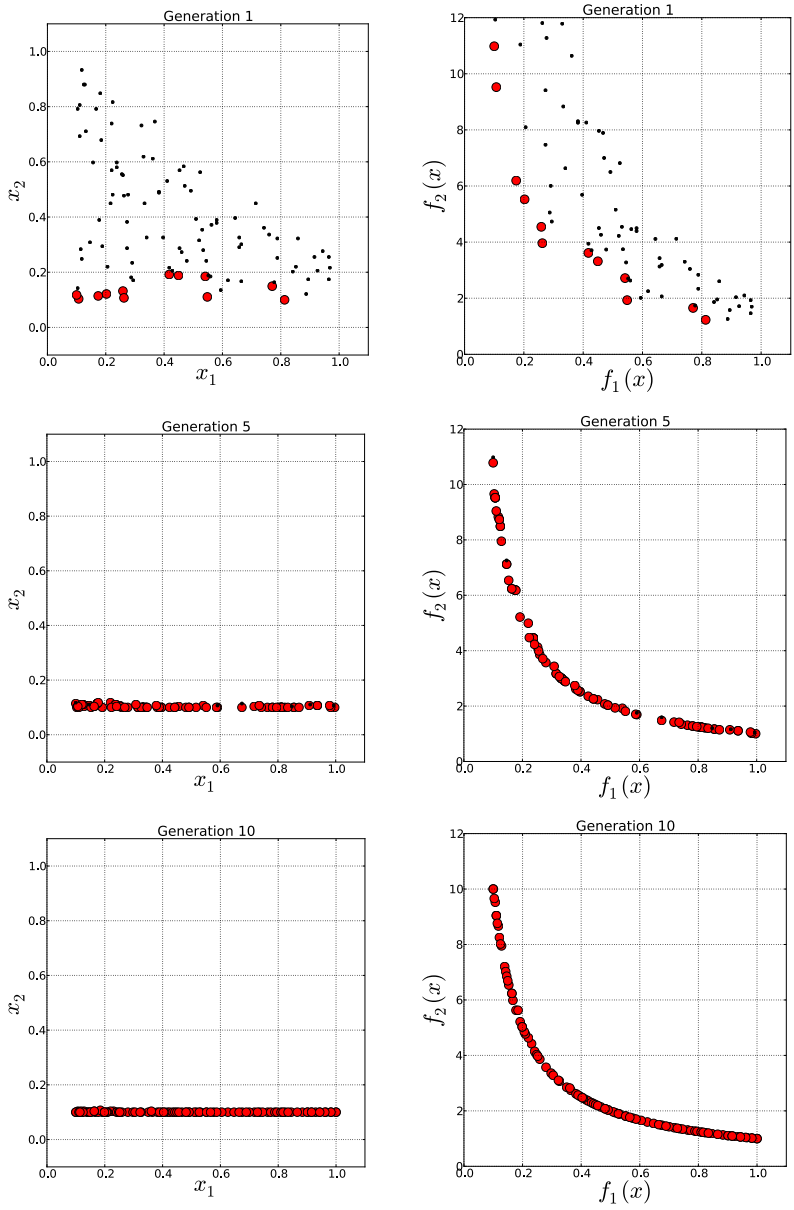


Figure 8.6: Search and Objective Spaces for Benchmark A at generation 1(top), generation 5 (middle) and generation 10 (bottom).

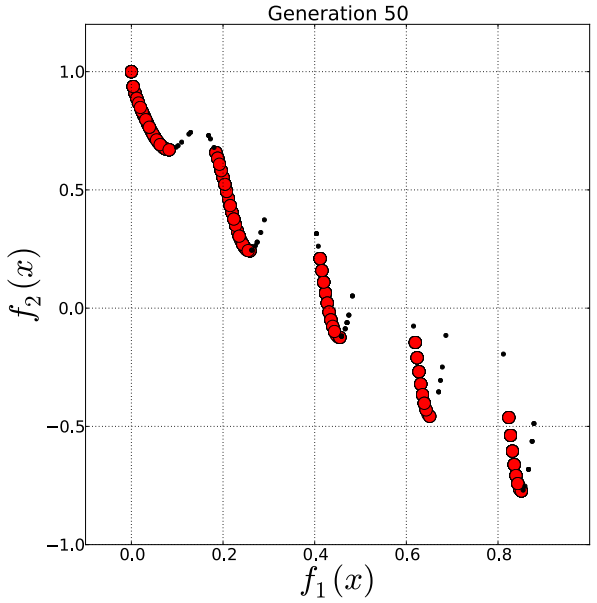


Figure 8.7: Objective Space for Benchmark B.

are called ZDT1 all the way to ZDT6. They all posses a high number of variables n and they vary in complexity. In particular, for this research we will be using ZDT3 as Benchmark B. The most important characteristic of ZDT3 is that it results in a very discontinuous Pareto front, composed of 5 different curves. Discontinuous Pareto fronts are of particular difficulty for search algorithms as they require global and local search methods to work simultaneously. Also the high number of variables (in the case of ZDT3 $n = 30$) means that the problem complexity is high. ZDT3 is described in the following equation:

$$ZDT3 : \begin{cases} Minimize & f_1(x) = x_1, \\ Minimize & f_2(x) = g(x) \cdot h(f_1(x), g(x)), \\ where & g(x) = 1 + \frac{9}{n-1} \cdot \sum_{i=2}^{n} x_i, \\ & h(f_1, g) = 1 - \sqrt{f_1/g} - (f_1/g) \cdot sin(10\pi f_1), \\ subject\ to & 0 \leq x_i \leq 1 \\ & n = 30. \end{cases} \tag{8.3}$$

Genetic variables for Benhckmark B were as follows:

Population Size (N)	100
Number of Variables	30
Number of binary digits	8 for x_i
Variable Domains	$x_i \in [0, 1]$
Mutation Probability (p_m)	0.1
End Condition	Number of Generations 50

Figure 8.7 shows the Objective space for Benchmark B. In it we can see that out implementation of NSGA-II was capable of finding an even distribution of individuals in all 5 curves of the Pareto front. This result also compares very well with the results found by Zitzler et al.

9

Parametric Models

In part I of this thesis we discussed parametric models from a architectural point of view, in this chapter we will go trough them in a more mathematical approach, starting from a mathematical definition. Daniel Davis provides such a definition in his PhD dissertation:

> "Returning to the Concise Encyclopedia of Mathematics, a parametric equation is defined as a "set of equations that express a set of quantities as explicit functions of a number of independent variables known as parameters". The mathematical definition can be refined by recognizing that the "set of quantities" in the context of design representation is typically geometry (although not always). Thus, a parametric model can be defined as: *a set of equations that express a geometric model as explicit functions of a number of parameters*."

(Davis 2013)

In our previous example problems we have not dealt with geometry, we have been using simple equations. The MinEx problem we saw on page 8.2 had two simple functions $f_1(x) = x_1$ and $f_2(x) = \frac{1+x_2}{x_1}$, in which x_1 and x_2 are the parameters. This set of equations conforms well to the first part of the definition we just saw, the definition for a parametric equation.

A parametric model would use its parameters to generate a geometrical object. We can make an example of a parametric model of a surface with two parameters. In this example the surface will be parametrized by means of a couple of Parabolas. These Parabolas will be determined by their heights

h_1 and h_2, which will be our parameters. Using These two parameters we will be able to generate the complete surface, calculate the position of any point in the surface, for any combination of h_1 and h_2 values.

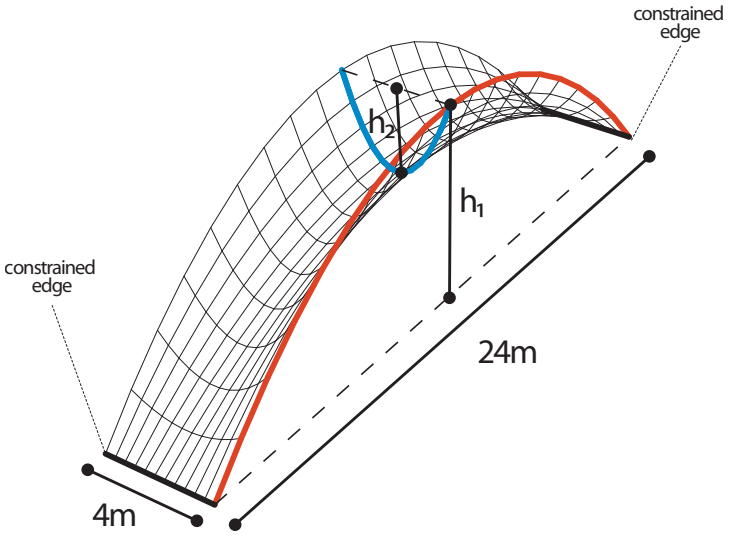


Figure 9.1: Example Parametric Model - Parametric Surface

Figure 9.1 shows a diagram of the parametric model, the two parabolas and the use of h_1 and h_2. The surface has a fixed rectangular shape in plan, the rectangle is 24×4 meters. The short borders of this rectangle are constrained, they will not change position in the Z coordinate, no matter which values h_1 and h_2 take. All of the other points in the surface are subject to changes in their Z coordinate.

The first parabola is anchored at the short edges of the rectangle (hence it is 24 meters long in plan), and it's height is determined by h_1. This parabola is the longitudinal one shown in red. We can see that the surface is symmetric, and that the first parabola is repeated at both ends of the surface. The second parabola (the transversal one shown in blue) is anchored at the midpoint of the first parabola and its height is determined by h_2.

The parameters on their own will not generate the whole geometry, they are merely the variables in the set of equations that will. In the parametric modeling techniques employed in architectural research and practice, and also the ones used in this thesis, these set of equations are computationally solved with the help of powerful CAD software that can generate even complex geometry with simple commands. Our current example is simple enough for us to use the equations of the parabolas to determine the geometry, this will allow us to see how a simple parametric model is computed. The parabola is usually expressed as a function of x and y cartesian coordinates:

$$\begin{aligned} y = f_{(x)} &= ax^2 + bx + c \\ &or \\ x = f_{(y)} &= ay^2 + by + c \end{aligned} \tag{9.1}$$

where a, b and c are coefficients that determine the shape and position of the parabola. But a more parametric equation, and an equation that is more directly related to our problem constraints is:

$$y = f_{(x)} = \frac{(bx - x^2) \cdot 4h}{b^2} \tag{9.2}$$

where h is the height of the parabola and b is it's base. In this case, given h and b we can get y values for all points along the x axis. Since our parametric surface is a tridimensional object, we will need to take this into account in our set of equations. They will be set up in such a way as to obtain the Z coordinate of a point determined by its X and Y coordinates, and of course our parameters h_1 and h_2. Since in our example, the plan of the surface is fixed to a rectangle measuring 24×4 meters, we can say that $0 \leq X \leq 24$ and $0 \leq Y \leq 4$. We can also say that b in the parametric parabola equation (equation 9.2) will be $b_1 = 24$ for the longitudinal parabola, and $b_2 = 4$ for the transversal one. The following set of equations defines our parametric surface:

$$\begin{aligned} Z &= \tfrac{(b_1 X - X^2) \cdot 4 \cdot (h_1 + H)}{b_1^2} \\ where \qquad H &= \tfrac{(b_2 Y - Y^2) \cdot 4 h_2}{b_2^2} \end{aligned} \tag{9.3}$$

where H is the height of a longitudinal parabola that has its midpoint at the transversal parabola at Y. In this sense we can see the set of equations as giving us a series of parabolas at X that have a height determined by h_1, h_2 and the transversal parabola.

So if for example we wanted to draw one surface using our parametric model, we would have to select value for our parameters h_1 and h_2, and then

could compute all values of Z for a grid of X and Y points, thus creating point cloud of our surface. If we then connect these points with lines we would obtain a 3D mesh representation of our surface like the one shown in figure 9.1.

9.1 Parametric models and Search Space

The most important thing to consider when creating a parametric model for the purposes of computational search is the possible outcomes present in the model, the kind of geometry that we will be including in the model (and thus the search) and the kind of geometry that would not be included. When we create the parametric model we are effectively defining the *search space* of our search problem, defining which set of geometry we will study. Designers generate parametric models in many different ways, all having advantages and disadvantages, but all parametric models have their limits, they can generate a wide range of geometry, but not all geometry. The limits of the model are defined by the parametric equations and the domains we give to our parameters.

In order to better understand this we will use the parametric surface example described above and see what kind of geometry we can get out of it, and what kind we cannot.

Figure 9.2 shows a series of surfaces that were all generated with our parametric model. 9.2(a) shows the resulting surface when both h_1 and $h_2 = 0$, we get a flat surface. If we fix $h_2 = 0$ but we assign a $h_1 \neq 0$ we obtain a single curvature surface as shown in 9.2(b). Double curvature surfaces are also possible, if we set both h_1 and $h_2 > 0$ or both h_1 and $h_2 < 0$ we get sinclastic surfaces or positive double curvature surfaces as shown in 9.2(c). On the other hand if we set $h_1 > 0$ and $h_2 < 0$ or $h_1 < 0$ and $h_2 > 0$ we obtain anticlastic surfaces or negative double curvature surfaces, as shown in 9.2(d).

Limiting the geometric possibilities that a model has is a good choice when using computational search methods. This allows the designer to only consider a very defined set of geometric possibilities at a time, if he so wishes. Not only is it possible to limit the geometry within dimensional values (for example limiting the surface from reaching heights above 10 meters), but also is possible to limit the geometry in a more qualitative way. Using boundary domains in the models parameters we can easily limit the possibilities of the model, we can prevent double curvatures, concave surfaces from the top, convex surfaces, flat surfaces, etc. The following table shows a few possible

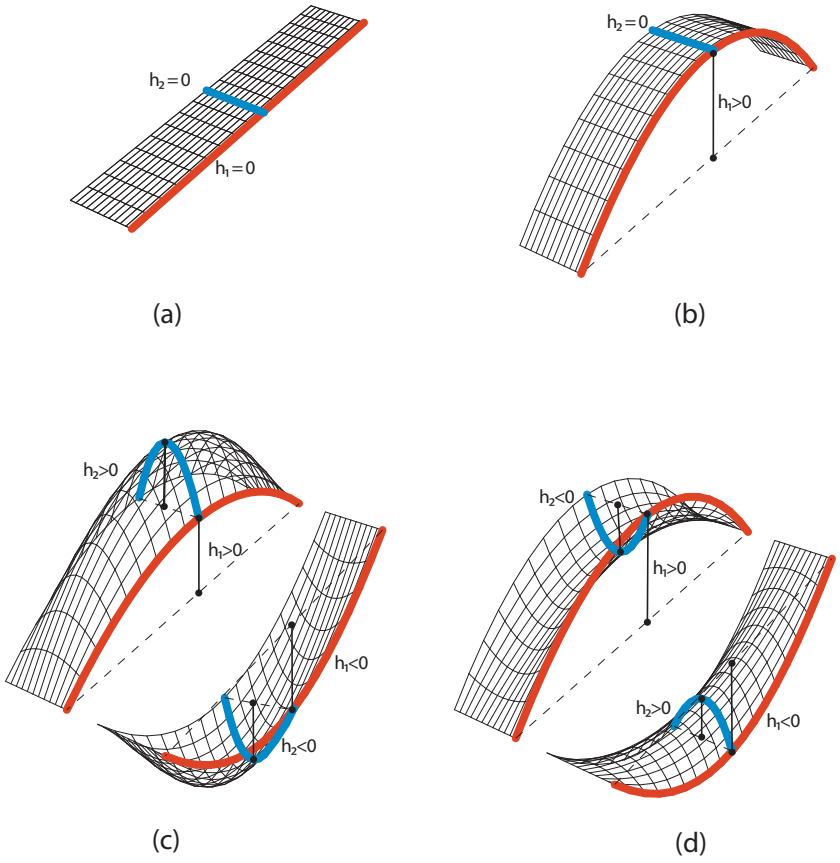


Figure 9.2: Parametric Surface possible Outcomes

domain combinations of h_1 and h_2 and the included and excluded geometry:

h_1 domain	h_2 domain	Included	Excluded
$-\infty < h_1 < \infty$	$h_2 = 0$	single curvature concave convex	sinclastic anticlastic
$h_1 > 0$	$h_2 > 0$	sinclastic convex	single curvature anticlastic concave
$h_1 < 0$	$h_2 < 0$	sinclastic concave	single curvature anticlastic convex
$h_1 > 0$	$h_2 < 0$	anticlastic convex	single curvature sinclastic concave

We can se that the parameter domains are a powerful way of controlling the search space. But of course the vast majority of the geometrical options are defined with the parametric equations themselves. No matter what limits or values we give to the parameters, in our example we will not obtain folded surfaces, multiple vertices or periodical surfaces. If we wanted to include one or all of those possibilities in the search space we would need to define a new model.

.

Part III

Applications of Computational Search

10

Shell Structures

10.1 Concrete Parabola-based Bridge Benchmark

The first use of Genetic Search in this PhD thesis parametric a single-objective GA used for a concrete shell bridge*. This first search process parametric developed as a Benchmark for the Genetic Algorithm, to test its ability to find optimal and sub-optimal individual solutions an a simple structural problem. In order to track the evolution and results of the GA a "fitness landscape" of the same structural problem is plotted and the GA evolution mapped on top of it.

10.1.1 Parametric Model

The Parametric model for this example is the exact same model shown in chapter 9 described by equation 9.3 and shown in figure 9.1.

As it was previously explained in chapter 9, this parametric definition of the surface guarantees that, by varying the values of h_1 and h_2, different configurations can be obtained: a completely flat surface 9.2a; positive double curvature surfaces 9.2c; negative double curvature surfaces 9.2d; and single curvature surfaces 9.2b.

*This example was developed by the Author and Mario Sassone and is published in a Chapter of the Book "Shells for Architecture: Form finding and structural optimization" edited by Sigrid Adriaenssens, Philippe Block, Diederik Veenendaal and Chris Williams (Méndez Echenagucia et al. 2014- IN PRINT). The Chapter entitled Computational Morphogenesis was written by the Author, Mario Sassone and Alberto Pugnale.

Since these two parameters effectively control the overall shape of the bridge, h_1 and h_2 are chosen GA the parameters. For both we have established a domain spanning from -40 m to +40 m. This means that the search space of this problem is two-dimensional and can be represented as a grid of values from 40 to 40 in its X and Y axis. We can thus assume that a vector of genetic variables x is made up of two variables: $x_1 = h_1$ and $x_2 = h_2$.

The parametric definition proposed above is then implemented in the geometry modeler. When the GA calls for a shape in terms of a set of x_1 and x_2, the CAD modeler generates the corresponding surface following equation 9.3, providing the object to be evaluated. However, for the FEM analysis a discrete model has to be generated. The geometry is discretized into a mesh, composed of shell or beam elements, depending on the type of structure. For this exercise, the shell is simplified as a grid-shell of comparable mechanical properties.

10.1.2 Structural Fitness Function

The construction of an FEM model for structural optimization, presents some differences with the ones used in normal analysis. Complex models require time consuming calculations, which represent bottlenecks in the flow of operations. The first requirement of FEM models for optimization is to be simple, with a number of elements strictly necessary and with a mesh correctly defined to evaluate the pertinent aspect of structural behavior. Even with powerful hardware setup, the repetition of hundreds, or even thousands of analyses might transform the optimization process into an extremely long task, if the model is not efficient.

There are basically two possibilities: (i) the use of customized finite elements solvers, developed in the same environment and (ii) the use of external applications, like commercial software. Both the alternatives have advantages and disadvantages, for this benchmark we will employ the first alternative. In shell analysis an important issue is the choice of the elements to use: in fact, even for a simple non-layered elastic shell, different formulations and approaches can be adopted. In the proposed application, the shell is approximated by a mesh of one dimensional beam elements, which geometric properties are defined in order to reproduce the characteristics of a continuous shell. This allowed to use a custom Python FEM code developed by Mario Sassone and capable of interacting seamlessly with the parametric modeler.

Displacements, strains, stresses, and strain energy are basically the effect of a load condition on an elastic structure. A stiff structure will show small

displacements and strains, while a strong structure will result in relatively small stresses, and both will have small strain energy. Displacements are a vector field, stresses are a tensor field, locally defined, and the strain energy is a scalar value, computed as an integral over the whole structure. Such quantities can be adopted as a measure of the structural performance. However, their differences will drive the optimization process to search for different optimal solutions. For this exercise, the maximum displacement of the whole structure is chosen as the fitness function to be minimized by the GA. As opposed to the strain energy, nodal displacements can reveal local, as well as global, weaknesses. Our fitness for this benchmark will be:

$$Fit_{structure} = max(\Delta_{Z_i}) \tag{10.1}$$

where Δ_{Z_i} is the deformation in the Z axis of the i_{th} node.

10.1.3 Fitness Landscape

The structural problem proposed for this benchmark is defined by only two parameters. This is purposely done in order to track and further explain the work done by the GA using a graphical representation. We first map out the solution domain of x_1 and x_1 by taking a two-dimensional parameter grid with grid points $P(x_1, x_2)$. By assigning a z value to each point of the grid, we convert it into a three-dimensional surface in which the z value represents the fitness calculated for the shape (individual) corresponding to that grid point P. For example:

- $P_{(0,0)}$ represents the completely flat surface, it has a maximum displacement in the z-axis (our fitness value) of 285 mm;
- $P_{(26,1)}$ has a maximum displacement in z of 21 mm;
- $P_{(40,40)}$ has a maximum displacement in z of 43 mm.

By repeating this operation for a series of individuals obtained by the discretization of the search space, we end up with a complete surface. This kind of three-dimensional representation of our problem and solutions is called a fitness landscape. Figure 10.1 shows the fitness landscape obtained in our Benchmark, which has many local minima, two global minima and an area of global maxima. This means that the problem can be considered a multimodal problem. Multimodal problems are notoriously difficult for gradient based or non-stochastic search methods, but they should be well within the possibilities of GA.

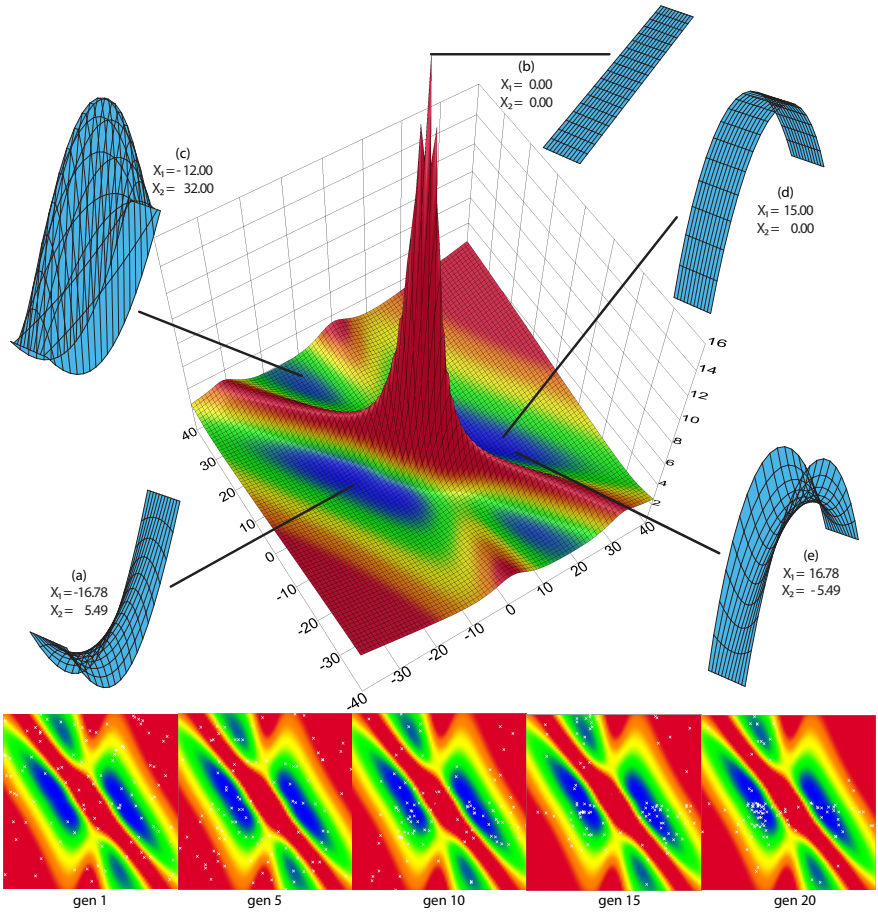


Figure 10.1: Fitness Lanscape for Parabola-based double curvature Benchmark, significant individuals and Genetic Algorithm evolution.

10.1.4 Genetic algorithm inputs

The GA used in this benchmark employes a binary coding of design parameters with mutation and elitism operators. It terminates after 20 generations with a population size of 100 individuals. The two variables (x_1 and x_2) are here coded into eight digit binary numbers or genes. Such genes are combined into one chromosome with the x_1 value positioned first and the x_2 second as explained in chapter 6. The single point crossover operator is used. We can resume the GA inputs of this benchmark in the following table:

Population Size (N)	100	
Number of Variables	2	
Number of binary digits	8 for x_1	8 for x_2
Variable Domains	$x_1 \in [-40, 40]$	$x_2 \in [-40, 40]$
Mutation Probability (p_m)	0.1	
End Condition	Number of Generations	20

10.1.5 Results

In Figure 10.1, we see a three-dimensional representation of the fitness landscape for this benchmark. Such a graphical tool allows us to study the exploration performed by the GA within the solution domain and to evaluate its efficiency by mapping generation by generation the search progress

The plan views of the fitness landscape at the bottom of figure 10.1 show the evolution of the GA at generations 1, 5, 10, 15 and 20. The individuals being considered by the GA are represented by small white exes on top of the plan views of the fitness landscape. We can see how the GA gradually concentrates its individuals in areas of interests, most particularly global and local minima seen in blue.

The figure also shows different configurations of the bridge structure in relevant points of the fitness landscape. It is of particular interest to see how the shapes of the local minima differ from one another, even if they possess similar fitness values. The search process aims to find global minima, but we have seen with this exercise that even other sub-optimal candidate solutions might be worth considering:

- Individual (a) is the global minimum as found by the GA after 20 generations. It is a hypar with a maximum displacement of 169mm. Because of the symmetrical nature of the problem, we can say that

individual (e) is the symmetric opposite to individual (a) and becomes a global minimum as well.

- Individual (b) represents the global maximum, a flat surface with a maximum displacement of 285mm.
- Individual(c) is a local minimum with very tall parabolas forming an highly irregular hypar.
- Individual (d) is a near optimal single curvature configuration.

This benchmark showed that the single objective GA is able to quickly point out global and local minima in a simple structural problem. Already in generation 5, the GA had found the global minima, and by generation 20 it had explored a much wider area, point out other areas of interest in the fitness landscape, thus demonstrating that the GA is an appropriate tool for multimodal problems.

The fitness landscape is proven to be an interesting and effective tool in the study of performance related problems. Its main drawback is the fact that it is limited to a two-variable problem, otherwise there would not be enough dimensions to properly represent the problem and its results. Another important drawback of this method is the calculation times. Since, in order to have a detailed landscape, a big number of simulations have to be carried out, it can be considered a "Brute Force" or exhaustive search method, and therefore its efficiency is not very high. However, for simple problems with small domains it proved to be an interesting tool for the study of parametric models combined with performance simulations.

10.2 Case Study 1: Concrete free-form Roof

After testing the GAs ability to explore a wide search space, and a multimodal objective space to find local and global minima, it was time to take one step further, and develop a multi-objective structural problem. A multi-objective search problem with contrasting objectives for structural design is developed and solved using NSGA-II[†]. The problem consists in the design of a 24×24 meter roof supported at its corners with a fixed and continuous thickness, with the objective of making it as rigid and as light as possible.

[†]This Case study was published as a part of an article on the International conference on Structures and Architecture 2013 in Guimarães, Portugal (Méndez Echenagucia, Pugnale & Sassone 2013).

So when compared to the previous benchmark, this case study has the important difference of being a multi-objective problem, and additionally, this case study does not employ the same parabola based parametric study. It employs more complex geometric possibilities involving NURBS geometry.

10.2.1 Parametric Model

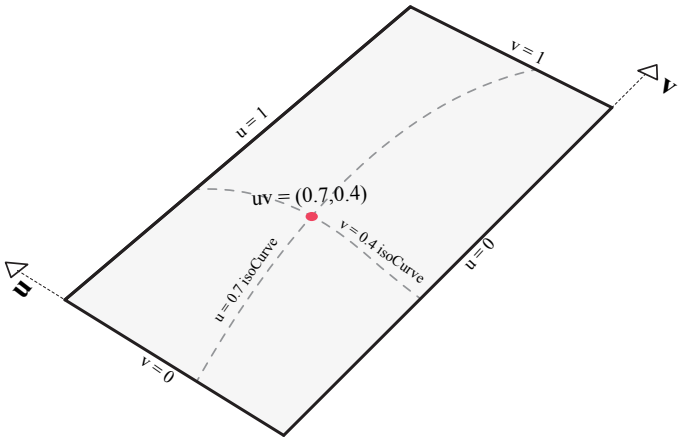


Figure 10.2: Non-Uniform Rational B-Splines - U and V parametrized surface.

In this case study the possibilities desired for the solution space went beyond what was possible with the previously studied parabola-based surfaces. Apart from single and double curvature surfaces, this case study parametric to include multiple curvature surfaces, often referred to as free-form geometry.

Non Uniform Rational B-Splines (NURBS) were introduced in the late 1970's mainly from the work of Pierre Bézier and Paul de Casteljau both working in the french automotive industry. They are the current standard for describing curves and surfaces in computer aided design (Rogers 2001). The representation of NURBS surfaces requires the use of two parameters commonly U and V, and so the X, Y, Z coordinates of any given point can be thought of as functions of U and V, e.g. $x = x_{(U,V)}; y = y_{(U,V)}; z = z_{(U,V)}$. So we can say that any point in this surface can be described by

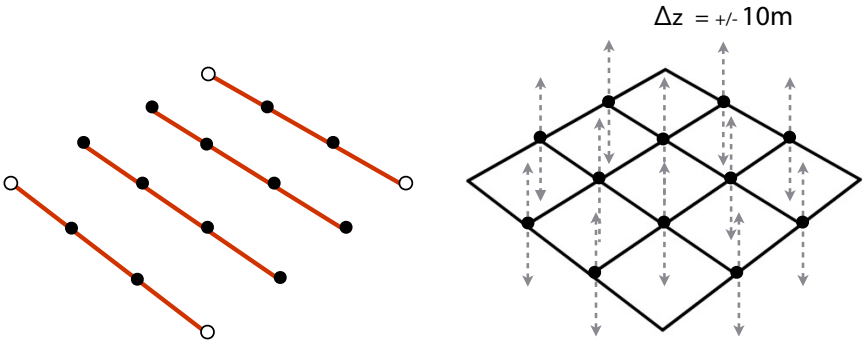


Figure 10.3: Parametric Model for the 24×24 roof problem.

its bi-parametric coordinates U and V (figure 10.2). This is a very similar approach when compared to the one described for the parabola-based surface in equation 9.3. The most important difference being that the NURBS surfaces uses not simple parabolas, but B-splines, capable of describing just about any curve we could think of, with just a few control points.

The NURBS geometry used throughout this PhD thesis is generated with the aid of the Rhinoceros, a commercial CAD software that contains a large library of functions that generate, edit and analyze NURBS. Rhinoceros also offers its users the possibility to customize the use of its functions by writing scripts that can access functionality in various ways and with multiple programing languages. One of these languages is Python, the same language we have been employing to develop our search algorithms and FEM solvers. This means that we can seamlessly call geometric functions as easily as we can call search or FEM functions.

Figure 10.3 shows the parametric model for this case study. The surface for this case study is generated by means of four NURBS curves (shown in red). These curves in turn define the NURBS surface that is the subject of study. The NURBS is built starting from a set of four spatial curves. The two curves at the ends have its end points fixed (seen represented as white filled dots in figure 10.3, while the other two are NURBS curves laying in vertical planes. As it will be shown bellow, the fixed points correspond to the structure supports. The four curves act as four vertical sections of the surface to be generated. The surface is defined as a NURBS passing through the section curves (this is called a lofted surface), with assigned polynomial

degree. Each curve is defined by four interpolating points, whose vertical position, the Z coordinate, is variable. Each variable point can move within the confines of a vertical line 20 meters long, they can move 10 meters above or 10 meters bellow the starting point $Z = 0$. By modifying the coordinate of the interpolating point, the section curves change and so does the surface. In such a way, a set of 12 real numbers is used to completely define the surface shape. The other NURBS parameters of the surface, as the degree of interpolating functions or the number and position of control points, are set constants in the problem.

We have established that there are a total of 16 control point in the surface, but 4 of them are not variable. Also we established that only the Z coordinate of these points is variable, so we can say that the number of variables in this problems is equal to 12. This gives us a vector of variables X where:

$$X = z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, x_{10}, z_{11}, z_{12} \tag{10.2}$$

We can also se this vector in a more general way:

$$X = x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12} \tag{10.3}$$

where in this case $x_{i_{th}} = z_{i_{th}}$.

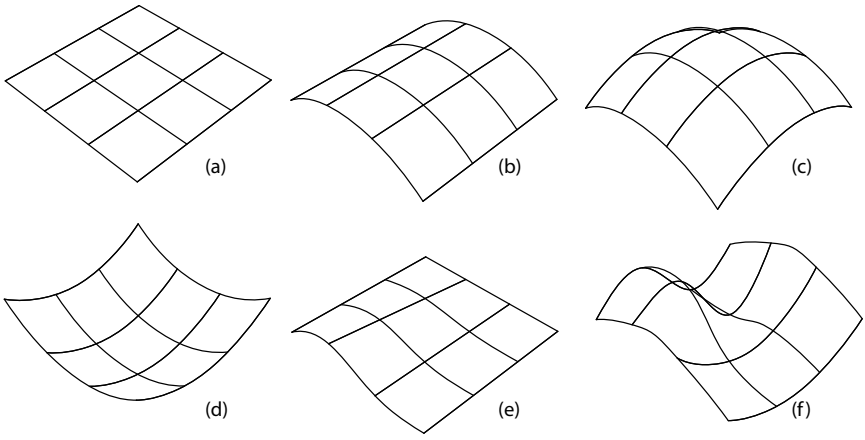


Figure 10.4: Possible individuals with the parametric model developed for the roof Problem.

The ranges of variability of each parameter define the set of potential solutions. As it has been already said, the length of the variable domain influences the search process and the possible outcomes. Figure 10.4 shows six of the possible surfaces contained within the above defined parametric model. Surface (a) is a flat surface, (b) is a single curvature surface, while (c) and (d) are positive double curvature surfaces in different directions, and (e) and(f) are surfaces with multiple curvatures, meaning they have both positive and negative curvatures within the same surface. Surface (e) has a complex curvature that is not very accentuated, while (f) has more pronounced curves.

10.2.2 Fitness functions

One of the important premisses for this Case Study is the use of multiple and contrasting objective functions. In particular for this case we want to search for rigid but light concrete structures. For every generated solution, two functions are used to calculate its fitness.

The first function is the same one used in the above benchmark, the maximum displacement in the z-axis. The FEM solver used in this example is also the same one used for the Benchmark, it is the FEM solver developed by Mario Sassone in the Python programing language. This enables a smooth communication between the parametric model, the FEM solver and the Search Algorithm. However, in this case, the geometry is not generated with the parabola function, but with the NURBS parametric model mentioned above, which leaves us with the problem of discretization of the otherwise continuous NURBS surface.

The FEM solvers used in this research are not capable of working with NURBS geometry. Geometry must be represented in small and flat shell elements. This means that we need to split up our continuous NURBS surface into small flat elements. Depending on the type of geometry, the only way to make sure that the resulting elements are flat is to generate triangular elements, not all of the resulting rectangular elements might be flat. Since we are working with free-form geometry, we discretized the NURBS surface into triangular elements using a set of Rhinoceros functions that were written for this exact purpose. The proper discretization of continuous surfaces can be a relatively simple task using this set of Rhinoceros functions, but as we will see later on this is not the case for surfaces containing windows or gaps.

The resulting discretized Mesh surface is loaded in its nodes with only a vertical load as shown in figure 10.5. The surface is only supported at its corners, that in this case are 24 meters apart in both directions. The support

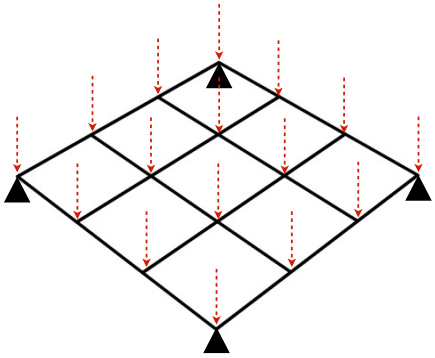


Figure 10.5: Loading and Node Constraint conditions for the 24×24 roof problem.

points are assumed to have pin joints, so they will not offer resistance to rotation.

The second fitness function measures the surface's weight. This function could be calculated by taking the surface's area and multiplying it by a fixed thickness (thus obtaining volume) and consequently multiplying this volume by the proper cubic weight of its material, concrete. However, since both material weight and thickness is the same for all surfaces, this function is simplified and the only value taken into consideration is the surface's area, the only variable in the calculation explained above.

We can sum up this multi-objective problem in the following way:

$$Case\ Study\ 1 \begin{cases} Minimize & f_1(x) = max(\Delta_{Z_i}), \\ Minimize & f_2(x) = S, \\ subject\ to & -10 \leq x_i \leq 10. \end{cases} \tag{10.4}$$

where x is the vector of variables, Δ_{Z_i} is the displacement in the Z-axis of the i_{th} node and S is the surface area.

10.2.3 Genetic algorithm inputs

This case study's genetic inputs were the following:

Case Study 1 Run#	1
Population Size (N)	10
Number of Variables	12
Number of binary digits	8
Variable Domains	$x_i \in [-10, 10]$
Mutation Probability (p_m)	0.1
End Condition	End after 500 generations

10.2.4 Results

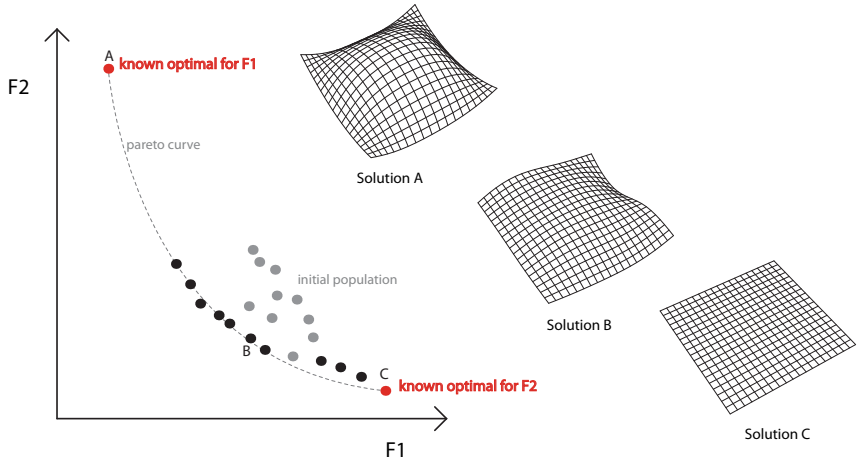


Figure 10.6: Objective space and Pareto front - 3 individual solutions for the roof structural multi-objective search problem.

Figure 10.6 shows the results of the search process for case study 1. The Pareto front, represented in the fitness space, contains the best solutions found by the algorithm: the lower branch of the curve contains solution that privilege lightness to stiffness, while in the left branch stiffer but heavier configurations can be found. In this benchmark, the best solution related to each fitness is known: the lightest shape is the flat shape, while the stiffer shape is a dome with the four central point at the top of the domain and the eight lateral points at the bottom. During the search process, the crowding distance algorithm tries to keep a good spacing between solution in

the front, but the two extremes (A and B) were not found. The knowledge of such extremes allows us to evaluate the efficiency of MOGA in terms of ratio between the size of the found front and the size of the actual front, including extremes.

The solutions spacing in the front is a good indicator of the variety of geometrical shapes and of the way different shapes answer to multi-objective requirements. A set of such shapes, related to the position on the Pareto front, is also shown in figure 10.6, together with the extreme cases. A direct representation of the position of shapes in the variable space is not suitable, due to the number of dimensions, but they can be compared one to another by the designer who is in charge to handle the produced material.

The fact that the extreme solutions (A and B) were not found could be due to an insufficient exploration during the search process (low number of individuals in the population or due to a small number of generations relative to the problem complexity) Since in this problem we do not know the exact shape of the true Pareto front, we cannot be 100% sure about the problem complexity, the true Pareto front could be almost discontinuous in the segments between the Front found and the A and B solutions. Meaning that there wold be no feasible solutions in those missing segments, and only one single solution in the extreme. This would make finding the extremes A and B highly difficult. What we can say for sure, is that the exploration in this case parametric insufficient for the extremes to be found, we cannot be sure about the problem complexity.

Solution (B) is an interesting compromise between rigidity and weight. It is a complex surface with multiple curvatures that assure its rigidity, but also it does not posses a high surface area, making it not very heavy, especially when compared with other more rigid solutions.

In order to better understand the relationship between problem complexity and number of calculations, the search algorithm is executed two more times with different GA inputs. The GA inputs for the second calculations were:

Case Study 1 Run#	2
Population Size (N)	30
Number of Variables	12
Number of binary digits	8
Variable Domains	$x_i \in [-10, 10]$
Mutation Probability (p_m)	0.1
End Condition	End after 160 generations

Meaning that there were close to 5.000 calculations (the same number

as the first run, but with a higher population size). And the third time the algorithm ran, the inputs were:

Case Study 1 Run#	3
Population Size (N)	100
Number of Variables	12
Number of binary digits	8
Variable Domains	$x_i \in [-10, 10]$
Mutation Probability (p_m)	0.1
End Condition	End after 100 generations

Meaning that there were 10.000 calculations, with a larger population size than the two previous search runs. Figures 10.7 and 10.8 show the Pareto Front and initial population for all three search runs. We can see that when compared to the first run, the second and the third have a slightly improved coverage of the Pareto front, since they are both closer to the two extremes (A and B). The most significant improvement is made in the direction of solution A, the improvement in the direction of A is not as noticeable. The Pareto front also seem to be slightly better, having solutions that are would dominate solutions in the first run. However, if we compare run 2 with run 3, the difference is hardly noticeable. We could even say that the second run has a better coverage of the Pareto Front.

Since there is a lot of randomness in NSGA-II, we cannot say for sure what the causes are for these differences in the search runs, but we can draw some observations. There is a small difference between the second and third runs in favor of the second, the third run having twice as many calculations as the second. This suggests that increasing the number of calculations is not a guarantee of improvement. Run 1 had higher number of generations than run 2, but an apparently insufficient number of individuals in the population. Run 3 had a larger population size than run 2, but a lower number of generations, and the data suggests that this run would have benefited form a few more generations.

There seems to be a significant relationship between population size and number of generations. Too many individuals in the populations seems to increase calculation times needlessly, while a small number of generations is also not appropriate. We can assume that these values are highly problem dependent, and that achieving an optimum number of generations and population size is not easy. This might also allude to the balance between exploration and exploitation described above. Higher individuals in population denote higher exploration, while higher number of generations are sings of higher exploitation.

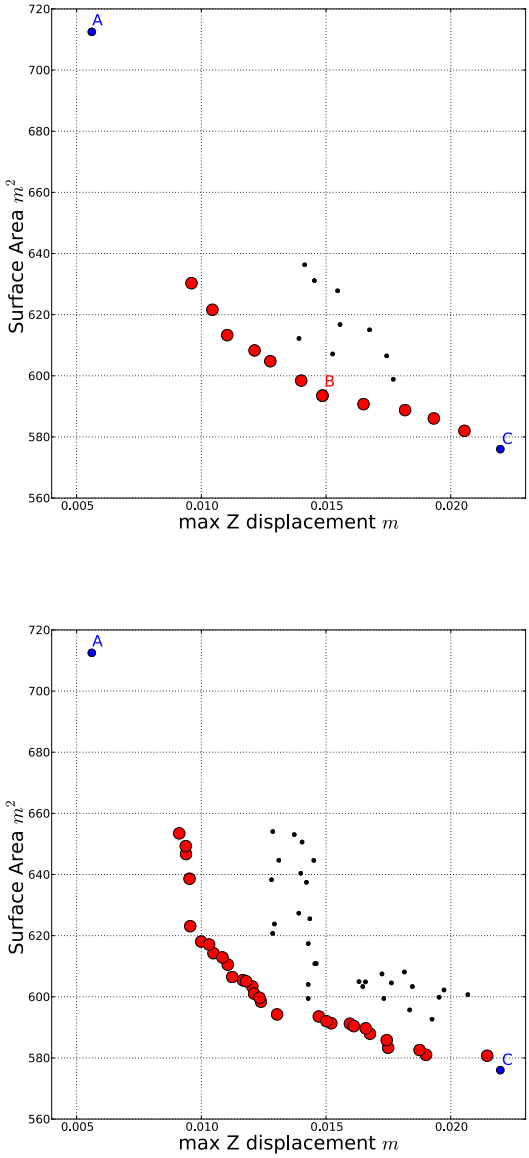


Figure 10.7: Objective spaces with Pareto fronts and initial population for Case study runs 1,2 and 3.

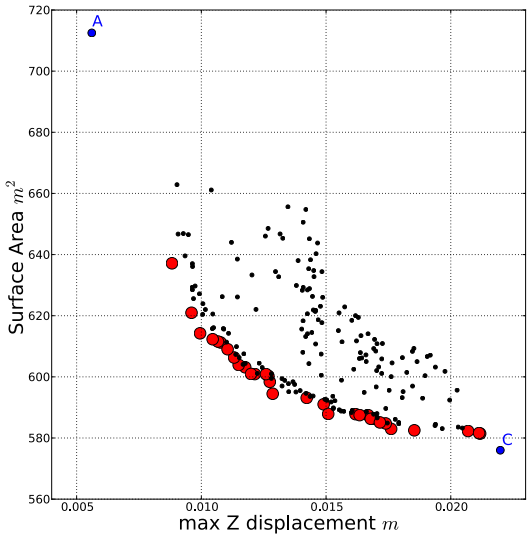


Figure 10.8: Objective spaces with Pareto fronts and initial population for Case study runs 1,2 and 3.

10.3 Case Study 2: Concrete free-form Bridge

The second case study shares many characteristics with the first one. It is also a multi-objective search problem with contrasting objectives for structural design and it is also studied using NSGA-II[‡]. The problem consists in the design of a 24×4 meter bridge supported at its ends with a fixed and continuous thickness, with the objective of making it as rigid and as light as possible. It follows the exact same premise as the benchmark but it uses free-form geometry.

10.3.1 Parametric Model

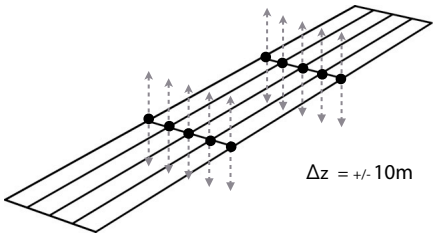


Figure 10.9: Case Study 2 Parametric Model - 24×4 Bridge.

Figure 10.9 shows the parametric model for Case Study 2. It is generated in the same way as the model for Case Study 1, its based on four curves, but in this case, the to curves at the end are fixed as flat lines. Therefore it has only eight variables, they are the z-axis coordinates of the eight control points that define the two middle curves. Also in this case the range of motion for the control points is 10 meters upwards and 10 meters downwards, making the total range of 20 meters. Also in this case, the possible outcomes include the flat surface, single and double curvature surfaces either positive or negative, and multiple curvature surfaces as well.

The Variable vector for Case Study 2 is as follows:

$$X = x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 \tag{10.5}$$

[‡]This Case study was published as a part of an article on the International conference on Structures and Architecture 2013 in Guimarães, Portugal (Méndez Echenagucia, Pugnale & Sassone 2013).

where $x_{i_{th}} = z_{i_{th}}$.

10.3.2 Fitness functions

The first fitness function in this problem is the maximum displacement in the z-axis $max(\Delta_z)$. Figure 10.10 shows the FEM model for Case Study 2. The surface is discretized into flat shell elements, and it is loaded at every node. It is modeled as a continuous thickness concrete shell, constrained with pin joints at its ends. The python FEM software developed by Mario Sassone is also employed in this case study for the calculation of the structure's displacement.

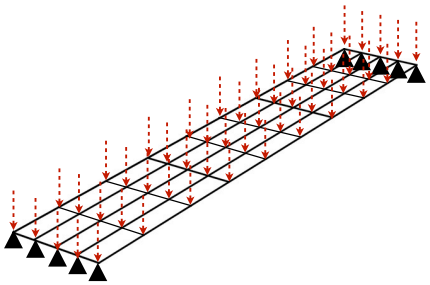


Figure 10.10: Loading and Node Constraint conditions for Case Study 2, the 24×4 Bridge.

The second fitness function is the same weight function, simplified as the surface area S. The problem can be expressed in the following way:

$$Case\ Study\ 2 \begin{cases} Minimize & f_1(x) = max(\Delta_{Z_i}), \\ Minimize & f_2(x) = S, \\ subject\ to & -10 \leq x_i \leq 10. \end{cases} \tag{10.6}$$

In this case there is only one known extreme, and that is the flat surface, having the minimum surface area from all possible solutions $S = 94m$.

10.3.3 Genetic algorithm inputs

For Case Study 2 the GA inputs were slightly different that for case study 1:

Case Study 2	
Population Size (N)	50
Number of Variables	8
Number of binary digits	8
Variable Domains	$x_i \in [-10, 10]$
Mutation Probability (p_m)	0.1
End Condition	End after 100 generations

The population size is fixed at 50 individuals and the number of generations set to 100. This means that we have 5.000 calculations. Eight binary digits means that we are dividing our range of motion (20 meters) into 256 steps, leaving us with a domain discretization of 7.8 centimeters. Since we have 8 variables and 256 possible values for each, that means that we have a total of 256^8 possible solutions, thats 1.84×10^{19}, a very big number.

10.3.4 Resutls

In figures 10.11 and 10.12 we can see the Pareto front for Case Study 2 in different levels of detail and at different stages of the search process. Even in this case the lightest shape is the flat shape (Individual A), but the stiffest does not have a theoretical significance. If the longitudinal section of the bridge were an arch with a shape perfectly corresponding to the pressure curve of the load, then the stiffest solution would have this shape and straight transverse section. It would be a barrel vault, or "flat arch". Since the parametric model we created does not allow such a perfect shape (due to the fixed distance of the generating NURBS curves), the only way to increase stiffness is to add some bending stiffness, through a transverse waved section. The shell, in this case, becomes a kind of ribbed arch, in which ribs increase the arch stiffness. Besides the extreme case of the stiffest shape, this considerations are important for other Pareto front shapes. Ribs, in fact, increase the stiffness and the weight at the same time. Figure 10.12 also shows a few significant solutions generated during the Search process for Case Study 2. The set of shapes depicted in figure 10.12 include some solutions coming from previous search steps (generation 22), instead than from the last only: those shapes do not necessarily represent local minima, but simply steps of the search process. However, they can play a role in interactive design, because they can be chosen as starting points of new search processes, through a redefinition of constraints and of the domain, suggested by the designer evaluations.

The 100_{th} generation contains the set of solutions that lie in the Pareto

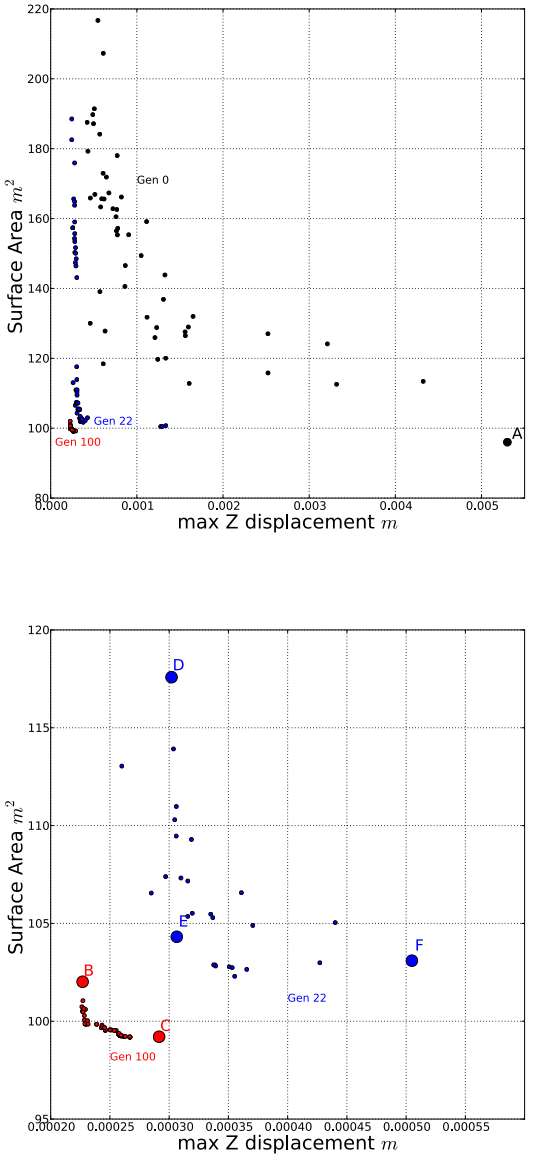


Figure 10.11: Objective space with Pareto fronts for the 100_{th} and 22_{nd} generations and the initial population for Case study 2.

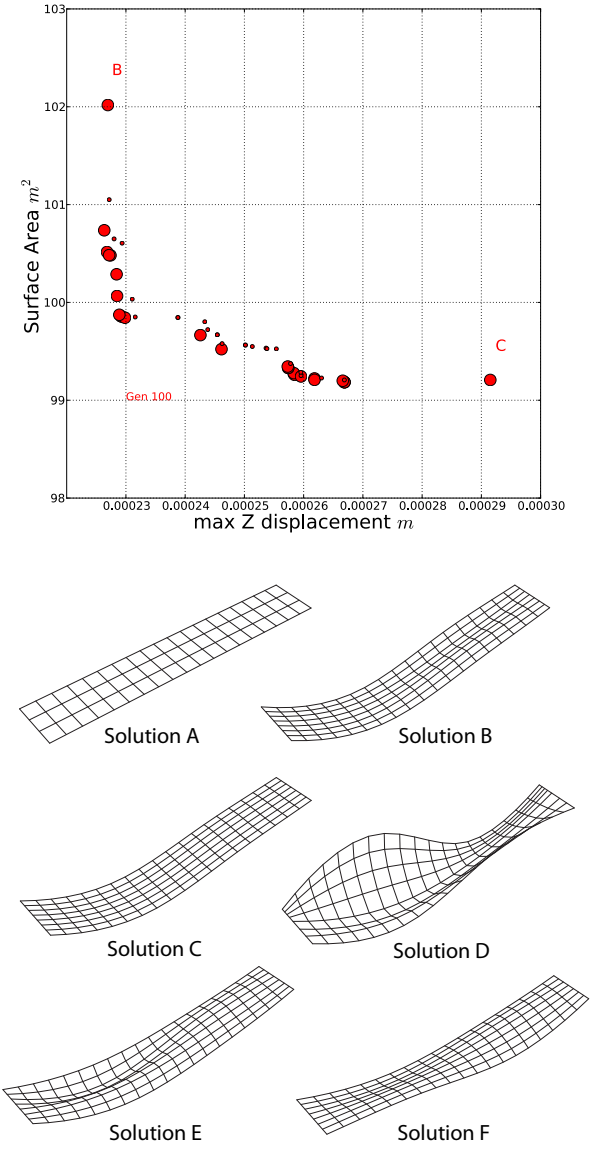


Figure 10.12: Objective space with Pareto fronts for the 100_{th} and 22_{nd} generations and the initial population for Case study 2 - Individual solutions A,B,C,D,E and F.

front, no other generation contained individuals that dominated any of these. If we take a look at the two Solutions presented in the 100_{th} generation (B and C), we see that they are not very dissimilar from each other. Solution C contains no curvature in the transverse section, making it lighter that solution B. The latter has a curved transverse section, giving it some bending stiffness, in fact increasing its fitness, but making it heavier.

The 22_{nd} generation is an interesting set of individuals to study. We can see that already after 22 generations the improvement from the initial population is significant. This set however, contains a much more diverse group of individuals when compared to the 100_{th} generation. Solution D has a very pronounced cross-section, while E and F have less marked ribs, solution F has a more continuous curvature, while E is wave-formed.

Solution E seems to be in an almost orthogonal position from solutions D and F. In fact, if we look at the overall shape of the Objective space, and we trace lines from solution A (not found by the search process) and the individuals in the Pareto front (generation 100), we note a different pattern that we had not seen before. In section 7.6 we saw that the shape of the Pareto Front can be a clue as to the level of contrast of a given problem. But in this case we see an almost orthogonal pattern, as if one function has very little influence over the other. This denotes a low level of contrast between the two fitness functions.

The difference in fitness for f_1 between solution A and B or C is gigantic, while the difference for f_2 is negligible when compared to the other distances we can see in the objective space. For example, the difference in f_1 between D and C is more that 6 times larger that the difference between A and C. These numbers tell us that we are gaining little weight from A to C, but on the other hand, we are reducing $max(\Delta_{Z_i})$ by 96%.

This low contrast is true for this Case Study and in these sections of the objective space, however as we can see in the 100_{th} generation, there is still contrast present in the problem. Other structural fitness functions could be investigated in order to see if the level of contrast changes, or if different measures give is the same patterns.

These long orthogonal segments might also be a clue as to why NSGA-II is having problems in finding solution A or solutions between A and the Pareto Front in the 100_{th} generation. If we a perfectly orthogonal pair of fitness functions, the Pareto front would be made up of only 3 solutions. This means that we would have empty segments between then, making it impossible for the MOGA to place solutions other than these 3 in the front. A pair of fitness functions that are almost orthogonal, but not exactly, should be a similar case. We can hypothesize that this is the behavior we have seen

in the last Case Study.

10.4 Masonry shells

Compression only structures, such as masonry and stone vaults, have a long and rich tradition, they are a fundamental part of the history of architecture and structural design. Since stone and masonry posses very little resistance to tension, their structural capacity relies a great deal on their geometry. They are designed to have compression only shapes, shapes that do not generate tensile stresses. The most iconic shape in this category is the arch, and more specifically the catenary arch. Shells, being three-dimensional objects can have more complex shapes than the arch. Traditional compression only shells were made in a vast array of shapes such as barrel vaults or sail vaults and they all have precise and well known geometry and proportions that make them compression only shapes.

Free-form compression only forms have also been studied and built in the past, perhaps more famously by Antony Gaudí. Recently the term structural form-finding has been given to the process of generating free-form compression only shapes by means of computation. As we saw in section 4, there are many ways of computing these shapes, and many of them do not involve digital computation. Several Digital form-finding processes exist, for example dynamic relaxation (Day 1965), force density method (Schec 1974), particle-spring systems (Kilian & Ochsendorf 2005) and Thrust Network Analysis (Block 2009). All of these methods tend to be very interactive, in that they need a user defined starting configuration, and many of them can be modified in such a way as to generate different compression only shapes almost in real time. They can be thought of as deterministic and gradient based methods.

The above methods and their applications demonstrate that compression only shapes can be very different from each other and from traditional masonry vaults. However, they are not completely free-form, they represent a sub-group of shapes that satisfy the compression only requirement. The above mentioned methods generate forms that reside within this sub-group, and the designer can interactively generate different shapes, exploring geometry that lies within this sub-group. From this point of view we can think of these form-finding processes as optimization methods rather than search methods. Genetic algorithms, while not being as fast and as efficient as the above mentioned methods in finding compression only forms, they are more powerful form a search point of view. They consider geometry groups or

sub-groups depending on the parametrization given by the user and they can consider these groups from other points of view, apart from the compression only shape. The selection of a GA or a form-finding method should be taken depending on the problem at hand and the user's needs.

In this section we will employ MOGAs to study masonry shell structures.

10.4.1 Structural analysis of masonry vaults

This case study introduces the use of masonry, a material that differs from reinforced concrete in that it does not have a high resistance to tension forces. Therefore, important changes needed to be made in the fitness function that studies its structural capacity and the FEM model that analyses its behavior.

In finite element structural simulations, materials and their mechanical properties are introduced by means of series of values, most important of which is the Elastic or Young's Modulus (E). The elastic modulus characterizes the stiffness of the material as the ratio of stress σ over deformation ϵ:

$$E = \frac{\sigma}{\epsilon} = \frac{F/A_0}{\Delta L/L} \tag{10.7}$$

where F is the force applied to the material, A_0 is the cross section area of the object to whom the force is applied, L is the length of the object and ΔL is the variation in length of the object due to deformation. E is expressed in units of force over units of area, this means it is a pressure unit, it is usually expressed in $Pa = N/m^2$, $MPa = N/mm^2$ or $GPa = kN/mm^2$.

The elastic modulus for reinforced concrete and masonry depend on their specific materials and configurations, but they can be around 30 GPa for concrete and around 10 GPa for masonry. In this way FEM calculations take into account the difference in stiffness of the materials. But the fragility of masonry under tension forces is not taken into consideration. Linear FEM calculations assume the material given is a perfectly elastic and symmetrical material, meaning that it has the same resistance under tension and compression forces. This might be the case for concrete under normal loads, but it is not the case for masonry, therefore we need to somehow take this into account when we study masonry shells.

A brief parametric study of masonry shells is made with the purpose if understanding how structural FEM simulations could describe masonry vaults, and how to take into account the characteristic of masonry described

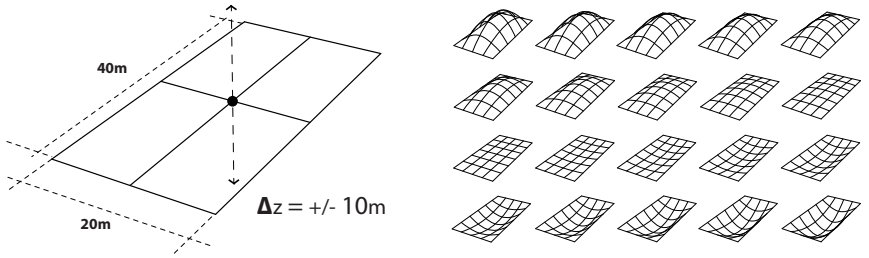


Figure 10.13: Parametric model for study of elastic potential energy and maximum tension of masonry shells - The 20 generated shells.

above. Figure 10.13 shows the parametric model made for this analysis, it contains 20 shells in total. The only variable in this analysis is the z-axis coordinate of the central control point of the surface. We can see that the generated surfaces are all positive double curvature surfaces, when $z > 0$ we get concave surfaces and when $z < 0$ we get convex surfaces. The only exception is the flat surface that occurs when $z = 0$.

Figure 10.14 shows the FEM model characteristics for this analysis, the shell is supported by pin joints along its edges, and each discrete shell element is loaded with its own weight. The surface is modeled with a fixed and continuous thickness of 5 centimeters. These FEM analysis are made by means of Oasys GSA, a commercial FEM software.

The purpose of studying concave and convex surfaces is to see the difference that the FEM analysis gives to the convex surfaces (that should be mostly in tension and therefore not viable for masonry vaults) and the concave surfaces that should be mostly in compression. The first analysis made is to plot the elastic potential energy U_e and the maximum tension stress $max(\tau^+)$ for all of the surfaces. We can call this a sensitivity analysis of the surface shape for these structural values.

Figure 10.15 shows the result of the analysis. We can see that the elastic potential energy U_e curve is symmetrical, it has the sane values for concave and convex surfaces, for example $z = -4$ and $z = 4$. This is to be expected since the FEM analysis considers the material to be perfectly elastic and symmetrical. The maximum tension $max(\tau^+)$ curve on the other hand is not symmetrical. We can see that the convex surfaces have higher tensions than their concave counterparts. Since the concave surfaces in this parametric model are not perfect compression only shapes, there are points in the

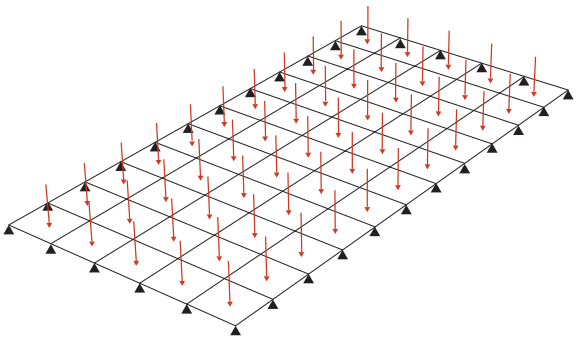


Figure 10.14: Loading and node constraint conditions for the parametric study, the 40×20 masonry shell.

surface which are in tension, but their values are lower that the ones found in the convex surface. This gives us the opportunity to create a fitness function that can identify surfaces that have low tension and low overall stresses.

The elastic potential energy describes the general condition of the structure. The higher the elastic potential energy, the higher the structure is being deformed, indicating a lower stiffness in the structure. For this reason, it is often used in GAs as a fitness function, it is a single number indicator of the overall stiffness. But we can see in this case that U_e is not a good indicator for masonry structures, because it does not distinguish between structures under tension or compression. We need to somehow include the tension into a new fitness function in order to help the GA make this distinction.

But maximum tension on its own is also insufficient for a proper fitness function. For example, if we try to compare perfect compression only shapes by calculating their $max(\tau^+)$ values, we would see that they both have $max(\tau^+) = 0$. This means that using only $max(\tau^+)$ as a fitness function would not allow us to distinguish between compression only shapes. Compression only shapes however, do have different compression stresses in them, and we can explore between them to find surfaces that minimize also compression stresses.

The following function parametric developed to incorporate both U_e and $max(\tau^+)$:

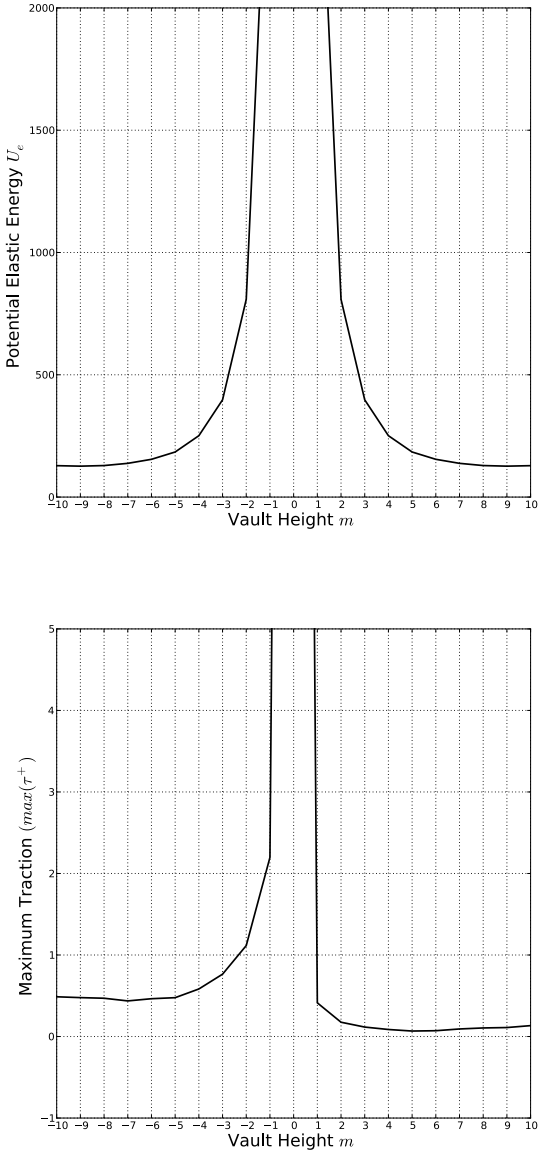


Figure 10.15: Potential elastic energy U_e for 20 masonry vaults - Maximum tension ($\max(\tau^+)$ for the same vaults.

$$fit = \frac{U_e}{U_{e,0}} + \left(\frac{max(\tau^+)^2}{max(\tau_0^+)^2} \cdot w \right) \tag{10.8}$$

where $U_{e,0}$ is the elastic potential energy for a reference shell, $max(\tau_0^+)$ is the maximum tension for that same reference shell and w is weight a coefficient. The function normalizes both U_e and $max(\tau^+)$ with a reference shell and then the tension is given a higher influence in the function by multiplying it by w. The reference shell should be a low performing shell, in this case the flat surface parametric selected.

Figure 10.16 shows the result of the parametric study for the fitness function (fit) described above, for different values of w. We can see that when $w \neq 0$ there is a good distinction between the convex and concave shell structures, but also there is a less noticeable but existing distinction between the different concave shells. This figure suggests that the developed fitness function is capable of studying masonry shells with the linear FEM analysis and with an appropriate distinction between tension and compression stresses.

10.5 Case Study 3: Free-form masonry roof

The first case study for masonry vaults is a shell roof for a rectangular building of 40 × 20 meters. This Case study has several similarities with the previous two case studies, it is also a multi-objective search problem with contrasting objectives for structural design and it is also studied using NSGA-II. The main difference, apart from the shell dimensions, is the material, and the fact that in this case the shell is supported along all of its perimeter.

10.5.1 Parametric Model

Case study 3 also employs NURBS geometry for the generation of the surfaces. In this case only nine control points were used, and only five of these are variable. Figure 10.17 shows the parametric model setup for case study 3. It shows the location of the five variable control points and their range of movement, they are free to move -10 or +10 meters from the flat configuration shown in figure 10.17. In this parametrization scheme the four corners of the shell are kept fixed, but it allows the edges of the surface to move freely. These edges are completely structurally supported, this means that

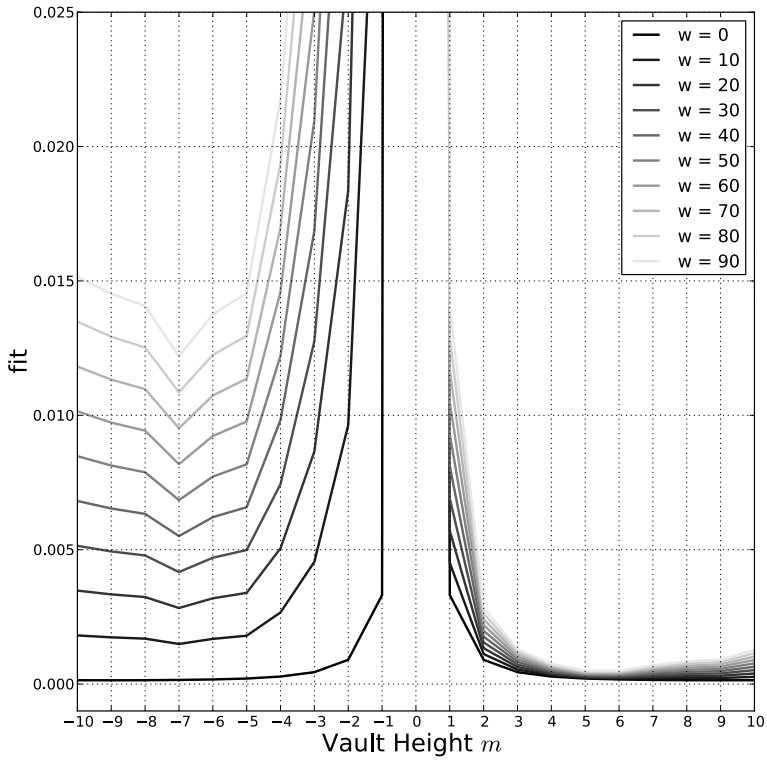


Figure 10.16: Parametric Study of the proposed Fitness function for masonry vaults with variable weight values w.

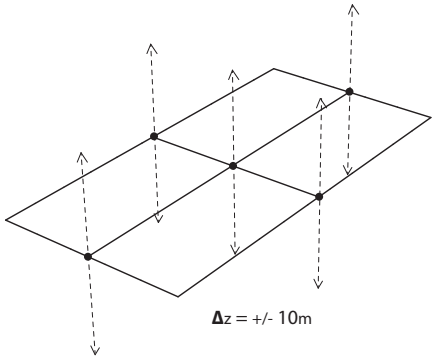


Figure 10.17: Case study 3 parametric model - 40×20 masonry shell.

they can move in space, but independently of the shape the edges take, they will always be supported by pin joints.

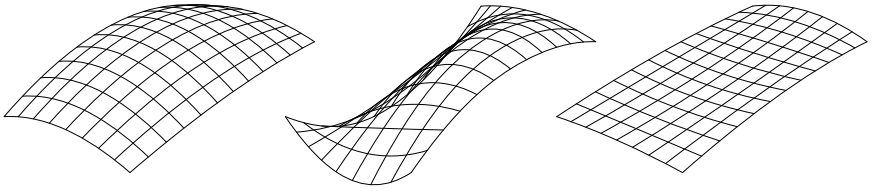


Figure 10.18: Possible individuals with the parametric model developed for case study 3.

In this parametric model all of the usual possible configurations are included: flat, simple and double curvature surfaces, both convex or concave as well as multiple curvature surfaces. Figure 10.18 shows a few possible configurations in this parametric scheme. We can see concave double curvature surfaces and complex multiple curvature surfaces, one with very noticeable curves and another that only slightly diverges from the flat configuration.

10.5.2 Fitness functions

This case study, as the two case studies before it, is intended to generate shells that are both light and structurally efficient. The first fitness function describes the masonry shell's structural capacity by means of the function described above in equation 10.8. The reference surface surface used to calculate $U_{e,0}$ and $max(\tau_0^+)$ is the flat surface. The FEM model used for this case study is the same one used for the above mentioned analysis, seen in figure 10.14, where the shells are supported by pin joint in all of the edges, and loaded with their own weight. The thickness of the shells parametric fixed at 5 centimeters.

The shell's weight parametric used as the second fitness function, and as in previous case studies, it parametric simplified as the surface area since thickness and material were fixed. The fitness functions for this case study can be expressed as:

$$\textit{Case Study 3} \begin{cases} \textit{Minimize} & f_1(x) = \frac{U_e}{U_{e,0}} + \left(\frac{max(\tau^+)^2}{max(\tau_0^+)^2} \cdot w \right), \\ \textit{Minimize} & f_2(x) = S, \\ \textit{subject to} & -10 \leq x_i \leq 10. \end{cases} \tag{10.9}$$

10.5.3 Genetic algorithm inputs

NSGA-II parametric used to search for solutions to this problem. It explores 100 generations with 50 individuals in the population. The GA inputs were the following:

Case Study 3	
Population Size (N)	50
Number of Variables	5
Number of binary digits	8
Variable Domains	$x_i \in [-10, 10]$
Mutation Probability (p_m)	0.2
End Condition	End after 100 generations

10.5.4 Results

After 100 generations NSGA-II produced the results shown in figure 10.19. The first objective space shows the complete figure with individuals from the first generation up to the last. We can see that the Pareto front has

a very orthogonal shape in this space, with solutions that are quite good at one function and not good at the other, as well as some solutions that are very good at both. This shows a low level of contrast between the two functions at this scale. The second objective space shows a detailed view of the solutions at the intersection of the orthogonal functions, solutions that are good at both functions.

We can also see some of shapes of the solutions in the Pareto front. With the exception of solution A, They are mostly concave and very symmetrical surfaces with different levels of height. Interestingly they all have arched edges. Solution C has a very low surface area while achieving a good amount of stiffness with its shallow concave shape. Solution A is a member of the Pareto front but mostly because of its very small area, its almost a flat surface and does not have a stiff shape for masonry construction. The stiffness difference between solutions C and D is not very big, but their difference in weight is more noticeable.

The stiffest solution in the entire set is solution B, it is the highest surface in the front, but not the heaviest in the set. We can expect the stiffest solution is not going to be the lightest, but the opposite is not true. Higher surface area does not mean a better shape is achieved, and also, since the surfaces are loaded by their own weight, the higher the weight the more load and perhaps a lower structural performance.

10.6 Case Study 4: Free-form masonry roof with variable thickness

The previous case study found optimal shapes for both stiffness and weight while having a fixed surface thickness. Case study 4 includes surface thickness as a problem variable, meaning that the GA can now search for stiffer solutions using the thickness, while also evaluating shell weight. Apart from this new variable, all of the other problem characteristics remain the same as in case study 3.

10.6.1 Parametric Model

Figure 10.20 shows the parametric model for case study 4. The variable control points make up the first five variables, and the shell's thickness is the sixth one. The shell's thickness can vary from a minimum of 5 centimeters to a maximum of 100 centimeters. The four corner control points of the

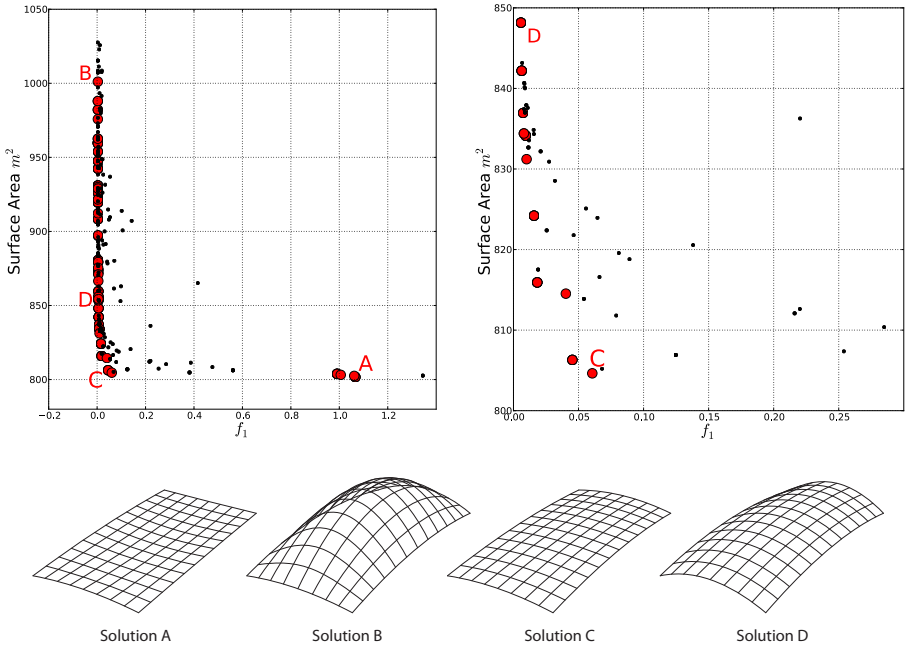


Figure 10.19: Objective space with Pareto fronts for all generations for case study 3 at different levels of detail - Individual solutions A,B,C and D.

NURBS surface remain fixed as was the case in the previous study. We can sum up the variables in this case study in the following variable vector X:

$$X = x_1, x_2, x_3, x_4, x_5, x_6 \tag{10.10}$$

where $x_{1-5} = z_{1-5}$ and $x_6 = thickness$.

10.6.2 Fitness functions

The first fitness function in this case study is related to the shell's structural performance. The structural capacity of the shells are measured by the function explained above in equation 10.8. The reference surface for the calculation of $U_{e,0}$ and $max(\tau_0^+)$ is the flat shell with the minimum thickness value (thickness = 5cm).

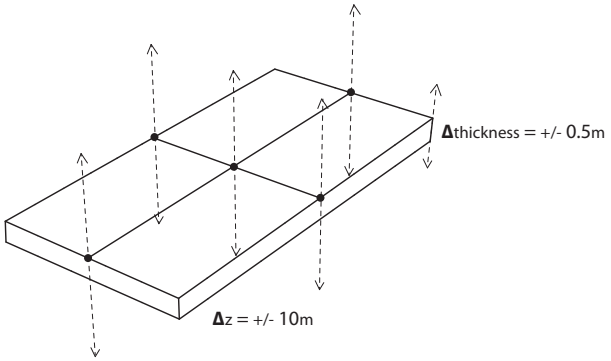


Figure 10.20: Case Study 4 parametric Model - 40×20 masonry shell with variable thickness.

The second fitness function measures the weight of the surfaces, but since in this case the thickness is variable, the usual simplification of surface is not viable, in this case the shell's volume is used. The fitness functions for this case study can be expressed as:

$$Case\ Study\ 4 \begin{cases} Minimize & f_1(x) = \frac{U_e}{U_{e,0}} + \left(\frac{max(\tau^+)^2}{max(\tau_0^+)^2} \cdot w \right), \\ Minimize & f_2(x) = V, \\ subject\ to & -10 \leq x_i \leq 10. \end{cases} \tag{10.11}$$

where V is the shell volume.

10.6.3 Genetic algorithm inputs

NSGA-II is employed for this case study as well. It explores 180 generation with a population size of 50 individuals. The GA inputs for this case study are detailed bellow:

Case Study 4		
Population Size (N)	50	
Number of Variables	5	
Number of binary digits	8	
Variable Domains	$x_{1-5} \in [-10, 10]$	$x_6 \in [0.05, 1.0]$
Mutation Probability (p_m)	0.2	
End Condition	End after 180 generations	

10.6.4 Results

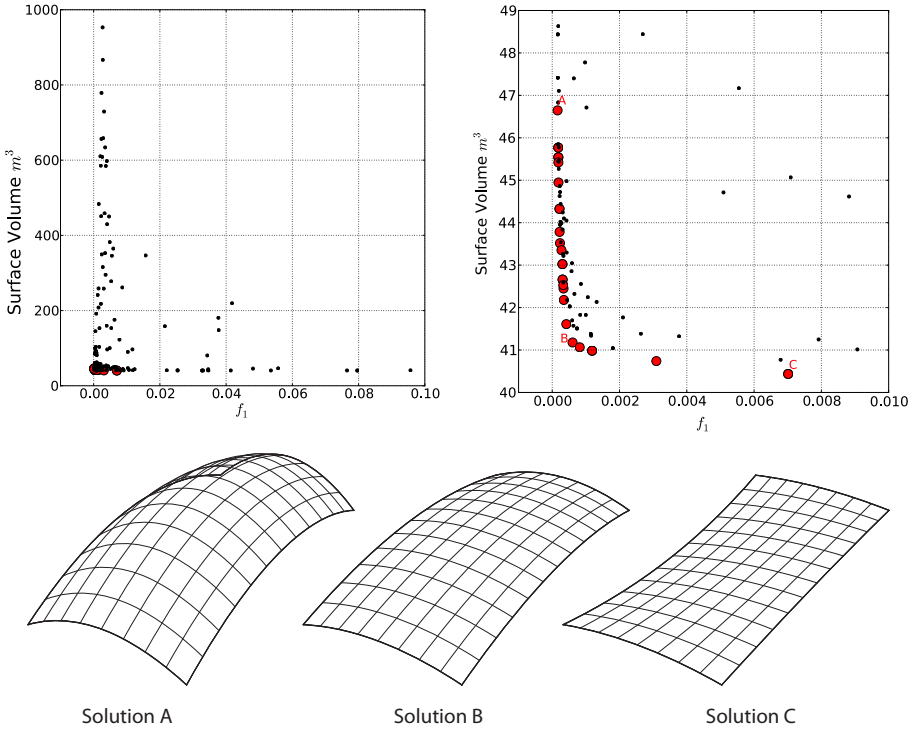


Figure 10.21: Objective space with Pareto fronts for all generations for case study 4 at different levels of detail - Individual solutions A,B and C.

Figure 10.21 shows the objective space with two levels of detail for case study 4. It shows an orthogonal configuration of this case study as well. The vast majority of the individuals in the Pareto front are quite far from the limits of the objective space, meaning that they are optimal when compared to the limits. The detailed view of the Pareto front shows very little contrast between the two functions. Be best performing solution from a structural point of view (solution A) is only 15% heavier that the lightest one in the front (solution C).

Solutions A and B are very similar double curvature concave shells with different heights. Solution C is an almost flat surface, but it is a hypar surface. It has concave cross-sections and convex longitudinal sections.

Since they share their fitness functions and materials, results for this study can be compared to those in case study 3. We can see that the Pareto individuals for the structural function reside in a much narrower range, they have better fitness values. This can be explained by the fact that this case study ran for 80 generation longer that the previous one. If we compare the two cases from a structural point of view, we find that solution A from case 4 has a better performance than the best one from case 3 (solution 3-B). In fact solution 4-A has a performance value that is twice as good as 3-B, and it is also lighter.

An important result that we can point out is the fact that all shells in the Pareto front are 5 centimeters thick. In fact, after just a few generations the GA considered only solutions with very low thicknesses, excluding solutions with high thicknesses. After generation 80, the entire population is made up of solutions with 5 centimeters of thickness.

11

Load bearing masonry walls

We have seen so far how Search algorithms help us gather useful information with complex shell surfaces. It stands to reason that search algorithms are naturally useful in complex geometries, and complex problems, and less so in more established problems with very simple geometry. But there are still very good advantages to the use of search algorithms in simple cases, their exploratory power can be put to good design use, and the information obtained might not always be obvious or otherwise easy to obtain. This is especially the case with simple geometries but with multi-disciplinary problems. In this cases, not matter how simple the geometry, the problem might be highly complex.

In this chapter we investigate the structural capacity of masonry load bearing walls will variable thicknesses and window arrangements. We will look into the complete structure of a rectangular building, considering different set of variables for all four sides of the rectangle. Windows are variable in quantity, dimension and position within the walls. As in previous structural case studies, we will contrast the structural performance of the masonry buildings with their weight, or more precisely, with the volume occupied by the masonry.

11.1 Structural analysis of load bearing masonry walls

As was the case with the masonry shell structures, the load bearing masonry walls need to be studied with great care to take into account the properties

of this material. Masonry has no resistance to tensile stresses and therefore the use of linear FEM calculations need to be studied with this issue in mind. Elastic potential energy U_e values alone can be misleading. A more comprehensive analysis, yet simplified for the early design stage, is presented in section 10.4.1. It makes the distinction between tension and compression forces, in oder to study the generated shape accordingly. The same approach is employed in this chapter for the study of masonry walls.

11.2 Parametrization of walls and windows

Simple geometries tend to have simple parametric models, especially with the use of sophisticated CAD software. However, a parametric study made for the purposes of search algorithms is only as good as the possibilities it can offer the designer. Parametric models should be able to generate all of the possibilities that the designer wishes to investigate without creating problems for the search algorithm.

In this case, the parametric model needed to include the possibility of generating different size rectangular windows, in any position in the wall. Also, the most important requirement in this model, would be the possibility to allow the search algorithm to select not only the shape and position of the windows, but the amount of windows as well. The search algorithm should be able to generate as many windows as required in each wall, and also have the possibility to generate no windows at all. This important issue represents an interesting challenge in the programming of the parametric model. Wall thickness on the other hand presented no challenges, as in the case of the shell surfaces.

11.2.1 Isomorphism: A failed parametric model

A first attempt to parametrize the above described model is made following the same ideas presented in (Wright & Mourshed 2009). Wright and Mourshed present a parametric model of a wall with variable windows for energy efficiency optimization. Figure 11.1 shows the parametric scheme used by Wright and Mourshed. They model the wall and their windows in a binary way. They discretize the wall area into square elements, then generate a binary number with the same number of digits as the number of elements. A window is assigned to those squares associated with a 1, and wall for those elements associated with a 0. In this way, the parametric model is capable of generating windows in variable numbers, size and position. However, the

generated windows are not strictly rectangular, as specified in our case. As seen in figure 11.1c, using this scheme, windows can be joined at their vertices, be discontinuous and also have wall elements inside them. This is not desired in our case, therefore a modification of the model has to be made.

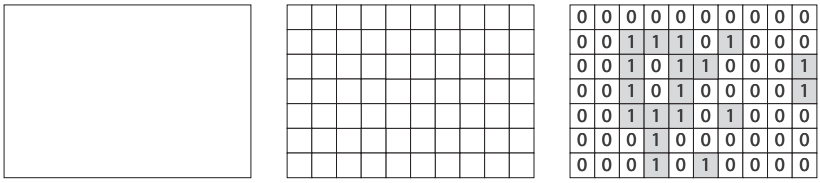


Figure 11.1: Parametric modeling scheme for wall with openings used in (Wright & Mourshed 2009).

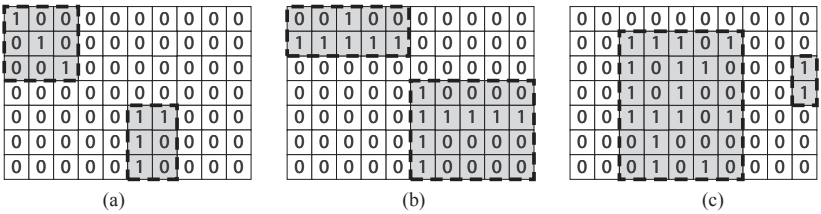


Figure 11.2: Parametric modeling scheme for wall with openings modified to include only rectangular and continuous windows.

Figure 11.2 shows the modification made to the model presented by Wright and Mourshed. An additional rule is implemented, where by window segments directly adjacent to each other, or window segments sharing a vertex, are combined into one single window. Figure 11.2c has the exact same binary input as figure 11.1c, but with very different resulting windows. With the addition of this rule, only rectangular windows are generated, and none of them have wall segments inside.

This modification however had negative unintended consequences, not present in the Wright and Mourshed model. Figure 11.3 shows 3 different binary digits, all of them generating the same window configuration inside

Figure 11.3: Three Isomorphic walls with different binary inputs.

the wall. This means that different genetic inputs generate exactly the same individual and therefore the same fitness function values. In genetic lexicon, this phenomenon is known as "encoding isomorphism" (Wang et al. 2006).

> "Encoding isomorphism means that chromosomes with different binary strings may map to the same solution in the design space. This leads to representational redundancy, which is not beneficial for the GA if the genetic operators cannot gain useful information from representational variants."

(Wang et al. 2006)

Another way of looking at encoding isomorphism is to look at the probabilities of generating different solutions at random. An ideal parametric model should provide the exact same probability of being generated to all individuals in its domain. All possible combinations of variable values should generate one single solution each, and all solutions being completely different from each other. When isomorphism starts being present in the model, there are a few solutions that have more instances of themselves in the parametric domain. Hence, there are higher probabilities of them being selected in a GA. Since this higher probability is not related to fitness values, but only encoding isomorphism, this higher probability is a problem. It influences the GA to select some solutions over others for the wrong reason. In the parametric model just presented, solutions with very large windows were far more likely to be selected that solutions with very small windows. The encoding isomorphism in the presented parametric model is so severe, that the GA, in hundreds of random initial individuals, is unable to generate a solution different from a wall containing only window squares and no wall areas. A better parametric model for this case study needs to be developed.

11.2.2 Window Area of Influence

A second attempt of parametrizing the above described problem is developed by means of subdividing the walls into areas of influence for each window. Figure 11.4 shows this concept for a single wall, subdivided (a) one single area, and (b) into 4 areas. Only one window can be drawn in each area, but the model is such that an area may or may not have a window.

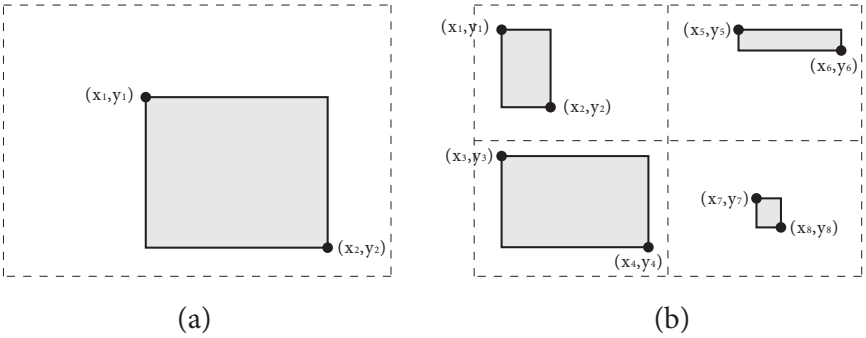


Figure 11.4: Parametric scheme following a window area of influence - (a) with one single area of influence - (b) with 4 areas of influences - All with 4 variables for each window.

Each window has 4 variables, x and y coordinates for 2 window corners. Since all windows are on a flat plane (the wall) there is no need for a third coordinate z. These coordinate values are normalized into values from 0 to 1. The four variables for each window can be expressed in the following way:

Variable	Point	Coordinate	Domain
x_1	1	X	0 to 1
x_2	1	Y	0 to 1
x_3	2	X	0 to 1
x_4	2	Y	0 to 1

In these terms, a window with variable values $x_1 = 0$; $x_2 = 0$; $x_3 = 1$; $x_4 = 1$ would result with the maximum sized window for that area. An additional rule has to he determined for the parametric model to be able to exclude windows from any given area. The rule establishes a minimum

window dimension, for example 40cm. In this case any windows containing x or y dimensions under 40cm would be excluded.

Following this parametric formulation we can obtain rectangular windows of variable size, dimension, position and number, just as we set out to do in the problem formulation. There are however two limitations:

- The maximum number of windows is determined by the number of areas defined in the model. As previously stated, some areas can be empty, so the total number of windows is variable, from 0 to the number of areas. This number would be user defined, so it represents a small problem.

- There is no possibility of generating one single window covering the entire wall if more than one area is defined. When more that one area is present in the model, windows cannot be combined into one large window. Attempts to create a parametric model capable of combining windows led to isomorphism problems and were abandoned.

These limitations mean that separate GA runs need to be made for different area configurations. The parametric model cannot change the number of areas during its search process.

11.3 Mesh Discretization of walls with windows

The previous structural case studies included only continuous surfaces. They were mostly complex and double curvature surfaces, but they were all continuous, they had no windows or openings of any kind. This means that they were discretizable in a continuous grid of rectangular or triangular elements. Surfaces with openings are more complicated to discretize. FEM software, such as the ones employed in this PhD research, require certain kind of mesh geometry in order to have accurate results. There are two important characteristics that meshes are required to have:

- The angles between two mesh edges cannot be very small, the more they are 90 degrees for rectangular elements, and 60 degrees for triangular elements, the better the results. In the case of Oasys GSA (the FEM software used for this case study) if angles are not adequately sized the program will throw and error and stop the analysis.

- All of the lines or edges in the original geometry, such as wall or window edges, need to be discretized in several lines, therefore more than two points are needed. A good rule of thumb is to use 4 points minimum.

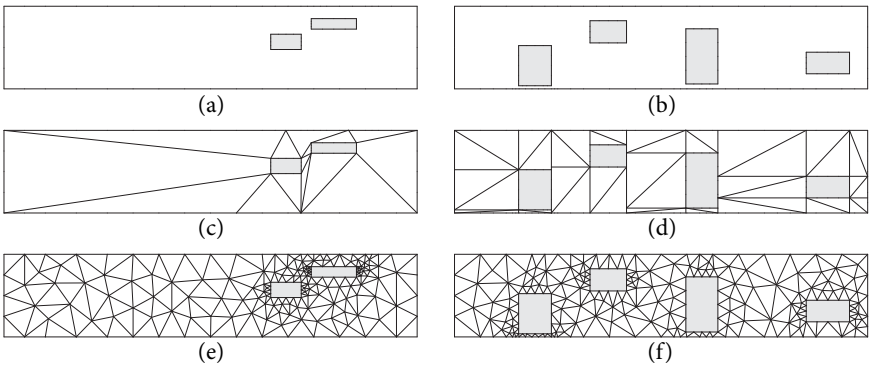


Figure 11.5: Correct and incorrect meshing of 2 walls with different window arrangements - (a) and (b) show the two original arrangements - (c) and (d) are the incorrect meshing of the two walls - (e) and (f) are the correct versions.

Figure 11.5 shows some examples of correctly and incorrectly discretized surfaces with rectangular openings. Since parametric models generate solutions that are quite different from each other, a meshing algorithm capable of respecting the above mentioned rules is needed. For the present PhD thesis Mehpy is used (Klockner n.d.). Meshpy is a Python wrapper for a meshing algorithm developed by Jonathan Richard Shewchuk called Triangle (Shewchuk 1996).

Triangle can generate meshes that contain openings, and that can respect the above mentioned rules. As its name suggests, Triangle generates only triangular meshes. For the purposes of this PhD research, a minimum angle of 30 degrees is imposed to all triangular mesh elements, and a minimum number of nodes of 4 is set for all wall and window edges.

Since triangle works only on two dimensions, in order to generate complete buildings, comprised of 4 separate walls, a few additional operations needed to be implemented. The four walls were calculated by Triangle in one single 2d mesh (figure 11.6), that is then folded into a 3d mesh by a special algorithm developed by the author. The first edge of the first wall

needed to be "welded" with the last edge of the fourth wall, as shown in figure 11.7. A slab or roof element also needed to be generated. A separate triangle calculation is therefore implemented, one that generated the slab elements, taking the wall elements as a starting points. In this way, the nodes generated for the slab coincided perfectly with the top nodes of the wall mesh. An additional "welding" operation is then carried out, this obtaining a single mesh that contains all 4 walls and the slab, and no redundant or overlapping nodes or edges are present. A diagram of the entire process is shown in figure 11.7.

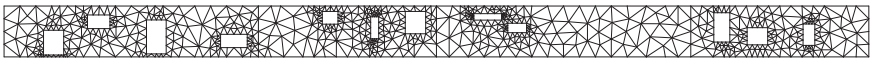


Figure 11.6: Unfolded walls for rectangular wall structure with windows.

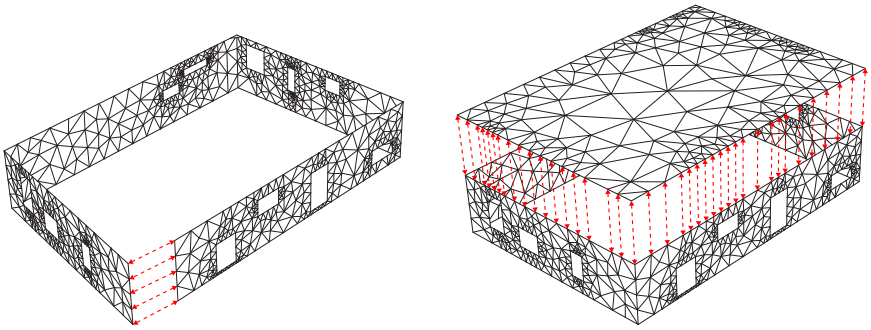


Figure 11.7: Folding and Welding process for wall structure meshes.

11.4 Case Study 5: Load bearing masonry walls

Case study 5 is a multi-objective search problem of a building supported by load bearing masonry walls. The building has a 20×14 rectangular plan and 6 floors, all containing offices. The structure is analyzed in only one

story, the idea being to study a standard floor. For this purpose the first floor is chosen, the ground floor is not chosen because it is a special floor, and not a repeatable one like the first floor. Each standard floor is 4 meters in height.

The purpose is to generate window and thickness configurations for each wall in a rectangular building that minimize the wall's weight and maximize structural performance.

11.4.1 Parametric Model

The building is parametrized following the window area of influence scheme described in section 11.2.2. Each wall is subdivided into 4 areas that cover all of the height. The thicknesses of the four walls are defined separately, so they are also variables. The parametric model for case study 5 is shown in figure 11.8.

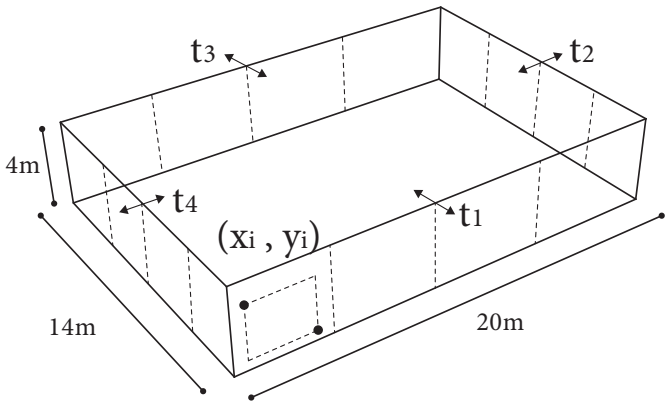


Figure 11.8: Parametric model for case study 5.

Since we have 4 variables for each window, and 4 areas for each of our 4 walls, there are in total 64 variables for the window configuration. To this we add the 4 thicknesses of the 4 walls and we get a total of 68 variables.

11.4.2 Fitness functions

Previous structural case studies in this thesis were mainly focused on roof structures and therefore were studied as such. In this case we are considering the main structural elements of an small office building. Hence it is necessary to subject the structure to more rigorous loads to determine their performance.

A total load is calculated from the buildings floor area and the number of stories. A load of 10kN were used for each m^2. With a floor plan of 20 $\times$ 14 m^2 $\times$ 5 stories, a total load of 14.000 kN is used. This represented the total vertical load applied to the structure. But the structure is also studied from a seismic point of view. Two separate loading cases are employed, each one of them includes the vertical load of 14.000 kN, plus one horizontal load equivalent to 10% of the vertical load. The difference between the two loading conditions is the direction of the horizontal loads, as seen in figure 11.9. The first loading case has a horizontal load parallel to the long dimension of the building, meaning that the 20m long walls would be the ones most responsible for resisting the horizontal load. The second loading case has an horizontal load parallel to the short dimension of the building, in this case being the 14m walls most involved.

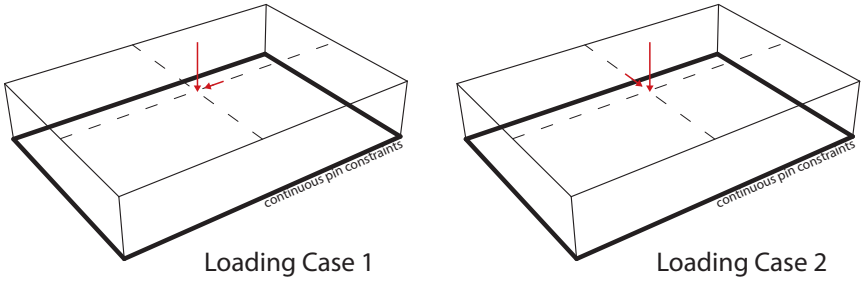


Figure 11.9: FEM model setup and loading cases for case study 5.

From the dimensions of the building, under equal thicknesses and window arrangements, we could expect that the second loading condition would be the most critical. However, since the parametric model used in this case study modifies window arrangements and wall thicknesses, it is not possible to know this beforehand. It is therefore necessary to calculate both

loading cases for all candidate solutions. The solutions are then evaluated considering the most critical case for each of them individually.

The masonry walls are studied with the same fitness function created for the shell structures, detailed in equation 10.8. As it was the case for the shell structures, a reference solution is needed in order to calculate our structural fitness function. More specifically, we need to set $max(\tau_0^+)$ and $U_{e,0}$. In this case, the most regular and strongest structure is selected, the structure without any windows and with the thickest walls. This selection implies the fact that the highest fitness values would be the one where $max(\tau^+)$ and U_e are lower or equal to $max(\tau_0^+)$ and $U_{e,0}$ respectively. The only modification would be the fact that it is calculated once for each loading case, and the one with the highest value is used. This modification can be expressed as follows:

$$\begin{aligned} & fit = max(f_{case1}; f_{case2}) \\ where \quad & f_{case1} = \frac{U_e}{U_{e,0}} + \left(\frac{max(\tau^+)^2}{max(\tau_0^+)^2} \cdot w \right) \\ & f_{case2} = \frac{U_e}{U_{e,0}} + \left(\frac{max(\tau^+)^2}{max(\tau_0^+)^2} \cdot w \right) \end{aligned} \tag{11.1}$$

This means that the FEM simulation is executed twice for each solution studied, and the fitness values used would be the highest one. This case study has significantly higher calculation times, not only because of the two FEM calculations, but also because the number of nodes, and elements in each solution is significantly higher than in the previous cases.

The second fitness function in case study 5 is the weight function, simplified as the wall's volume. The search problem put forth in this case study can be defined as follows:

$$Case\ Study\ 5 \begin{cases} Minimize & max(f_{case1}; f_{case2}), \\ Minimize & f_2(x) = V \end{cases} \tag{11.2}$$

11.4.3 Genetic algorithm inputs

The number of variables in this case study is significantly higher than the ones seen in previous case studies, and the calculations times are higher as well. This represents a big challenge for the MOGA. In this chapter we present the results found with the following genetic inputs:

Case Study 5		
Population Size (N)	50	
Number of Variables	68	
Number of binary digits	8 for window points (x_{1-64})	6 for thicknesses (x_{65-68})
Variable Domains	$x_{1-64} \in [0, 1]$	$x_{65-68} \in [0.05, 1.0]$
Mutation Probability (p_m)	0.2	
End Condition	End after 100 generations	

A population size of 50 individuals and 100 generations might or might not be enough for us to find a final Pareto front. The results obtained in this chapter are presented as found, no metric to determine convergence have been done.

11.4.4 Results

Figure 11.10 shows the objective space and Pareto front for case study 5. It shows a Pareto front that is not as orthogonal as some of the ones we saw in other structural case studies, suggesting a higher level of contrast between f_1 and f_2 in this case.

Figure 11.11 shows 3 examples of Pareto dominant solutions for case study 5. Solution A is a structurally high performing solution, but it is a heavy one. It is characterized by thick walls and very small windows. Structurally high performing solutions in the front have very small windows, or none at all in the short walls. This suggests that they can develop the most strength for loading case 2 in this way. As previously mentioned, loading case 2 is expected to be the most critical under most conditions, and since f_1 selects the most critical case as a fitness value, it is reasonable to expect the small walls to be the object of the most attention, but this is not always so. Another important characteristic of high performing solutions for f_1 is high thicknesses. But thickness is not found to be distributed evenly among the four walls, in most cases, the thickest wall is the long wall opposite the horizontal load in loading case 2. This is most evident when looking at solution A. This fact is evidence that not only walls parallel to the horizontal walls are structurally significant, when thicknesses are high in perpendicular walls, they begin to contribute significantly to buildings structural rigidity.

Solution C is the best performing solution for f_2 (the lightest solution in the Pareto front). As is to be expected, light solutions tend to have very small thicknesses on all 4 walls and larger windows in them as well. Solution C shows the largest windows in the longer walls, while still preserving some thickness for the same long wall opposite the horizontal load in case 2.

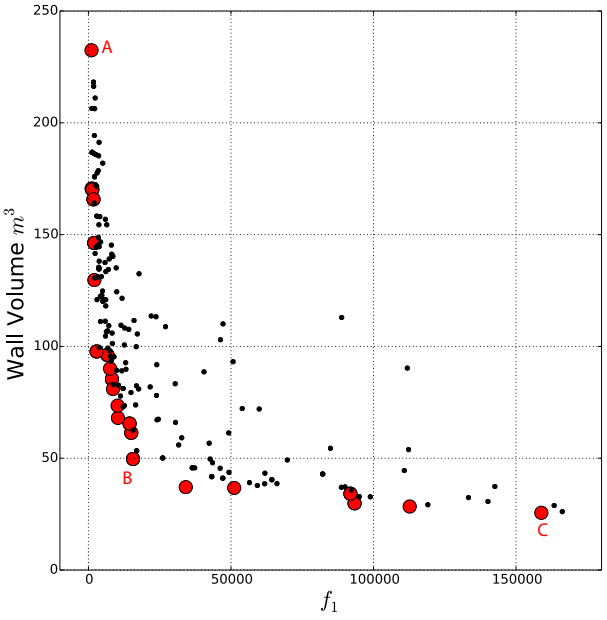


Figure 11.10: Objective space for case study 5 - Pareto Front in red.

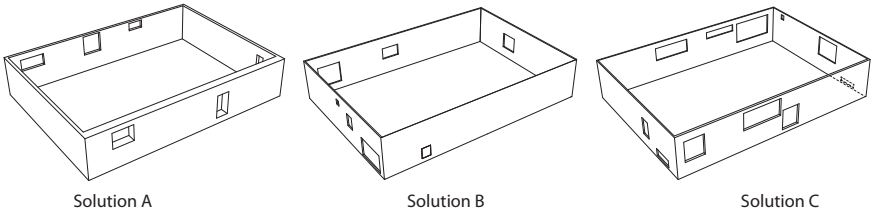


Figure 11.11: Case Study 5 result structures.

Solution B is a good compromise for f_1 and f_2. It is also a very interesting exception to most other solutions in the front. It does not have a thick long wall, on the contrary, it has a very thin one. Most other solutions in the Pareto front are similar to solution A, It also has an interesting diagonal window pattern in one of its short walls. This seems to suggest that a higher level of exploration could accomplish very interesting window patterns that increase rigidity while maintaining very low thicknesses.

Another interesting feature of the results of case study 5 is the fact that solution B is quite different from other solutions close to it in the front. With as many as 68 variables, it is not possible to plot the search space for this study, but if we did, we would find that non-dominated solutions are quite far from each other in it. Pareto-optimal solutions are quite different from each other, and do not form a continuous pattern in the search space. From an architectural point of view, to gather this set of very different but high performing solutions is arguably quite useful, and can significantly increase performance in the early stages of design without hindering creativity.

12

Acoustic Design of Concert Halls

> "Charles Garnier, designer of the Opéra Garnier in Paris, said in his book "The Grand Opera in Paris", that he had pursued diligently the elusive factors of good acoustics, but he confessed that he finally trusted to luck, "like the acrobat who closes his eyes and clings to the ropes of an ascending baloon." "Eh bien!" he concludes, "Je suis arrivé" He went on, "The credit is not mine, I merely wear the marks of honor. It is not my fault that acoustics and I can never come to an understanding. I gave myself great pains to master this bizare science, but after fifteen years of labor, I found myself hardly in advance of where I stood the first day... I had read diligently in my books, and conferred industriously with philosophers - nowhere did i find a positive rule of action to guide me; in the contrary, nothing but contradictory statements. For long months, I studied, questioned everything, but after this travail finally i made this discovery. A room to have a good acoustics must be either long or broad, high or low, of wood or stone, round or square, and so forth... Chance seems as dominant in the theatrical [opera house] world as it is in the dream world in which a child enters Wonderland!""

(Beranek 2004)

Charles Garnier's sarcastic comments on the difficulties in the acoustic design of a concert spaces retain some validity even today. He started construction on the Paris Opera house a few years before Wallace Clement Sabine, the father of modern architectural acoustics, was born. Sabine would

later collaborate with McKim, Mead and White in the design of the Boston Symphony hall which is still considered one of the best for its acoustics. Since the pioneering work of Sabine, much more is known about the propagation of sound waves in confined spaces, the perception of sound by the human ear and brain, as well as the acoustical preferences of the audiences that attend these concerts. However, as acoustician Lawrence Kirkegaard puts it "What is yet to be learned could be more important than what we already know" (Beranek et al. 2010). Many uncertainties still make the design of concert auditoria a notoriously complex task even in contemporary architecture.

The purpose of acoustic design of concert spaces is to create the conditions for the enjoyment of music in the room, to enrich the experience of the musician and concert goer. This means generating the acoustics that the is user's preference for the music being played. Subjectivity is highly present in this field, and it is partly responsible for its complexity. Concert hall design is part architecture, part physics and part psychology. The user's opinion of a hall may be affected not only by the physical characteristics of the room, but also by many other unrelated issues such as cultural background, education and taste, and even by the reputation of the room.

Over the last 100 years, with the work of many experts that carried out interviews, questionnaires and laboratory experiments, the listener's and the musician's preferences in the acoustics of rooms destined for specific kinds of music have been studied, and some important characteristics have been laid out. The first and most important of which is reverberation, but it is by no means the only one:

> "In concert hall acoustics there are at least five independent dimensions. This was first established by Hawkes and Douglas (1971) and in the last three decades the nature of these different dimensions has been refined. The major concerns are that the *clarity* should be adequate to enable musical detail to be appreciated, that the *reverberant response* of the room should be suitable, that the sound should provide the listener with an *impression of space*, that the listener should sense the acoustic experience as *intimate* and that he/she should judge it as having adequate *loudness*. This list is by no means complete or definitive. It omits any reference to tone colour or timbre, which is certainly also important. Yet the five qualities: clarity, reverberance, spatial impression, intimacy and loudness provide a useful starting point for discussion."

(Barron 2009*a*)

In this passage, Barron describes five major acoustical attributes considered to be important to the listener's preferences in concert halls. Figure 12.1 shows the questionnaire he administered to study the acoustical quality of British concert halls. In it we can see the five attributes or "dimensions" and their characterization by the user. When it comes to clarity, rooms can be muddy or clear, they can be dead or alive when it comes to reverberation, the spatial envelopment can be expansive of constricted, the room can be remote or intimate, loud or quiet. These are subjective impressions by the listeners of the sounds they heard, the physical phenomenon occurring during the concert.

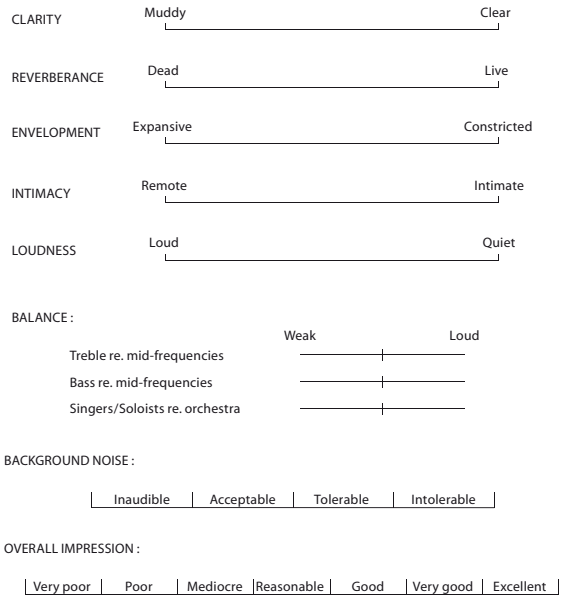


Figure 12.1: Questionnaire used by Barron for subjective survey of British concert halls.

A fundamental part of the puzzle of room acoustics is the connection between the physical characteristics of musical sounds and the listener's preferences, the relationship between subjective preferences and objective

measurements of sound. Physical studies and descriptions of sound fields inside concert spaces have been extensively made. Recordings of music, gunshots and balloon bursts have been used to measure and quantify different aspects of the physical phenomena at play in these spaces during the enjoyment of music. These recordings have been translated into countless measurements, visualizations and methods to try and make this connection between subjective and objective. An agreement among acousticians as to the best objective description of subjective preference is still not existent (Bradley 2011). However, most concert halls today are designed following one of these methods, most commonly the Room Acoustics Parameters described in the ISO 3382-1 standard (ISO 3382-1:2009 International Standards Organization, 2009).

The acoustic design of a concert hall can be defined as the definition of the shape, dimensions, materials and functional configuration of a space, with the purpose of creating the acoustic conditions over the entire audience area that best reflect the subjective preference of the listeners. The most important variables involved in this case are the shape and the materials of the surfaces that reflect the sound from the orchestra to the audience, the volume that reverberates the direct sound creating the hall experience.

Another important characteristic of the rooms is the absence of acoustical defects like echoes or background noise. While these defects have nothing to do with acoustical preference, they surely can lessen the concert experience. Nothing can ruin the concert experience more that the presence of external noise like car horns or ambulance sirens. This issue however is more easily solved, by means of insulating materials, sound barriers and generally avoiding vibration transmission. Noise control however has very little relation with room shape. Disturbing echoes on the other hand are the subject of room shape studies. Sound reflections, commonly from concave surfaces, can be heard as separate sounds (not as reverberation) if they are much louder than the rest of the reflections, and if they are too distant from the original direct sound. Therefore special attention is also given to the room shape in order to avoid sound concentrations and strong late reflections.

12.1 Concert Hall Types

Concert hall shapes have long been the subject of study, with particular attention being placed on existing room types that have good reputations for

their acoustics*. Concert halls and Opera Houses are an interesting case in the study of architectural typology. A very large number of concert spaces can be traced back to a small number of types, designers have not strained very far from established designs. Acoustics is considered by architects as a complex and somewhat obscure art, and until recently they followed established recipes, in the form of known types. Concert spaces are very expensive to build and even to repair, so the responsibility imposed on the designers is very high. All of these factors combined give us a scenario in which four simple concert hall types can describe almost all existing concert spaces. Existing concert halls can be classified into four main room types: Shoeboxes, fan shaped rooms, hexagonal rooms and vineyards (Meyer 2013). Equal volume, dimensions, materials and even reverberation could be achieved in all of these concert hall types, but they all have a particular sound.

Figure 12.2: The Musikverein in Vienna - photo credit: Andreas Praefcke.

The most important type of concert hall, the one with the highest number of examples is the "Shoebox" room. They are mostly characterized by their

*A discussion of concert hall types form a more theoretical point of view is presented in section 3.2. A historical review of concert hall types is presented in (Meyer 2013)

rectangular plan shape with the orchestra in one end of the room, their parallel and vertical sidewalls and their flat ceiling (thus the name shoebox). Some of the most renowned concert halls, such as the Boston Symphony hall, the Concertgebouw in Amsterdam and the Musikverein in Vienna (see figure 12.2), are shoebox rooms.

Most of these historical spaces are characterized by an important presence of ornamentation in the form of sculptures, bas-reliefs and frames. These tend to be volumetrically complex and irregular, thus guaranteeing a good level of sound diffusion or sound scattering from their surfaces. In other words, most sound waves are not reflected specularly from the surfaces, but are diffused in many directions. While this is not a characteristic that is present only in shoebox rooms, it is considered to be important in this type, in order to reduce the risk of flutter echoes caused by parallel walls.

Shoebox rooms can have a wide variation in acoustic reputation, from the most renowned to others with very bad reviews. This means that just the parallelepipedal shape is not enough to obtain a high acoustic quality. Studies have been made as to the correct proportions that the shoebox room should have in order to obtain acoustical quality[†]. Most of the best performing rooms in this type tend to be narrow and long rooms, as opposed to wide and short. This might sound counter-intuitive in the sense that long rooms would have a good number of listeners far away from the orchestra, while short rooms would have them closer. But apparently, listeners prefer to have a good number of side reflections coming form the sidewalls, so distance from the sidewalls becomes more important.

Another important room type is the Fan shaped room. As the name explains, the Fan shaped rooms can be best described by their fan or trapezoidal shape in plan, with their angled sidewalls opening up away from the stage. They can be seen as an indoor version of the greek and roman amphitheaters, they have the stage in the center of a series of concentric circles. They provide a visual intimacy in the room, since they minimize the distance between audiences and the stage. They are known to have little initial reflections in the middle of the room, due to their angled sidewalls. Many of them have a concave curved back wall that can cause sound concentrations on the stage[‡]. Examples of this type are the Fredric R. Mann auditorium

[†]See for example (Klosak & Gade 2008, Méndez Echenagucia, Astolfi, Shtrepi, van der Harten & Sassone 2013*a*)

[‡]See for example Beranek's account of the design of the Aula Magna for the Universidad Central de Venezuela, in which we recounts the solution proposed to these problems (Beranek 2004).

in Tel Aviv or the Aula Magna in Caracas.

A third room type outlined by Mayer is the hexagonal concert hall. The hexagonal rooms are characterized by their hexagonal plan shape, thus having angled sidewalls opening behind the orchestra and closing towards the room. The angled sidewalls offer the same advantages and disadvantages as the ones discussed in the fan shaped rooms. But the presence of the walls closing behind the audience provide a different experience, these walls provide a greater number of side reflections to seats positioned in the center of the room. Examples of thus type are Barbican Concert Hall in London or the Bunka Kaikan in Tokyo.

The Berlin Philharmonie is considered to be the first Vineyard room. This type is characterized by having the orchestra in the middle of the room, and by the presence of many audience terraces and inclined walls providing early reflections to seating positions directly in front of them. As it is explained in chapter 3, this type is very much appreciated by audiences and musicians for their intimacy, the vicinity of the audience to the stage, but also by its acoustical quality. The presence of a great number of terrace walls seems to substitute for large sidewalls that provide early sound reflections.

12.2 Room Acoustics Parameters

In the late 1890's, Sabine developed the concept of the Reverberation Time (RT) that became the basis for the study of room acoustics to come. Reverberation time was defined as the time it takes sound to decay by 60 dB after the sound source was switched off. Sabine noticed that this decay time was related to room volume and characteristics, and developed the following formula to calculate it:

$$RT = \frac{0.16 \cdot V}{A_{tot}} \quad \text{(s)} \tag{12.1}$$

where V is the room volume and A_{tot} is the total absorption of the room and can be calculated as:

$$A_{tot} = \sum_{i=1}^{N} S_i \cdot \alpha_i \tag{12.2}$$

where N is the number of surfaces in the room, S_i is the surface area of the i_{th} room surface and α_i is the absorption coefficient for the i_{th} sur-

face. Since materials have different absorption qualities at different sound frequencies, RT changes significantly for different frequencies in the same room. It is typically measured for several octave bands like 62, 125, 250, 500, 1000, 2000, 4000 and 8000 Hz. Reverberation times, as well as most other acoustical parameters are calculated from the measured impulse response. The impulse response is defined by the ISO 3382-1 standard as the "temporal evolution of the sound pressure observed at a point in a room as a result of the emission of a Dirac impulse at another point in the room" (ISO 3382-1:2009 International Standards Organization, 2009).

Since Sabine's development of the reverberation time, a great number of Room Acoustics Parameters, or objective measures, describing the sound field inside concert have been developed [§]. Most of these parameters describe aspects of the sound field at a single position inside the room based on the measured impulse response, and most of these try to describe one aspect of the subjective impression of the listener. Each parameter is associated with a Just Noticeable difference (JND) and a prescribed or optimal value, minimum-maximum acceptable values or a range of values preferred. Preferred parameters and their optimal values are usually selected on the basis of the purpose of the room, for example, rooms intended for opera, chamber or symphonic music are generally studied by practitioners with different parameters and using different optimal values for them. While there seems to be a general agreement among acousticians on JND values, there is no general consensus on which parameters best describe subjective preference of the listeners, nor the optimal values for them (Bradley 2011). For example, some contrasting optimal values are prescribed by Beranek (Beranek 2004) and Barron (Barron 2009*a*).

Bradley describes four main categories of acoustical parameters (Bradley 2011): Decay times, clarity measures, sound strength and measures of spatial effects. Other authors include other categories such as "Intimacy" or "Warmth", but these will not be included in this PhD research. In this chapter we will go through Bradley's categories and the objective parameters that describe them. They are all calculated by means of the measured impulse response.

[§]These parameters are the subject of the ISO 3382-1 standard (ISO 3382-1:2009 International Standards Organization, 2009). A historical description of the parameters is presented in (Lacatis et al. 2008). A comprehensive study of each parameter, their Just noticeable Differences and proposed optimal values or ranges is given in (Abdou & Guy 1996).

12.2.1 Decay Times

Decay times describe the way sound levels decay over time. The reverberation time RT, the first important acoustical measurement is of course a decay time, but it is not the only one. From RT on, other decay times were developed with different objectives, but they all describe reverberation in some form or another.

Studies show that the decay time parameter that best describes the subjective impression of reverberation while listening to music is the Early Decay Time EDT (Barron 1995). Mike Barron describes the development of EDT:

> "Atal, Schroeder and Sessler (Atal et al. 1965) conducted subjective tests in which subjects were asked to match artificially reverberated speech and music, with the comparison being made between decays which were linear (regarding sound level) and non-linear. For these artificially reverberated sounds the decay rate over the first 160 ms was found to relate most closely to perceived reverberation. When recordings were made in two concert halls, the subjective reverberation time matched most closely the initial reverberation time measured over the first 15 dB of the decay. Jordan (Jordan 1970) subsequently proposed in 1970 measuring the decay rate over the first 10 dB of the decay, naming it the early decay time (EDT)."

(Barron 1995)

Abdou and Guy describe EDT as the "slope of best fit straight line to sound level decay curve from 0 to -10 dB, extrapolated to -60 dB"(Abdou & Guy 1996). EDT simply extends the decay rate of the first 10dB to the full 60dB in order to be compared to the more historical and traditional RT. EDT values can be either longer or shorter that the RT value at that same point in the room. Barron describes this behavior and the room characteristics that define it (Barron 1995).

As all other acoustical parameters, it is very important to understand how sensible is the human hearing system to variations in EDT, how big a difference is perceived by the listener. This values are called Just Noticeable Differences (JNDs). The ISO 3382-1 standard reports a JND for EDT of 5%.

Beranek writes that the concert halls that obtained the best subjective ratings on his interviews had a mid frequency EDT value between 1.8 to

2.05 seconds for the occupied room measurements and between 2.45 and 3.1 seconds for the unoccupied room (Beranek 2004). These values are not that dissimilar from the values prescribed by Barron, from 1.8 to 2.2 seconds in the occupied rooms (Barron 2009*a*).

12.2.2 Clarity measures

Beranek defines clarity in the following way:

> "When a musician speaks of "definition" or "clarity", he means the degree to which the individual sounds in a musical performance stand apart from one another."

(Beranek 2004)

Clarity measures can then be said to be the objective measurements that try to describe the acoustical conditions in which individual sounds in a musical performance stand apart from one another. The way most of these measurements try to describe these acoustical conditions is by considering the ratio between early arriving energy and the reverberant later sound. There are many such parameters, Bradley includes four of them in his list of clarity measures: Definition D_{50}, Clarity C_{50} and C_{80}, and Centre time T_s. Out of these parameters, the one most used for the description of clarity for musical performances is C_{80}.

Also called the early to late ratio, C_{80} was developed by Reichard, Abdel and Alim in 1974. C_{80} is calculated on the basis of the measured impulse response, and it is defined by the following equation:

$$C_{80} = 10log\left(\frac{\int_0^{80} p^2(t)\mathrm{d}t}{\int_{80}^{\infty} p^2(t)\mathrm{d}t}\right) \quad \text{(dB)} \tag{12.3}$$

where $p(t)$ is the instantaneous sound pressure of the impulse response at a given measurement point. As the formula shows it is a logarithmic ratio of the energy measured before 80 milliseconds and the energy measured after 80 milliseconds. That means impulse responses with a high number of early reflections and low reverberation will give a high and positive C_{80}. Impulse responses with low early energy and high reverberation will have a low and negative C_{80}. Impulse responses that obtain the same amount of energy

before and after 80 milliseconds will have a C_{80} value of 0 dB. The ISO 3382-1 standard reports a JND for C_{80} of 1 dB.

In his study of 58 concert hall and opera houses, Beranek notes that rooms that obtained the best subjective ratings on his interviews had a mid frequency C_{80} value between 0 and -3 dB (Beranek 2004). These values are significantly different that those that Barron considers to be acceptable C_{80} values. Barron indicates values from -2 to 2 dB to be in an acceptable range (Barron 2009*a*).

12.2.3 Sound Strength

Sound strength is fairly self explanatory, these parameters look into the loudness or strength of the sound arriving at listeners form the source. The source's output has a big influence on the strength of the sound arriving at listeners, but a great deal of this also has to do with the room. Sound strength is considered by many acousticians to be one of the most important attributes to determine the acoustical quality of a room (Bradley 2011), with Beranek going as far as declaring that room volumes and materials should be calculated considering G as well as RT (Beranek 2011).

Sound strength is generally expressed by two objective measures, Sound Pressure Level SPL and sound strength parameter G. Of these two, the vast majority of acousticians use parameter G. It is defined in the ISO 3382-1 standard as "as the logarithmic ratio of the sound energy (squared and integrated sound pressure) of the measured impulse response to that of the response measured in a free field at a distance of 10 meters from the sound source", noting that the impulse response should be measured with an omnidirectional acoustic source. This definition can be expressed in the following equation:

$$G = 10log\left(\frac{\int_0^\infty p^2(t)\mathrm{d}t}{\int_0^\infty p_{10}^2(t)\mathrm{d}t}\right) \quad \text{(dB)} \tag{12.4}$$

where $p_{10}(t)$ is the instantaneous sound pressure of the impulse response measured at a distance of 10 meters in a free field. The standard also provides this alternate method for determining G when using an omnidirectional sound source of which the sound power level is known. In that case G can be obtained from the following equation:

$$G = L_p - L_W + 31\text{dB} \quad \text{(dB)} \tag{12.5}$$

where L_p is the sound pressure level measured at the desired point and L_W is the sound power level of the sound source used to do the measurement. G values can be positive, when the measured energy is greater that the free field energy at 10 meters, or negative, when the opposite is true.

Beranek explains that loudness in the room is mostly related to the total absorption A_{tot} and that approximately 50% of that absorption is determined by the area occupied by the audience (Beranek 2011). So the number of listeners in the audience, and the area they occupy in the room are a fundamental attribute to consider when designing concert spaces.

The JND for strength parameter G is defined in the standard as 1 dB. Optimal ranges of G can be varied when consulting different authors. Beranek considers that G values for mid-frequencies should range from 4 and 7.5 dB in large symphonic spaces, and should be even higher in spaces for chamber music. Barron on the other hand prescribes a minimum G value that is determined by the source-receiver distance (Barron 2009*b*). According to Barron, for receivers from 10 meters to the source onwards, a minimum value G_{min} is calculated by means of the following equation:

$$G_{min} = 10\log(100/r^2 + 2.08e^{-0.02r}) \quad \text{(dB)} \tag{12.6}$$

where r is the source-receiver distance. This formula prescribes a G_{min} value close to 4 dB for receivers 10 meters from the source, 2 dB at 20 meters, and 0 dB at 40 meters. Above 40 meters Barron keeps a minimum value of 0 dB.

12.2.4 Measures of Spatial Effects

The collaboration between Harold Marshall and Mike Barron in the 60's and early 70's gave start to the study of the spatial aspects of sound in concert spaces.

> "Before 1960 audible spatial effects were associated with the late reverberant sound (Kuttruff 2000); the experience of sound in a cathedral space clearly supports this connection. A long reverberation time and room surfaces that scatter sound were thought to enhance the spatial effect. Then in 1967 Marshall suggested that strong early reflections from the side were a component of sound in halls with the best acoustics. Whereas in the past there had been no guidelines available regarding the appropriate shape for symphony concert halls, here was a criterion with consequences for auditorium form. Marshall's ideas also

provided an explanation for the high reputation of traditional rectangular plan halls...

...Spatial impression was found to involve a sense of the source becoming broader for loud sounds, as well as a sense for the listener of being surrounded by sound, a sense of envelopment. The two components of spatial impression are called "source broadening" and "listener envelopment"."

(Barron 2009*a*)

The source broadening effect is mostly referred to as Apparent Source Width (ASW) and Listener Envelopment as (LEV). It is now known that ASW is related to early arriving lateral reflections, and that LEV is more related to late arriving lateral energy. ASW is most commonly studied by means of the Early Lateral Fraction parameter LF_{early} and Inter-aural cross correlation of the early-arriving sounds $IACC_{early}$, most commonly expressed as $1 - IACC_{early}$. LEV is most commonly measured by the late-arriving lateral sound strength (GLL)(Bradley 2011).

It is generally accepted that the early sound is most determined by the shape of the room while the reverberant sound is much less dependent on shape. Since this PhD thesis is most concerned with the early design phase and more specifically with architectural form, out of the two spatial effects we will consider only ASW. As it was previously mentioned ASW is studied by the use the means of LF_{early} and $IACC_{early}$.

Developed by Barron and Marshall (Barron & Marshall 1981) LF_{early} is the ratio between the early arriving lateral sound energy and the early arriving energy from all directions. It can be expressed as the following function:

$$LF_{early} = \frac{\int_5^{80} p_L^2(t)\mathrm{d}t}{\int_5^{80} p^2(t)\mathrm{d}t} \tag{12.7}$$

where $p_L(t)$ is the instantaneous sound pressure in the auditorium impulse response measured with a figure-of-eight pattern microphone. Looking at this equation, we can see that LF_{early} is a dimensionless quantity and that it can have values from 0 to 1.

$IACC$ measures the correlation between the impulses response measured inside the two ears of a dummy head. These correlations can take into

account different time intervals, in this case we are talking about $IACC_{early}$ so we consider an interval between 0 and 80 milliseconds, the same one used for LF_{early}. $IACC_{early}$ can be expressed by the following equations:

$$IACF_{early}(\tau) = \frac{\int_0^{80} p_l(t) \cdot p_r(t+\tau)\mathrm{d}t}{\sqrt{\int_0^{80} p_l^2(t)\mathrm{d}t \int_0^{80} p_r^2(t)\mathrm{d}t}} \tag{12.8}$$

$$IACC_{early} = max.of\ |IACF_{early}(\tau)| \quad \text{for} \quad 1 \ \text{ms} < \tau < 1 \ \text{ms}$$

where $p_l(t)$ and $p_r(t)$ are the left and right pressure impulse response measured inside or near the ears of a dummy head and τ is the time interval or time shift.

In his review of acoustic objective measures, Bradley explains what is known about the relationship between LF_{early} and $IACC_{early}$:

> "LF_{early} and $1 - IACC_{early}$ measures are conceptually quite different and it is not initially obvious that they are related to each other. However measurements of both quantities in 15 different halls (Bradley 1994) have shown that hall average values are significantly correlated in the octave bands from 125 to 1000 Hz inclusive, but not in the 2000 and 4000 Hz octave bands...
>
> ...The two types of measures do assess some similar aspects of the sound fields, but there are other aspects that do not create the same variations in these two types of quantities. One can speculate about the cause of the differences. LF_{early} values are derived from simple energy summations, but $1 - IACC_{early}$ values involve cross correlations of signals that could be influenced by interference effects that may not be reflected as changes in LF_{early} values. The important question is, are these audible differences and hence important to perceptions of concert hall sound quality? It seems possible that moving to an adjacent seat might produce measurable changes in $IACC_{early}$ values but not in LF_{early} values. Again, how do such changes relate to what we can hear? We need to understand the importance of the differences in these two types of quantities to know which best tells us about the subjectively important aspects of the spatial characteristics of halls."

(Bradley 2011)

Beranek found that when compared to LF_{early}, $IACC_{early}$ better predicted the subjective preferences in his interviews. But these calculations were made by using the average LF_{early} and $IACC_{early}$ values and not single listening positions. The choice between these two parameters is then not a clear one. For this PhD thesis LF_{early} was employed as a measure of ASW.

12.3 Total subjective preference and Room Acoustics Parameters

As we have seen in the previous sections, there are many subjective qualities to a concert space, and many more objective parameters that try to describe those qualities. But is there a way to describe the overall or total subjective preference of a room? Can we obtain a single number that describes the overall quality of a concert hall?

In two separate studies, Beranek and Ando try to answer this question by employing a weighted sum of a few parameters in order to get a single number. Beranek (Beranek 2004) attributes the sound quality of a room to the following parameters in these respective percentages:

$1 - IACC_{early}$	25%
EDT	25%
SDI	15%
G_{mid}	15%
Δt_1	10%
BR	10%

where SDI is the Surface Diffusivity Index, Δt_1 is the Initial Time Delay Gap, G_{mid} is the G value for mid frequencies and BR is the Bass Ratio. On the other hand Barron describes the relative importance of one subjective quantity over the other in the following way:

> "Several independent subjective quantities are important and people have their own personal bias in terms of what is for them most important. This implies that there is no single quantity that is most important. Rather, several measurable quantities are important and in a well-designed hall values for each quantity

> need to be within acceptable limits throughout the auditorium. The concert hall experience is definitely multi-dimensional."

(Barron 2009*a*)

In this PhD thesis, no single acoustical parameter or subjective quality is given preference over any other. In this research the four subjective families described above are treated as separate and contrasting functions, and the search for an optimal concert space is done by means of a multi-objective search algorithm. Therefore, the objective functions are evaluated with the Pareto approach.

13

Spatial distribution of Room Acoustics Parameters

The distribution acoustical quality inside concert spaces is not uniform (Akama et al. 2010, de Vries et al. 2001, Pelorson et al. 1992, Fujii et al. 2004). Source to receiver distances, as well as local conditions such as vicinity of sidewalls, balcony overhangs and balcony fronts make for substantial differences between listening positions. These differences can be measured by using the standard ISO 3382-1 room acoustic parameters. These measurements show that the differences between these listening positions can very well go beyond the JNDs (Barron 2005, Akama et al. 2010).

A clear objective of good acoustic design is to provide optimal acoustics to all listening positions inside a room. In this chapter we will look at the distribution of parameters in concert spaces and how room shape influences this distribution.

13.1 Past studies on distribution

Several studies have been made as to the variability in room acoustical parameters in different positions in concert spaces, in other words on their spatial distribution (Akama et al. 2010, de Vries et al. 2001, Pelorson et al. 1992, Fujii et al. 2004). In these studies distribution maps and histograms of different acoustical parameters are plotted in order to study their distribution. Although the object of the study is not a statistical fact, but a physical phenomenon at different points in the space, statistical methods such as histograms can be employed to best understand this distribution.

Akama et al. studied the spatial distribution for three different concert halls of varying shape, volume, and seating capacity. They measured the impulse response in almost every seating position in these rooms, and plotted both distribution maps and histograms of the distribution of several monaural parameters at mid frequencies. They calculated the width of the distribution for 90% of the values and found that this width is several times bigger that the JND, especially for EDT and C_{80} where it could reach more that 5 times the JND, less so for RT where it could reach less than 3 times the JND (Akama et al. 2010).

Akama et al. also performed normality tests on the histograms for RT, EDT and C_{80} values in these three rooms. Normality tests are meant to study the distribution in a histogram and determine wether or not the distribution can be considered normal. They found that many of the distributions, especially those involving C_{80} at mid frequencies, could not be said to have normal distributions.

Fujii et al. measured the impulse responses in many listening positions in two Japanese concert halls with similar shape but with different volume, seating capacity and surface types in their sidewalls (one more irregular than the other) (Fujii et al. 2004). They plotted distribution maps for the Sound Pressure Level (SPL), subsequent reverberation time (T_{sub}), Initial Time Delay Gap (Δt_1) and $IACC$. They confirmed that distributions of these parameters are wider their respective JNDs. They also concluded that the scattered reflections of the sidewalls of one of the rooms had an influence on the distribution, particularly by decreasing SPL and increasing Δt_1 values near the walls.

13.2 Measurements of Distribution

The ISO 3382-1 standard (ISO 3382-1:2009 International Standards Organization, 2009) parameters all measure values for single points in the room. So in order to quantify their distribution inside the room we need to resort to statistical or mathematical methods. This section discusses several methods for the study and quantification of this distribution. Additionally we look into the distribution of optimal values inside the room, rooms should not only be acoustically as uniform as possible, they should uniformly distribute optimal acoustics.

13.2.1 Average values

It is very common for acousticians to display room measurements in terms of whole room averages. They convey a general description of the room in a single, easy to understand value. However, mean values can be misleading because they can be the result of very different values. This might not be the case for reverberation times, but many other acoustical parameters such as Clarity and Sound Strength can vary greatly inside a single room (Barron 2005, 2013). This also means that two very different halls can share the same mean values. Barron makes the example of two british halls, one with favorable subjective impressions, and the other very much unliked by audiences, and points out that this could not be predicted by their average objective values. This means that their average values are quite similar despite their very different subjective impression(Barron 2005).

13.2.2 Standard deviation

The ISO 3382-1 standard (ISO 3382-1:2009 International Standards Organization, 2009) cites the Standard deviation (σ) as a way to describe spatial variance. The standard deviation is good way to present measurements in a room. For example, the mean value of C_{80} for all listening positions, accompanied by the standard deviation of those values from the mean, paints us a much clearer picture of how the clarity parameter is distributed inside this room.

It is worth noting that standard deviations have very little meaning in non-normal distributions. When we have asymmetrical distributions in terms of their mean value, the standard deviation can be misleading, and should not be taken into consideration. As we have seen above, not all distributions of acoustical parameters inside concert spaces are normal distributions. Hence the standard deviation is not a perfect indicator of distribution.

When we study how *optimal* values of acoustic parameters are spatially distributed inside the room (uniformly or not), the standard deviation is not a viable option. If we consider two rooms with the same average C_{80} but with different standard deviations, we can be sure that the room with the lowest standard deviation is the more uniform room. But if we want to figure out which room has the largest distribution of optimal C_{80} values, and the optimal C_{80} is not equal to the two rooms average C_{80} values, then the standard deviation is not going to help us. It could be the case that the room with the highest standard deviation has more listening positions with

optimal C_{80}.

Comparisons between rooms are quite useful in the design of a new room, especially if one considers the use of optimization or automated search algorithms. Uniformity for uniformity's sake is not enough to design a room. We need the evenly distribute an optimal value (typically a value related to subjective preference). The use of optimal values to determine a room's form presumes an agreement among the designers about these optimal values. Currently acousticians do not agree upon definite optimal values for most of the performance parameters contained in ISO 3382-1, and more importantly many lack criteria for preferred values (Bradley 2011). There is also disagreement on the correct parameters to use. However, since all current parameters describe the acoustical quality in a single point in the room, they all have a need for uniformity measures. They all also need optimal values relating to subjective preferences, so in order to design a room, an agreement among the people involved has to be reached.

13.2.3 Percentage of satisfied receivers

Barron (Barron 2005) proposed the use of the percentage of "satisfied" receivers as another possible number to describe the acoustic quality of a whole room. Receivers will be considered satisfied in a different way for each performance parameter in question. Generally speaking a receiver will be considered satisfied when it obtains a parameter value within a pre-selected optimal range. For example, if we select an optimal range of EDT from 1.8 to 2.2 seconds, then we will consider satisfied all of the receiver positions in which we can measure an EDT within that range. Some acoustic parameters are frequently studied in relation to the source-receiver distance. In this case the determination of a satisfied receiver can be made by means of a function that considers the source-receiver distance. For example, Barron proposes a minimum acceptable sound strength G value that is in relation to the source-receiver distance (Barron 2005). That value is determined by equation 12.6. So in this case we would consider satisfied receivers who obtain a G value equal of higher than G_{min}. In order for the percentage of satisfied receivers to work properly, a high number of measurements are required. This costly and labour intensive, but this can be expected on any measure of spatial distribution.

The percentage of satisfied receivers seems to be a good way of comparing rooms in terms of the spatial distribution of sound quality, the room with the highest percentage should be considered the best room. Since it is a percentage, it also has the advantage of communicating quite well the

degree in which the room is uniform or dis-uniform. The percentage of satisfied receivers however does not consider the degree in which the dissatisfied receivers are dissatisfied. Since this percentage looks only into the number of receivers that lie inside an optimal range, it does not measure the difference between the measured data and the optimal values, we do not get an idea of how far the unsatisfied are from being satisfied.

Figures 13.1 and 13.2shows an analysis of the unsatisfied receivers for a series of different shoebox rooms (A,B,C and D) with various room length to width ratios and heights. The analysis consists in the incremental widening of the optimal range from which the percentage of satisfied receivers is calculated. In the x axis we see how this range increases in size, and in the y axis we see the percentage of satisfied receivers corresponding to each range. By plotting the percentage as it increases with the wider ranges we get an clear picture of how close or how far receivers were from the original range. In other words, the faster the rate of increment of the percentage, the less unsatisfied the unsatisfied receivers are.

The first issue we can discuss about this graphs is the fact that the rate of increment of each room. We can see that the rates are always not constant, we see the percentage increase in curves with varying slope angles. If we look at the curves for EDT we can see that rooms C and D have higher initial numbers of satisfied receivers, but A and B have a very high rate when compared to C and D. So we can say that C and D have unsatisfied receivers that are more unsatisfied that those from A and B.

If we look at the case of C_{80} this behavior is perhaps even more evident. Rooms B and C have a higher initial percentage than room A, but room A has a much higher rate of increment, to the point that room A surpasses rooms B and C after a few range increments and reaches 100% first than them. We can generalize the results of this study by saying that a higher number of satisfied receivers does not guaranty that the unsatisfied receivers will be less unsatisfied. On the contrary, it could be argued that, in some cases, a high percentage of satisfied receivers is achieved at the expense of the unsatisfied ones. So if we have to compare rooms A and B for example, it's not so easy to say which is best, one that has a good number of satisfied receivers and the rest are completely unsatisfied (room B), or one that has a lower number of satisfied but the unsatisfied are not that unsatisfied (room A).

The unsatisfied analysis reveals that there is no straight forward relationship between the percentage of satisfied and the status of the unsatisfied. Hence, there is a need for a more precise tool in the study of the distribution of optimal values inside rooms. One that includes the unsatisfied receivers

in the equation.

13.2.4 Histograms Study

We have seen above that past studies on the distribution of acoustic parameters inside rooms employ the use of statistical tools such as histograms in their studies. We will begin this section by plotting the histograms for our test rooms A, B, C and D.

Figures figs. 13.3 to 13.7 show histograms of the distribution of RT, EDT, C_{80}, G and LF_{early} values at mid frequencies in rooms A, B, C and D. In the x axis they show the respective parameter values discretized in small sections and in the y axis they show the occurrence or frequency in which these values fall into. The charts also show the width of 90% of the distribution. We can also see in grey the optimal values for each parameter suggested for the symphony concert hall. The optimal values for each parameter and the source of the value (name and publication of author who suggested it) are presented in the following table:

Parameter	**Optimal Range**	**Reference**
RT	1.8 to 2.2 (s)	(Barron 2009*a*)
EDT	1.8 to 2.2 (s)	(Barron 2009*a*)
C_{80}	-2 to 2 (dB)	(Barron 2009*a*)
G	Barron's min G curve to ∞ (dB)	(Barron 2009*b*)
LF_{early}	0.1 to 0.35	(Barron 2009*a*)

RT histograms are shown in figure 13.3, they show that the distributions are usually very uniform, it is widely known that RT is mostly uniform throughout the room. With the exception of a few outliers in room B, most of the RT values lie within 2-3 JNDs from each other. In addition, the distributions seem to be symmetrical. Other parameters have less uniform and symmetrical distributions.

Figure 13.4 shows the EDT distributions, they show much wider distributions in terms of their JNDs, form 8 to 15 JNDs in width. We can also note that distributions are not very symmetrical.

Figure 13.5 shows the C_{80} distributions. Looking at these histograms we can explain the results of the unsatisfied analysis shown in figures 13.1 and 13.2. We can see why rooms B and C have a higher initial percentage of satisfied receivers. Many of their listeners lie inside the optimal range, but they are distributed in a wide area (12.5 and 9.4 JNDs respectively).

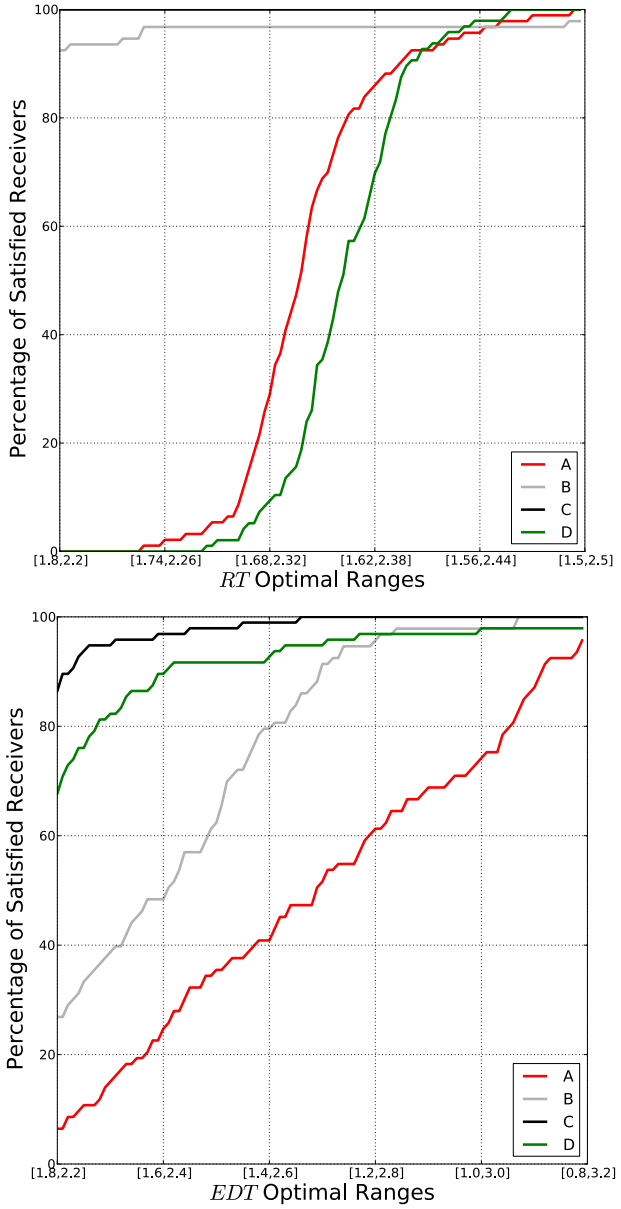


Figure 13.1: Unsatisfied analysis for RT, EDT, C_{80} and LF_{early} for rooms A and B.

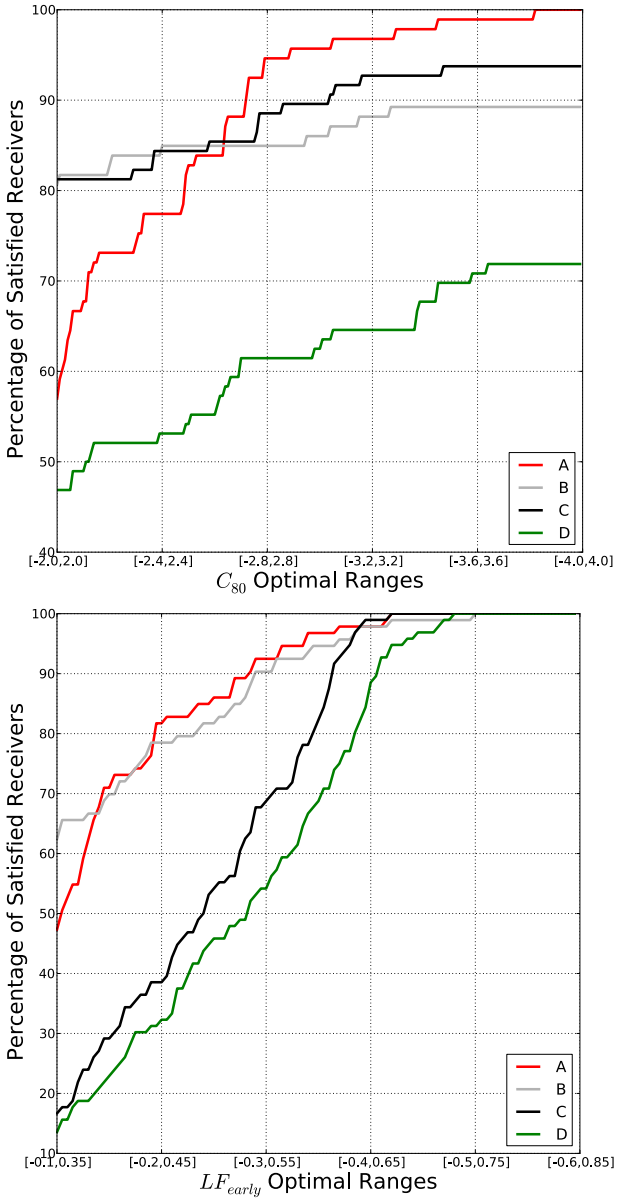


Figure 13.2: Unsatisfied analysis for RT, EDT, C_{80} and LF_{early} for rooms C and D.

Room A has a narrower distribution of 8.5 JNDs but it is not centered in the optimal range, however it is very close to the optimal values, hence its unsatisfied receivers are not as unsatisfied as those in rooms B and C. We can also see that these distributions are quite asymmetrical and far from a normal distributions.

13.2.5 Difference weighted sum

Based on Ando's theory of subjective preference (Ando 1983) Beranek devised an "objective method" that rated rooms according to the mean values of orthogonal objective attributes (Beranek 2004), in Beranek's case they were $IACC$, Δt_1, G_{mid}, EDT, BR and SDI. As it was outlined above, his rating is based on a weighted sum of 6 parameters. The rating method can be outlined with the following set of equations:

$$S = S_1 + S_2 + S_3 + S_4 + S_5 + S_6 \tag{13.1}$$

where S is the total subjective preference of an acoustical environment and $S_{i_{th}}$ is determined by:

$$S_i = a_i |x_i|^{3/2} \tag{13.2}$$

where:

$$\begin{array}{l} x_1 = 1 - IACC_{early} \\ x_2 = log(\Delta t_1 / \Delta t_{1,pref}) \\ x_3 = G_{mid} - G_{mid} \qquad \text{(dB)} \\ x_4 = log(EDT / EDT_{pref}) \\ x_5 = log(BR / BR_{pref}) \\ x_6 = log(SDI / SDI_{pref}) \end{array} \tag{13.3}$$

where $IACC$, Δt_1, G_{mid}, EDT, BR and SDI are the average values of those parameters, $IACC_{pref}$, $\Delta t_{1,pref}$, $G_{mid,pref}$, EDT_{pref}, BR_{pref} and SDI_{pref} are the optimal or preferred ranges of values of the same parameters, and where:

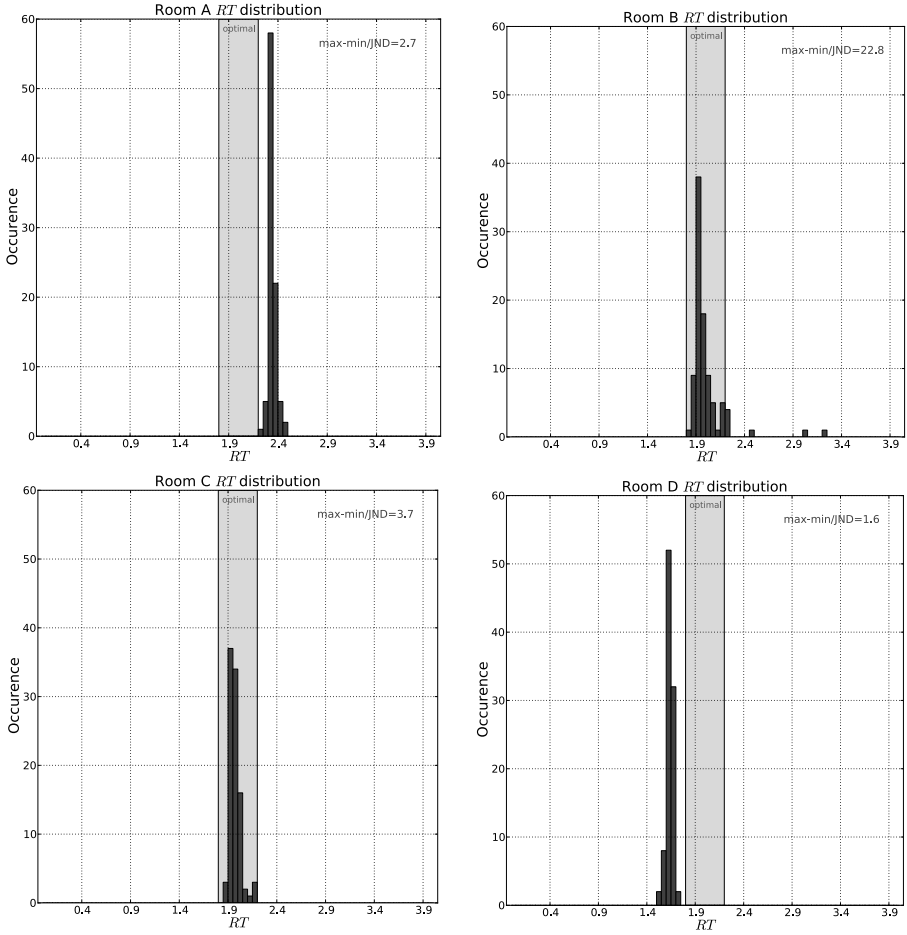


Figure 13.3: Histograms of the distribution of RT for rooms A, B, C and D.

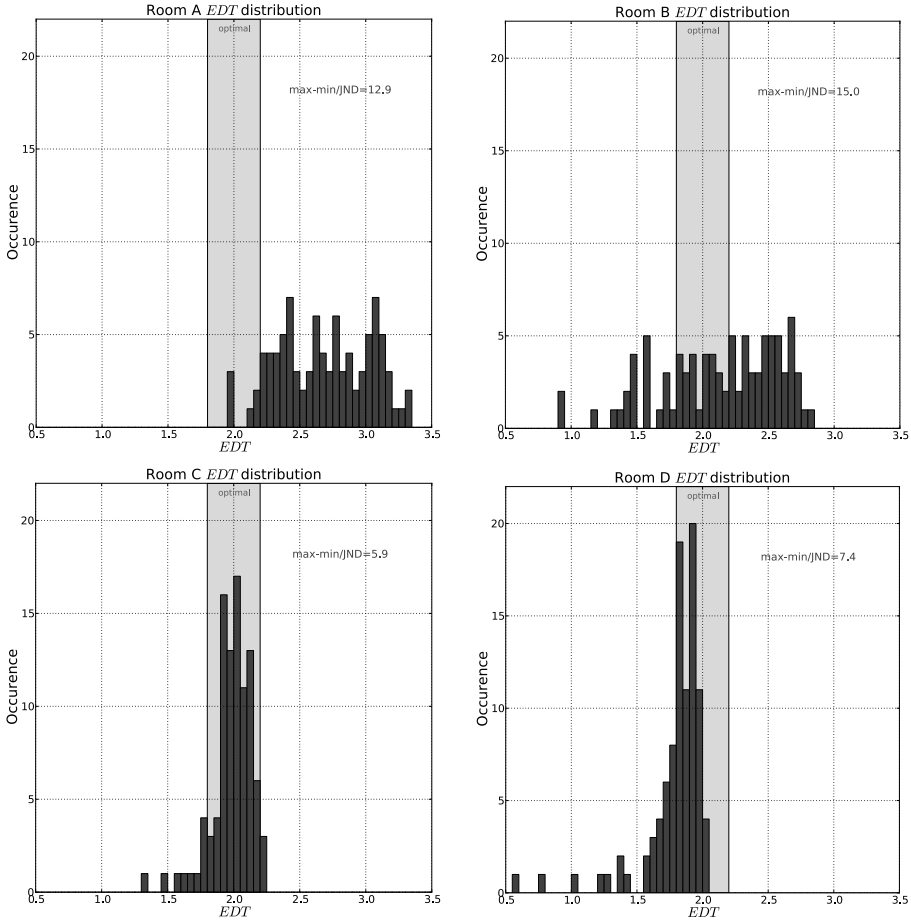


Figure 13.4: Histograms of the distribution of EDT for rooms A, B, C and D.

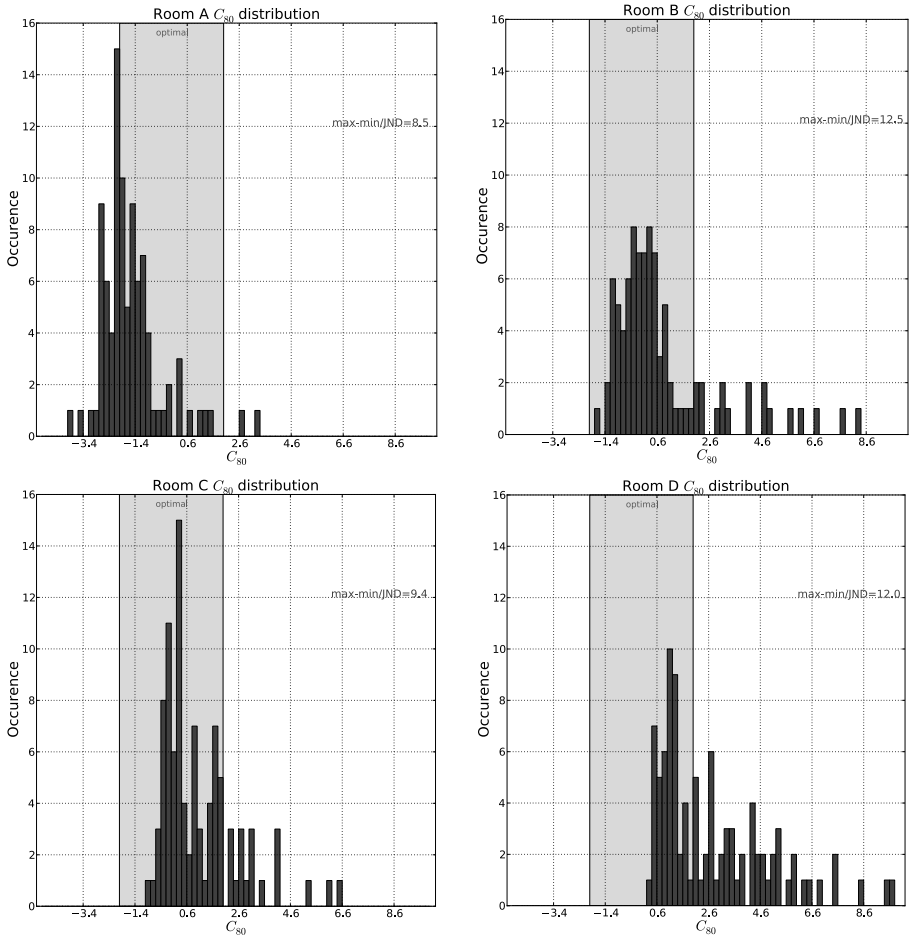


Figure 13.5: Histograms of the distribution of C_{80} for rooms A, B, C and D.

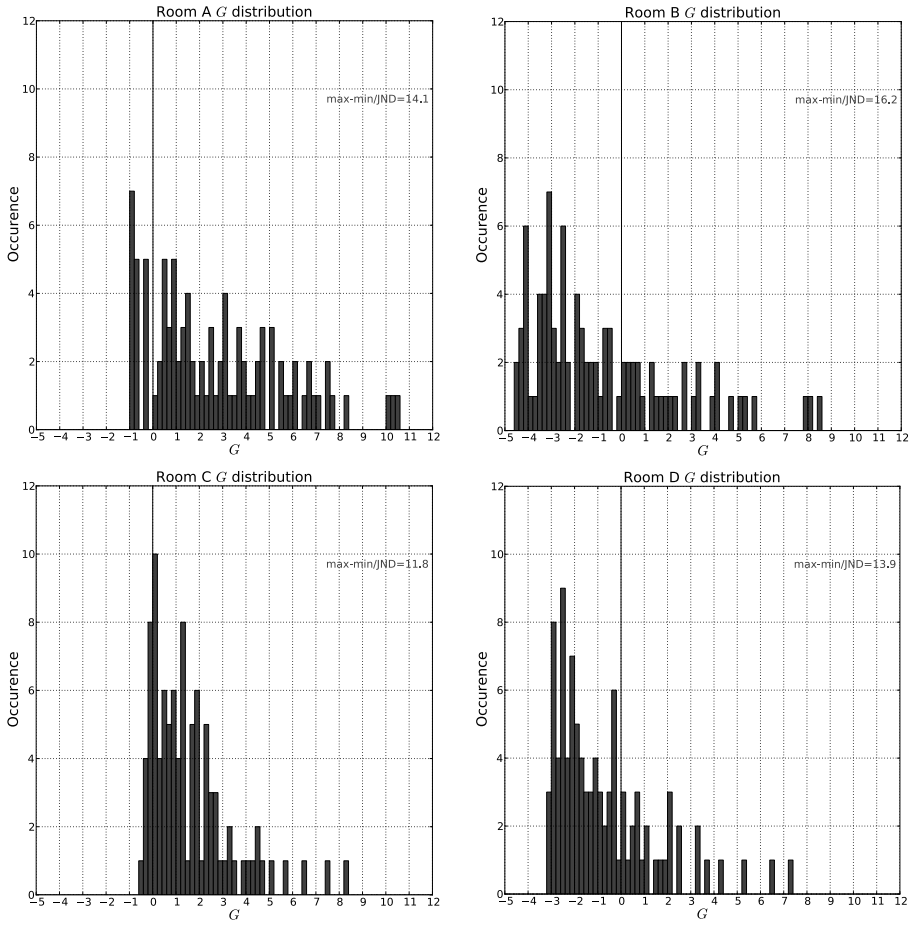


Figure 13.6: Histograms of the distribution of G for rooms A, B, C and D.

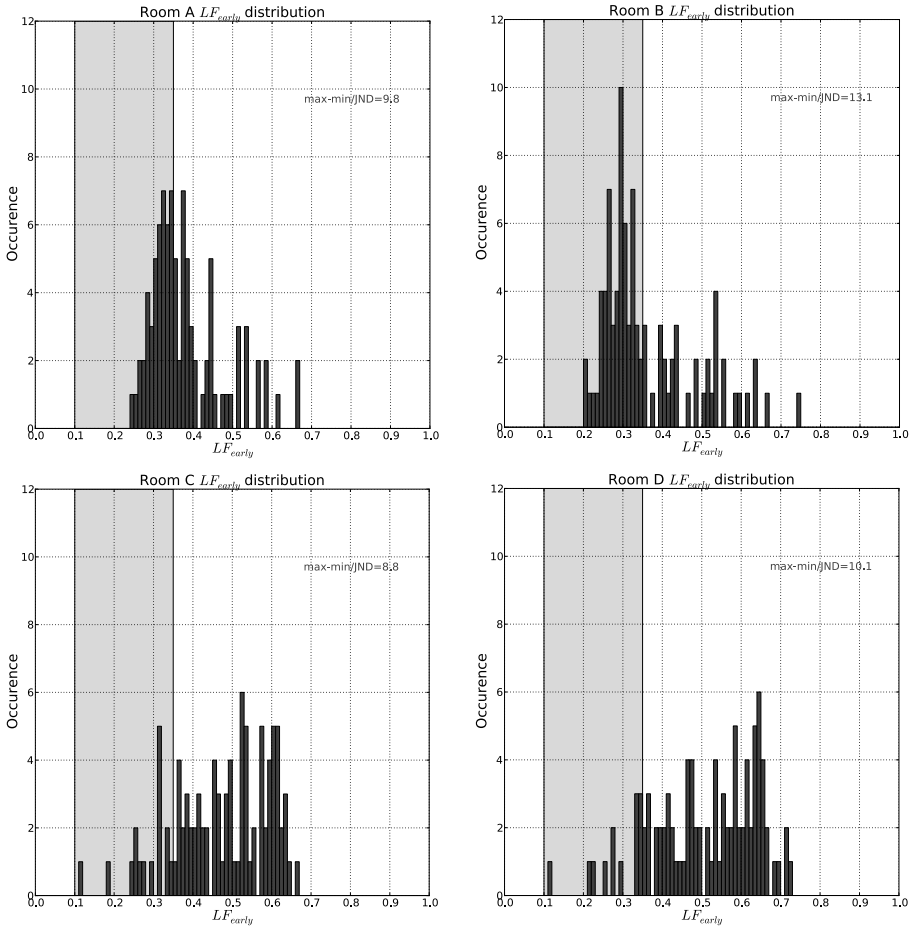


Figure 13.7: Histograms of the distribution of LF_{early} for rooms A, B, C and D.

$$
\begin{aligned}
a_1 &= 1.2 \\
a_2 &= 1.42 \\
a_3 &= 0.04 \quad \text{for} \quad G_{mid} < 4.0 \\
&= 0.07 \quad \text{for} \quad G_{mid} > 5.5 \\
a_4 &= 9 \quad \text{for} \quad EDT < 2.0 \\
&= 12 \quad \text{for} \quad EDT > 2.3 \\
a_5 &= 10 \quad \text{for2.2} \quad \text{sec} \\
a_6 &= 1
\end{aligned}
\tag{13.4}
$$

An important detail to take note of in this formulation is that Beranek is using the average value of each parameter and uses that value to calculate the total subjective preference S. Beranek publishes a series of graphs that show the weighed value of S_i for possible average EDT values. Figure 13.8 shows the graph for S_4, the graph for EDT.

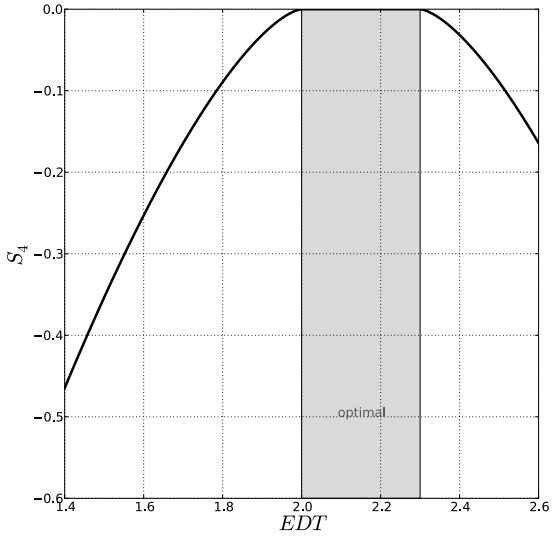


Figure 13.8: Beranek's weight factor Parabola for EDT.

With this formulation, Beranek considers all values inside the preferred range to have equal subjective significance, while values outside of the pre-

ferred range obtain exponentially lower S_i scores. This exponential decay in subjective preference is also present in Ando's 1983 study. He explains this decay and his subsequent use in subjective preference calculations by means of figure 13.9 (a and b) and the following caption:

> "Scale values of subjective preference obtained by the paired-comparison test for simulated sound fields in an anechoic chamber. Different symbols indicate scale values obtained from different source signals (Ando 1983). Even if different signals are used, a consistency of scale values as a function of each factor is observed, fitting a single curve".

(Ando 2007)

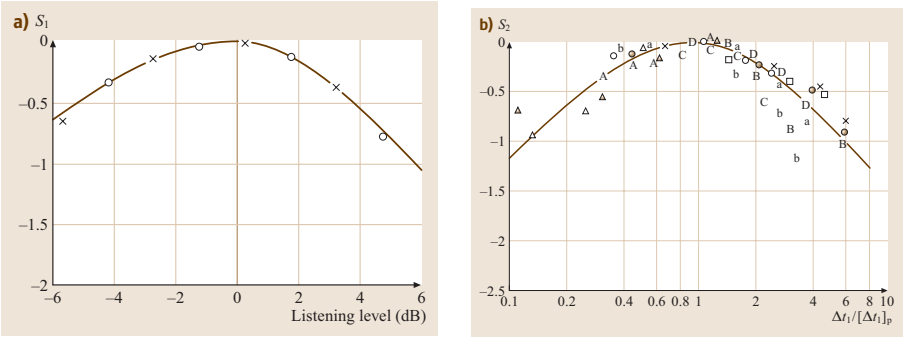


Figure 13.9: Ando's weight factor Parabola for Listening Level LL (a)and Δt_1 (b).

In other words, Ando found that subjective preference decays in an exponential way as parameter values move away from preferred parameter values. This way of quantifying decay gives us an opportunity to measure more accurately the level of satisfaction of any listening position, based on the distance between the measured value and the preferred value and Ando's curve.

In this PhD thesis an equation to study the spatial distribution of sound quality in concert halls based on the difference of measured and optimal values is proposed. Using Ando's theory as employed by Beranek, we sum the weighted differences of optimal and parameter values, for all receivers in

the room. More accurately, we calculate the integral of a function $f_{(i)}$ that is determined by the rooms distribution multiplied by a weighting factor determined by Ando's curve. Figure 13.10 shows the distribution function $f_{(i)}$ for the EDT parameter in rooms A, B, C, and D.

The total distribution score D_i for each parameter i will thus be determined by the following equation:

$$D_i = \int_{-\infty}^{+\infty} f_{(i)} \cdot S_i \, \mathrm{d}x \tag{13.5}$$

where S_i is the subjective weighting curve determined by:

$$S_i = |x_i|^{3/2} \tag{13.6}$$

and where the x_i values are:

$$\begin{aligned} x_1 &= log(EDT/EDT_{pref}) \\ x_2 &= C_{80} - C_{80_{pref}} \\ x_3 &= G - G_{pref} \\ x_4 &= log(LF_{early}/LF_{early_{pref}}) \end{aligned} \tag{13.7}$$

where EDT, C_{80}, G and LF_{early} are the measured parameters values and EDT_{pref}, $C_{80_{pref}}$, and $LF_{early_{pref}}$ are the mid points of the preferred parameter ranges. As already mentioned the G parameter is considered in respect to the source-receiver distance d, so G_{pref} is calculated using Barron's formula (see equation 12.6). The values for S_i can also be expressed by the curves in figure 13.11.

The preceding formulation can be simplified in order to avoid the approximation of the histogram into $f_{(i)}$. The actual values measured in the room can be used in the following set of equations:

$$D_i = \frac{1}{n} \cdot \sum_{i=1}^{n} S_{i,n} = \frac{1}{n} \cdot \sum_{i=1}^{n} (|x_{i,n}|)^{3/2} \tag{13.8}$$

where n is the number of receivers measured in the concert hall. The room that obtains the highest D_i value is considered the room with the most uniform distribution of optimal parameter i values.

If we use equation 13.8 to study C_{80} for rooms A, B, C and D we get different results than those obtained by looking at the percentage of satisfied receivers. Figure 13.12 shows normalized values for both measurement types for C_{80} in our test rooms. The comparison shows that the relationship between these rooms has changed substantially. Room A for example is

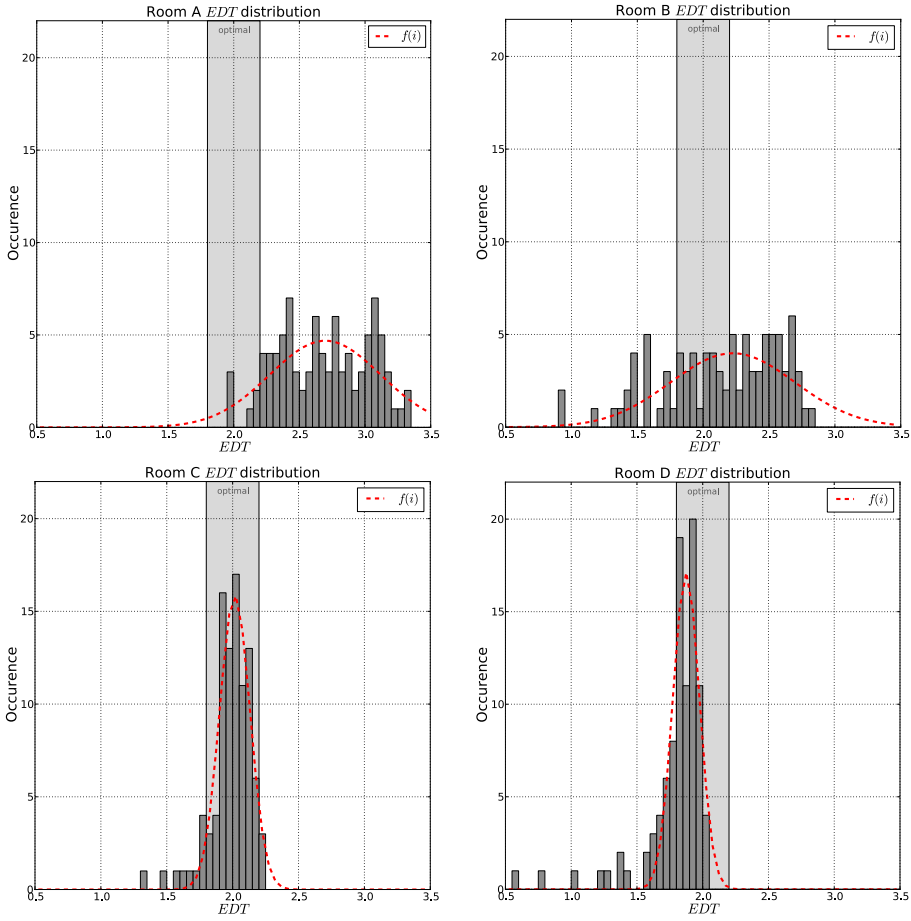


Figure 13.10: Distribution functions $f_{(i)}$ for the EDT parameter of rooms A, B, C and D.

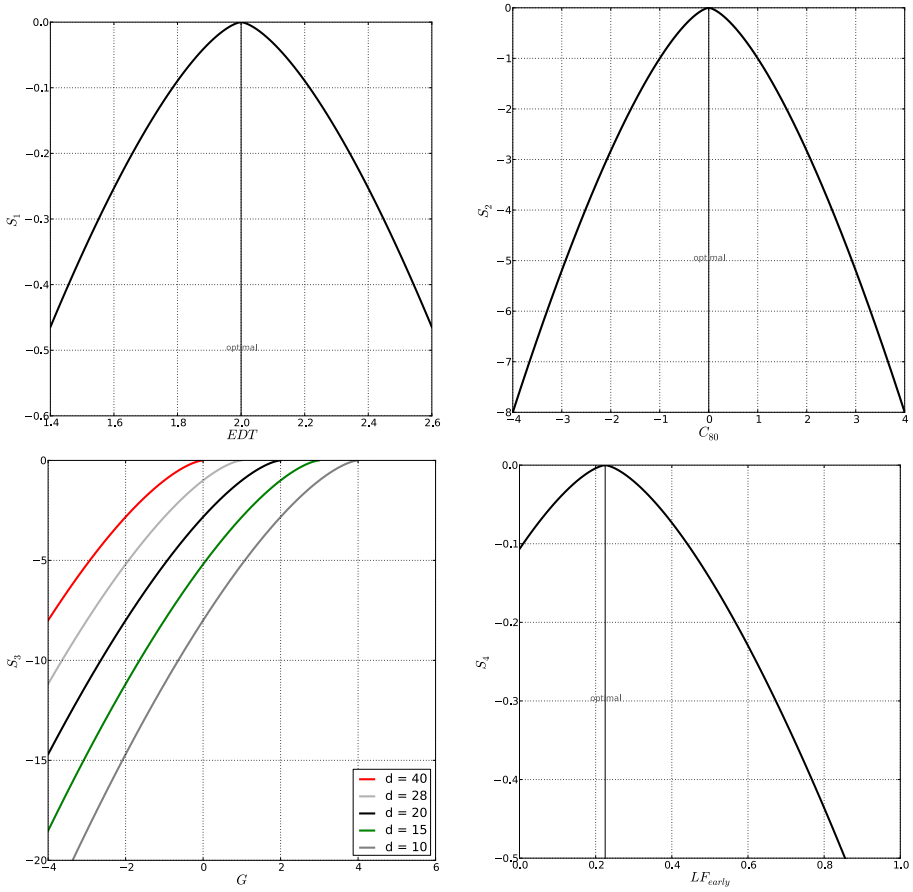


Figure 13.11: Proposed weight factor curve for EDT, C_{80}, G and LF_{early}.

much closer to rooms B and C, gaining a lot of distribution fitness. Room B on the other hand looses distribution fitness in the new measurement. This new measurement seems to better reflect what wee see in the unsatisfied analysis (figures 13.1 and 13.2) and the histograms for C_{80} (figure 13.5).

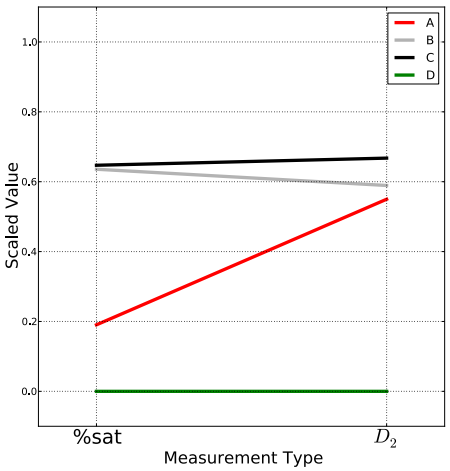


Figure 13.12: Percentage of satisfied receivers for C_{80} and D_2 comparison.

The formulation proposed in this PhD research differs from Beranek's and Ando's in that it takes into account room distribution (all of the measured values in the room) and not just the room average values.

13.2.6 Discussion

When it comes to expressing room parameters in a single value, or just a few numbers we can employ different techniques. While very common, the mean value is appropriate only for Reverberation Times, and not for other parameters that have a greater variation inside the room. The mean value accompanied by the standard deviation is a much better choice, but its difficult to make comparisons based on those numbers if the means are different. It is also not statistically correct to use standard deviations in non-normal distributions.

Comparisons can be made by using optimal values as references. Barron's percentage of satisfied receivers while having advantages lacks information about the unsatisfied receives. We can say that using the percentage of satisfied receivers has the added value of communicating to the designers the extent to which the receivers are inside the optimal range quite clearly. On the other hand the standard deviation is not as clear as to how much improvement it would still need to obtain an acceptable result.

The percentage of satisfied receivers however, fails to inform about the conditions of receivers outside the optimal range. As figures 13.1 and 13.2 suggests, rooms can have high numbers of satisfied receivers at the expense of unsatisfied ones, distributing sound quality unevenly. A new approach is proposed, making use of the difference between measured values from all positions in the room, and weights them according to Ando's theory of subjective preference(Ando 1983).

13.3 Parametric study of concert Hall Types

Klosak and Gade studied the effect of shoebox room proportions (width/length) have in acoustical parameters by means of computer simulations of 24 different rooms (Klosak & Gade 2008). They plotted mean values of several acoustic parameters accompanied by their standard deviations and the 90th percentile. They found that the proportions of the room had a significant influence on the distribution of acoustic parameter values.

Bradley(Bradley 2011) describes four main categories of acoustical parameters: Decay times, clarity measures, sound strength and measures of spatial effects. In this section we will investigate the distribution inside concert halls by means of a parametric study using one parameter from each of these four categories. A large number of computer simulations of concert halls were made for each room typology, and each one of the simulation results was analyzed with the difference weighted sum method described above.

13.3.1 Selection of Types

The most built room type across the last century has been the shoebox, and in recent years the vineyard room has been increasing in popularity(Meyer 2013). This PhD thesis presents a parametric study of the distribution of acoustical parameters in shoebox, fan shaped and hexagonal rooms.

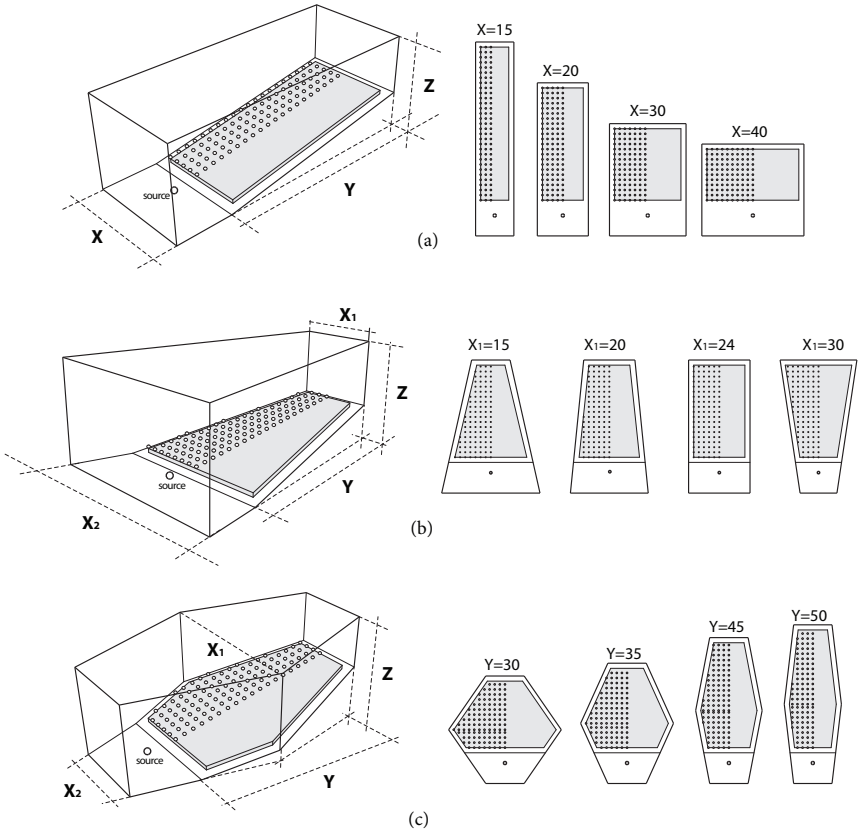


Figure 13.13: Shoebox, Fan and Hexagon geometry parametrization.

Each typology is studied in terms of their form, volume and proportions. The variations in shape are different for each typology and will be detailed bellow, but all the rooms studied have the same audience area, so a comparison of results between room types is also possible. These room types were selected because of their use in concert halls throughout the world, and because their geometry makes it possible to make such a parametric study.

Many variations of each room type will be computationally generated and studied for later comparison. Only the general geometry of the room will be studied, no balconies canopies or surfaces will be included in the simulations. All of the room types will include a 10% inclination in the audience area.

While vineyard rooms are also an important type in today's concert halls, their geometry is far more complicated to study. There are too many variations of vineyards to include in a single study, their geometry is hard to describe by using few variables. For this reason they were left out of this study.

13.3.2 Methodology

Geometry Parametrization

Figure 13.13(a) shows the parametrization of the shoebox room geometry. audience area remains fixed as geometry in plan is changed, Y decreases as X increases. The two geometric parameters that determine the shape of each shoebox room are the height of the room (Z) and the width/length ratio (X/Y).

Figure 13.13(b) shows the parametrization of the fan shaped room geometry. Also in this case audience area remains fixed, so as X_1 increases X_2 decreases. The room length (Y) remains unchanged. The two geometric parameters that determine the shape of each fan room are the height of the room (Z) and the X_1/X_2 ratio.

Figure 13.13(c) shows the parametrization of the hexagonal room geometry. In order to maintain audience area fixed in this type of room the room's length (Y) decreases as the room's width at the center of the room (X_1) increases. The room's width at the ends of the room (X_2) remains fixed. The two geometric parameters that determine the shape of each hexagonal room are the height of the room (Z) and the width (at mid room)/length ratio (X_1/Y).

Type	V_1	Dom	V_2	Dom	Step Size
Shoebox	X	15 to 35m	Z	15 to 25m	1m
Fan	X_1	13 to 35m	Z	15 to 25m	1m
Hexagonal	Y	20 to 50m	Z	15 to 25m	1m

Table 13.1: Parameter domain and discretization for each room type.

Fitness landscapes

The parametric study was done by calculating the acoustic parameters for a series different rooms of varying shape, and comparing results. Since each concert hall typology was parametrized using 2 variables (V_1 and V_2), the different room shapes studied can be expressed as a combination of V_1 and V_2. If we create a grid made up of V_1 in its X axis and V_2 in its Y axis, we can plot the performance of these rooms in the Z axis and create what is known as a fitness landscape. Table 13.1 shows the variable domains for V_1 and V_2 for all room types as well as the size of the variation for each variable at each step.

Acoustical simulation

The acoustical simulations were done with Pachyderm Acoustical Simulation (van der Harten 2011) plugin for Rhinoceros. Pachyderm Acoustical Simulation is a collection of acoustical simulation algorithms for use in Rhinoceros, ranging in purpose from prediction to auralization. Among its features are a growing number of simulation algorithms that can be performed using mesh or NURBS models. Pachyderm combines the source image method with a the ray-tracing technique.

Room materials study

The prime goal of this chapter is to understand the influence of room shape in the distribution of acoustical parameters inside performance spaces. But room shape is not the only factor that affects this issue and it is not easy to study room shape independently. We can understand this difficulty by looking into Sabine's formula *. It teaches us that reverberation times depend on room volume and total absorption area, which in turn depends on the

*see formula 12.1 on page 192

room's surfaces area and the materials of these surfaces. A room's shape affects its volume and the area of its surfaces, so in order to study room shape we would have to keep the materials absorption coefficient fixed. However, keeping absorption fixed would not be sufficient to maintain a fixed total absorption, in order to do so material surfaces would also have to be fixed. This means that, if we want to study the effects of only the room's shape, we would have to use geometries that had the same room surfaces as well as keep materials fixed.

Klosak and Gade studied the effect of shoebox room proportions (width/length) have in acoustical parameters by means of computer simulations of 24 different rooms (Klosak & Gade 2008). They plotted mean values of several acoustic parameters accompanied by their standard deviations and the 90th percentile. They used 3 groups of shoebox rooms that kept room surfaces, materials and volume fixed, consequently also keeping the reverberation time fixed. The only modification between rooms inside each group was the width to length ratio. They reported noticeable changes in acoustical parameters and their distribution inside these rooms.

The study presented in this chapter involves room changes that go beyond the width/length ratio, they also include room volume and total surface area. But in order to do this we need to clarify the influence of the material in sound quality distribution. Figures 13.14 and 13.15 shows two sets of histograms for the distribution of C_{80} in rooms A, B, C and D. The first set (seen in the left column) shows the histograms of the rooms with a fixed absorption coefficient for all of the surfaces in the room excepting the audience. The absorption coefficient for those surfaces in this set was fixed at 10%. The second set of histograms (seen in the right column) show the histograms for the same rooms but with a variable material absorption coefficient. The materials are set in such a way as to keep reverberation time fixed with the changing of the room's surfaces and volume. In other words, the absorption coefficient for these rooms is selected by using the following equation:

$$a_{room,f} = \frac{(\frac{0.161 \cdot V}{RT_f}) - (S_{audience} \cdot a_{audience,f})}{S_{room}} \tag{13.9}$$

where $a_{room,f}$ is the absorption coefficient for frequency f of all the surfaces in the room except the audience area, V is the room volume, RT_f is the desired reverberation time for frequency f, $S_{audience}$ is the surface area of the audience, S_{room} is the sum of all areas of all the surfaces of the room and $a_{audience,f}$ is the absorption coefficient of the audience for

frequency f. For the 500Hz and 1000Hz octave bands $RT_f = 2$ seconds [†].

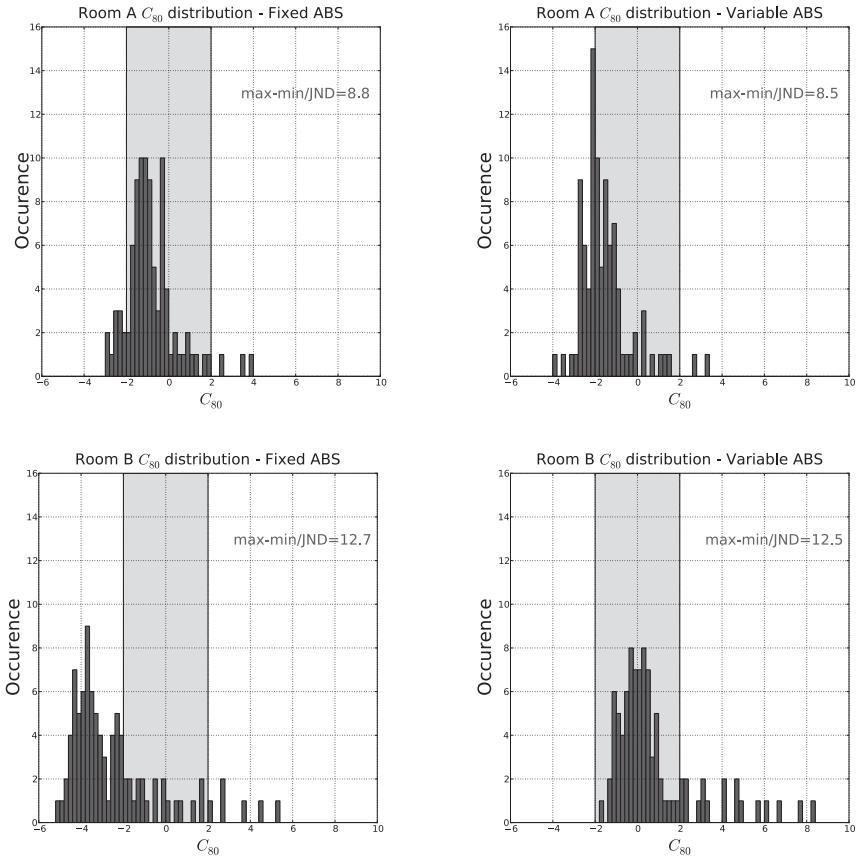


Figure 13.14: Histograms for rooms A and B for C_{80} distribution with fixed and variable material absorption coefficient.

These rooms all belong to the same shoebox parametric model explained above, hence they all share the same rules. Most importantly, in all of these rooms the audience area is kept constant. For these reasons we can say that Rooms A, B, C and D serve as a guide of what is happening with all of the rooms in terms of the material's influence.

[†]This value was taken from (Barron 2009*a*).

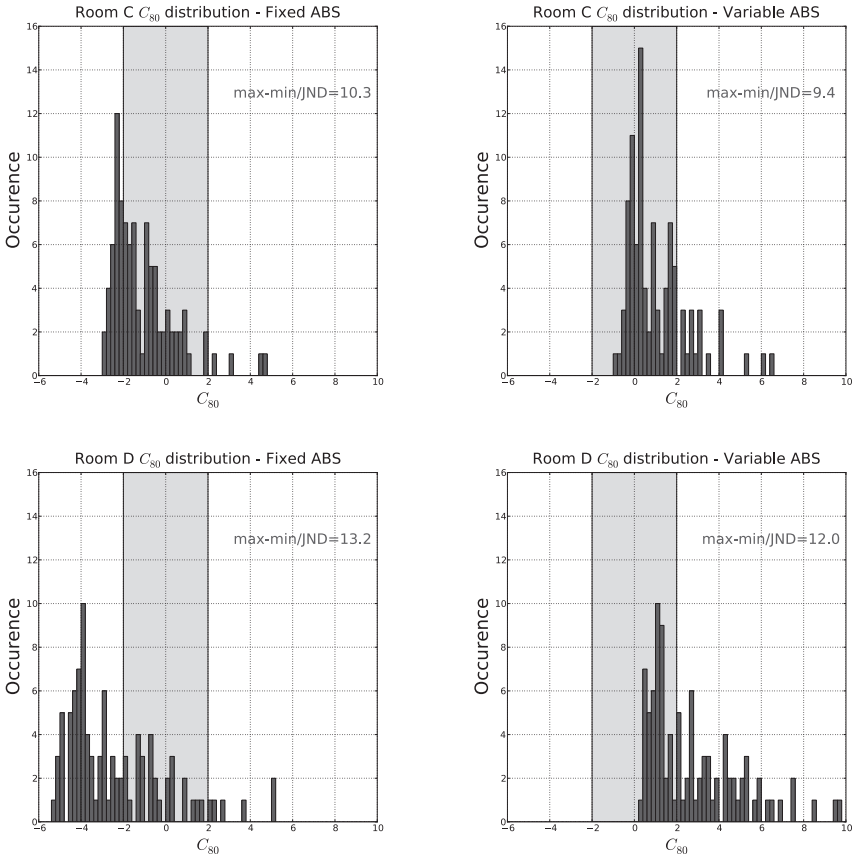


Figure 13.15: Histograms for rooms C and D for C_{80} distribution with fixed and variable material absorption coefficient.

Figures 13.14 and 13.15 is organized in such a way as to have room A with fixed absorption on the left and room A with variable absorption on the right of the same row. This is the same for rooms B, C and D, thus facilitating comparisons. By comparing the histograms on the left with those on the right we can see that the distributions for these two rooms are similar, they have a very similar 90% width to JND ratio. The comparison reveals one important difference, room distributions are traslated from the other as is evident by comparing how they relate to our optimal range. With the exception of room A, all histograms on the right have similar distributions to those on the left but with higher C_{80} values. Room A seems to show no movement and very little variation in width to JND ratio.

The fact that the distribution is very similar tells us that the material has little influence in the distribution of acoustic parameters, but only on their position from optimal values. This is also true for EDT, G, LF_{early} and $T30$ with different degrees. In some cases the influence of material is more present in distribution, but generally speaking we can say that material does not have the largest importance.

This fact should not be completely surprising, given what we know about RT and room shape. Barron (Barron 2009*a*) explains that the most important factors in the determination of RT in a room are room volume and audience area. As we have discussed above, our parametric model keeps audience area constant throughout the variations, including those for rooms A, B, C and D. That leaves only room volume as the determining factor for RT, and leaves out material absorption.

Material selection for parametric study

Having studied the influence of materials in the distribution, and seen how the most important factor is not material but room shape, we can decide wether to use a fixed material absorption or to change material to keep reverberation time fixed. Since material absorption affects not distribution but values in relation to the optimal range, we have to conclude that in order to find the shape that best obtains the best distribution of *optimal* values we need to study the combination of material and shape. For this reason, keeping the material fixed would strongly limit the scope of the study.

In this chapter we chose to change the absorption coefficient of the room's surfaces (with the exception of the audience) using the approach described in equation 13.9. Because this approach keeps reverberation times fixed, we can exclude RT as a parameter to study and concentrate on EDT,C_{80}, G and $LFearly$.

The diffusivity of the room's surfaces also influences the distribution of sound quality. However in this case, since it has little influence on the reverberation time, we decided to keep a scattering coefficient fixed at 50% for the room surfaces, except for the audience. The audience was given a 70% scattering value. These numbers have been taken from (Lam 1996).

13.4 Case Study 6: Parametric Shoebox, Fan and Hexagon

Case study 6 is the typological parametric study developed with the method described above. The subject of this case study are the 3 room typologies presented above. In this case, the search process is not done via a MOGA but with an exhaustive search algorithm and the fitness landscape technique. This is due to the fact that this problem is subdivided into room types, and each room type has only 2 variables. Hence an exhaustive search within normal variable domains is not too time consuming, and it allows us to better understand and diagram the problem.

The parametric study results are studied in two ways:

- Comparisons within room types. This is a comparison of rooms in the same type, we study the incidence of the room variations in shape has on the distribution of quality for all four of the acoustic objective parameters.

- Comparisons between types. This is a comparison of the distribution of acoustic parameters between the 3 different room types. The purpose is to examine the particular qualities that make each type better or worse in each acoustic parameter distribution.

Each comparison has its own representation method.

13.4.1 Comparison within room types: Fitness landscapes

The concept of the fitness landscape was explained in section 10.1.3. In this comparison we use a fitness landscape for each acoustical parameter and each room type.

Shoebox Rooms

Figure 13.16 shows the fitness landscapes for EDT, C_{80}, G and LF in the shoebox room type. As shown in table 13.1, the shoebox rooms have two variables: x_1 that represents the room width, and x_2 that represents the room height. All landscapes thus have the x_1 variable in the x axis and x_2 in the y axis. The z axis of the landscapes vary, each one indicates a different acoustic parameter. The first one (upper left) shows D_1, the second (upper right) shows D_2, the third (lower left) shows D_3 and the fourth shows D_3 , all calculated with equation 13.8.

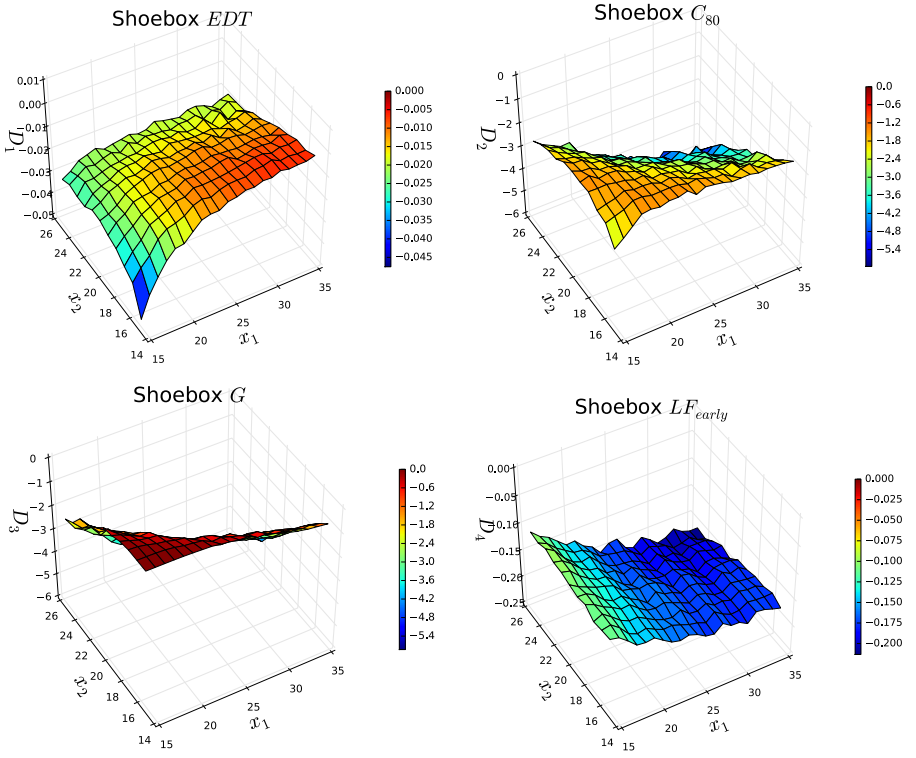


Figure 13.16: Fitness landscapes for the shoebox room type for EDT, C_{80}, G and LF_{early}.

Shoebox rooms show sensitivity to both room width/length ratio and

height when it comes to EDT, C_{80} and G, while LF_{early} seems to be only affected by the width/length ratio. This is a reasonable result since LF_{early} is the only parameter that is completely independent of reverberation.

C_{80} and G have the highest values when rooms are long, narrow and not very high, while EDT shows a preference towards short,wide and also not very high rooms. The overall trend shown in the landscapes for C_{80} and G seems to be very similar, but it can be argued that for different reasons. In the case of G the preference can be explained by the room volume, G values tend to be lower as volume increases. And since this parameter has no maximum optimal value, the higher the G the better the D_3. In the case of C_{80} this can be explained by the presence of higher early reflections when the rooms are narrow. This is perhaps most important in our case study because there are no other surfaces in the room that provide early reflections, apart from the sidewalls and the ceiling, there are no canopies or balcony overhangs. In the absence of early reflections C_{80} values tend to be low, and EDT values tend to be high (Barron 1995). That also explains why EDT preference goes towards wider rooms. Taller rooms have too much reverberation, and therefore, too long $EDTs$ for our optimal range (1.8-2.2s).

Early Lateral Fractions are at their highest D_4 values when the rooms are narrow, regardless of room height.

Fan shaped rooms

Figure 13.17 shows the fitness landscapes for EDT,C_{80}, G and LF in the fan shaped room type. In the case of the fan shape room, table 13.1 shows that the variables for the fan are: the room's width at the end of the room, and the room's height. All landscapes thus have the x_1 variable in the x axis and x_2 in the y axis. In the same way as the shoebox example, the z axis represents distribution values for our four acoustical parameters calculated with equation 13.8.

It is important to notice the way these room's geometry changes in plan when x_1 increases. When $x_1, 24$m is low, the room is an inverted fan, when $x_1 = 24$m it is a shoebox, and when $x_1 > 24$m it is a fan shaped room. This is shown in small plan views in figure 13.13(b).

The most important variation this progression in the angle of the sidewalls from the point of view of the sound source. When the room has a regular fan shape the sidewalls open up from the stage towards the room, when it is an inverted fan the walls are closed towards the room. An open angle towards the room produces a higher number of early reflections arriving

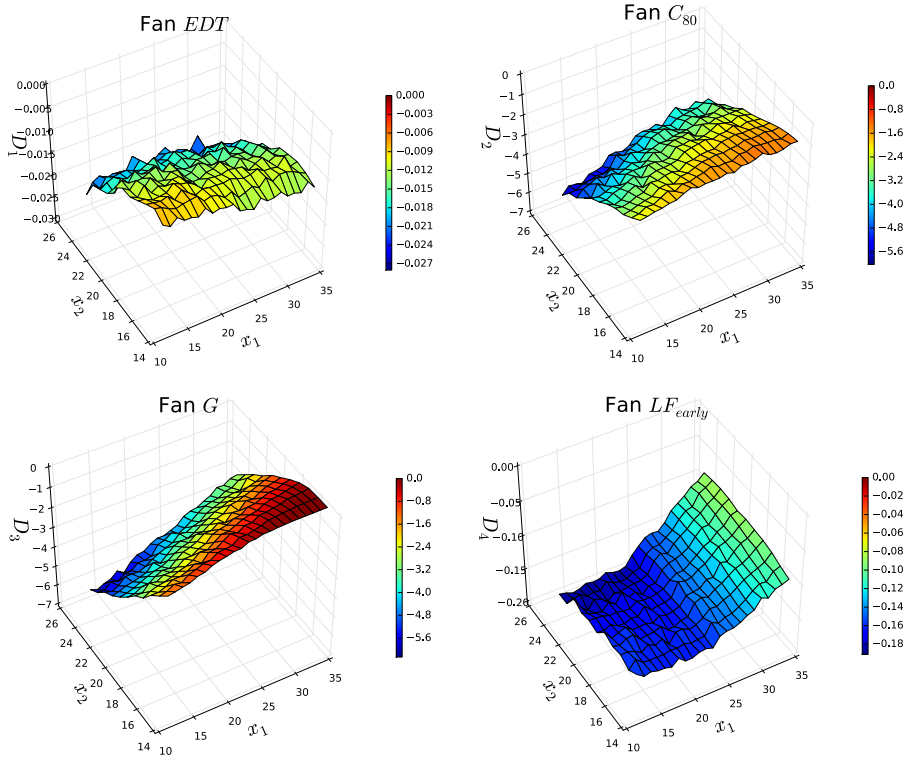


Figure 13.17: Fitness Landscapes for the Fan shaped Room Type for EDT, C_{80}, G and LF_{early}.

the audience. While a closed angle is synonymous with higher reverberation. This behavior is shown in figure 1.3 on page 23. It shows the spatial distribution of the G parameter in 4 variations of hexagonal room with equal room height. High G values are represented in blue and lower ones in red. We can see that an open angle around the stage increases G values in the last rows of the room. We can also see that parallel sidewalls around the stage have the opposite effect, regardless of the angle of the sidewalls around the audience.

This behavior explains the results obtained in this parametric study of the fan shaped rooms. Room height x_2 also has a big importance in the results of some parameters. D_2 and D_3 values are at their highest when an open angle is present around the stage and the room is low, showing little contrast between C_{80} and G (as seen in the shoebox example). EDT seems to have a lower sensitivity to the room shape in plan than it has on room height, however, we find the highest D_1 values when the inverse fan is present and the room is low. This is consistent with Barron's explanation on $EDTs$ (Barron 1995) and the fact that inverse fans have lower early reflections.

Room height has a low but present incidence on LF_{early} values, and the highest D_4 values are found when room is a tall and open fan.

Hexagonal Rooms

Figure 13.18 shows the fitness landscapes for EDT, C_{80}, G and LF in the hexagonal room type. The two variables for this room type (seen in table 13.1) are the room length x_1 and room height x_2. These are the two variables plotted in the x axis and y axis of the landscapes in figure 13.18. As usual the z axis describes D_i.

Also in this case the variation in plan is important to understand the results. As seen in figure 13.13(c), as room length increases, room width at the middle of the room decreases to maintain audience a fixed area. Room withs at both ends of the room are always fixed. This means that sidewall angles in plan vary significantly as x_1 increases. In this case there is never a closed angle from the stage towards the audience, as was the case in the fan shaped rooms. The angle becomes closer to parallel sidewalls as the room length increases, but never getting to the point of becoming parallel.

Results in this study seem to be related in good measure to that movement in plan, showing a higher number of early reflections and a higher G when the room becomes narrower and sidewall angles become more closed, but never flat. Room height also has an important role in the results of all

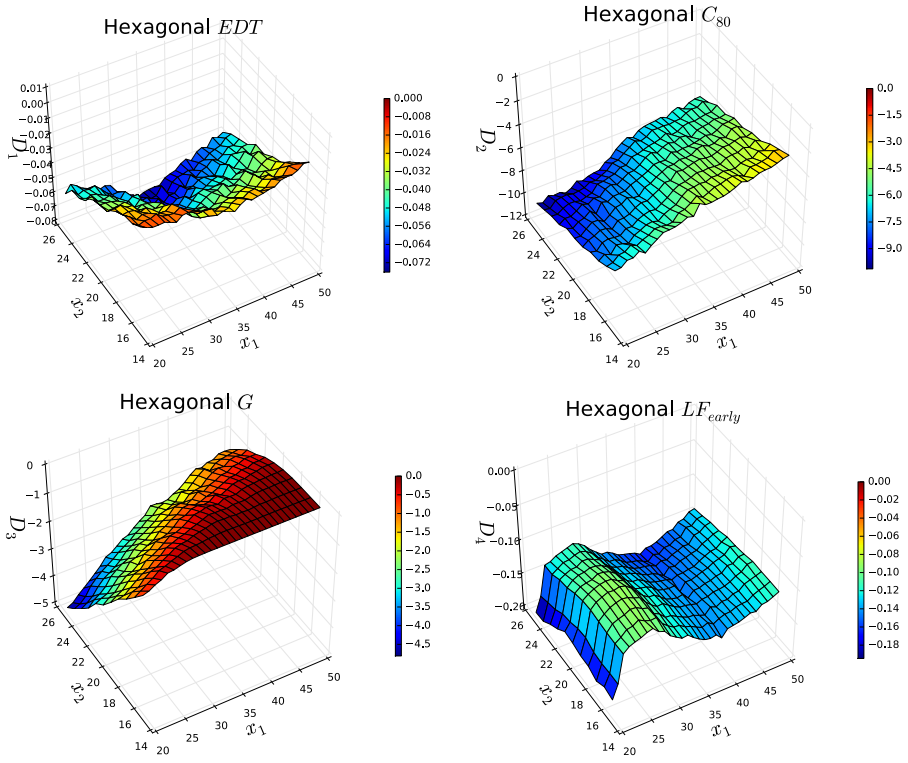


Figure 13.18: Fitness Landscapes for the Hexagonal Room Type for EDT, C_{80}, G and LF_{early}.

parameters with the exception of D_4. As seen in the previous types, D_2 and D_3 values seem to be very much related.

D_1 results show an interesting trend in plan. D_1 values tend to be high either when x_1 values are low or high, but not in mid values. D_1 values are lowest when the rooms are high, but they are also low when x_1 are close o 35 meters. Shoebox rooms have high D_1 values when they are very wide, this can also explain why they are high when rooms are wide in the hexagonal rooms. High D_1 values when x_1 values are high are harder to explain. Results suggests that rooms in x_1 mid range that are not ideal, they seem to have too little lateral reflections but not a high reverberation, while narrow hexagonal rooms regain lateral reflections and reverberation seems to be near optimal.

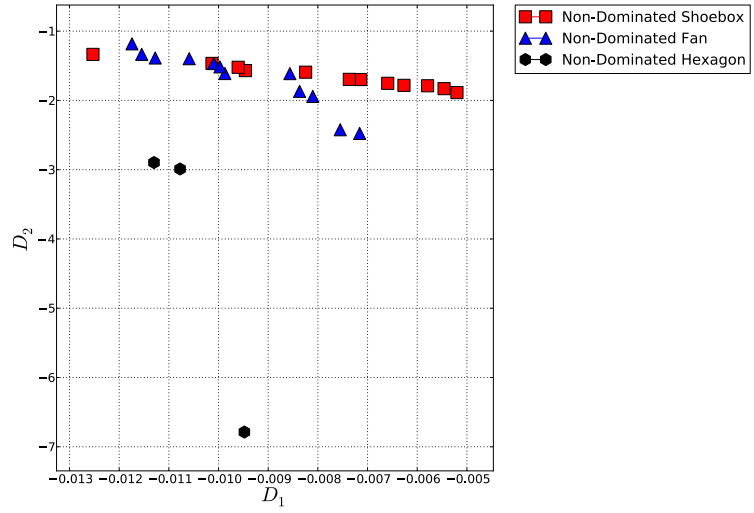
Also in this case, D_4 values are almost unrelated to room height, they vary only with x_1 values. As is the case with D_1, D_4 numbers are highest in two different sets of rooms. When x_1 is close to 25 and 50 meters, D_4 numbers are high. The D_4 landscape suggests that a good lateral energy is achieved either in very narrow rooms (as seen in the shoeboxes) or in rooms that have an optimal opened angle around the stage (as seen in the fan shaped rooms). Interestingly, when the room is too wide, the angle around the stage is of no help to improve lateral energy and D_4. This could also be related to the angle around the audience, but with a lesser degree of importance.

13.4.2 Comparison between types: Pareto fronts

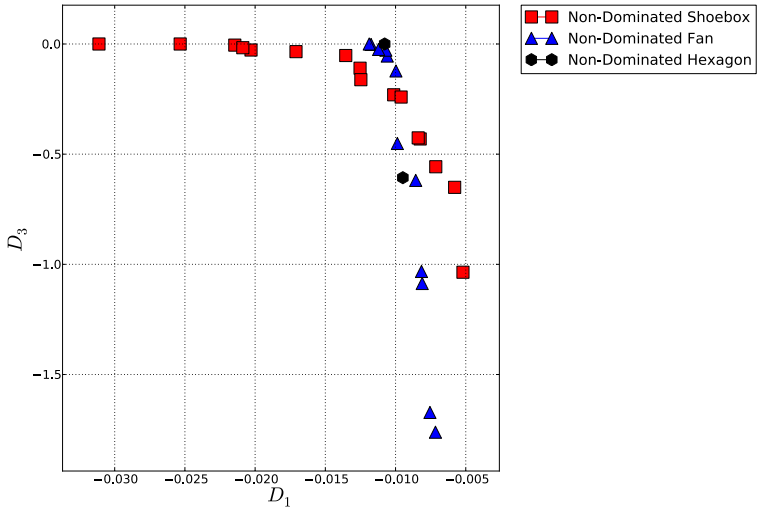
A more direct comparison between the different types can be seen by studying their Pareto fronts. Since all room types were studied with the same objective functions (D_1,D_2,D_3 and D_4), we can plot an objective space containing all of the room types, and in the objective space, single out the non-dominated individuals. Since there are 4 objective functions, it is not possible to have a single graph representing all functions, objective spaces will be shown in pairs of functions.

Reverberation vs. Clarity

Figure 13.19a shows the Pareto fronts of all 3 room types for the EDT and C_{80} parameters. We can see that shoeboxes have the highest D_1 values and both shoeboxes and fans have high D_2 values. Hexagonal rooms have lower values for both D_1 and D_2.

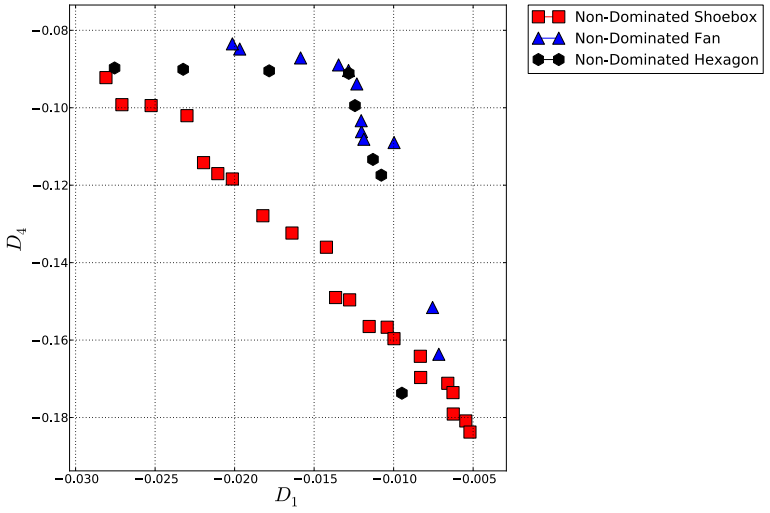


(a) EDT vs. C_{80}

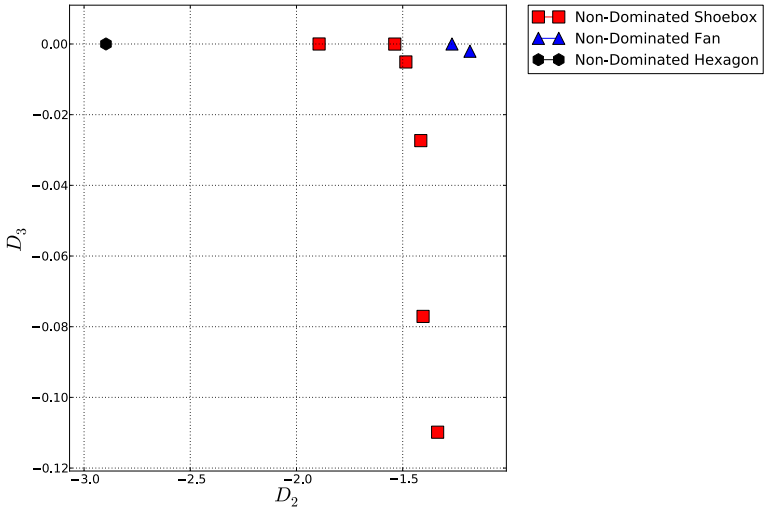


(b) EDT vs. G

Figure 13.19: Pareto Fronts comparisons of the shoebox, fan shaped and hexagonal rooms.

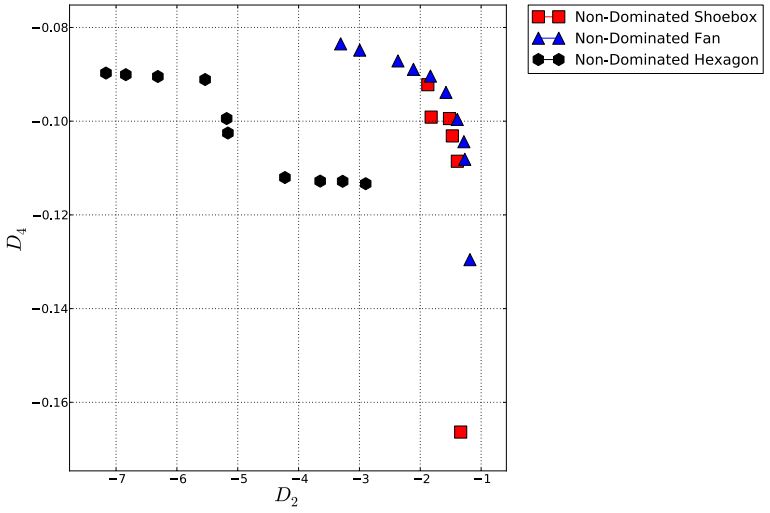


(a) EDT vs. LF_{early}

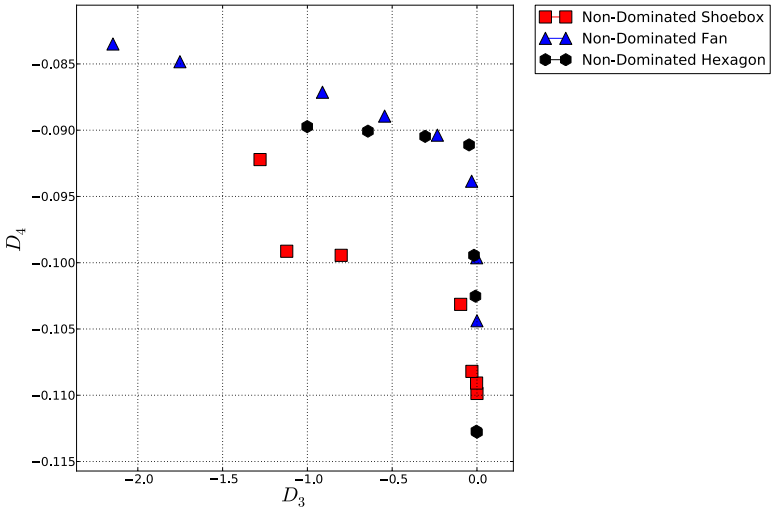


(b) C_{80} vs. G

Figure 13.20: Pareto Fronts comparisons of the shoebox, fan shaped and hexagonal rooms.



(a) C_{80} vs. LF_{early}



(b) G vs. LF_{early}

Figure 13.21: Pareto Fronts comparisons of the shoebox, fan shaped and hexagonal rooms.

There is contrast between D_1 and D_2, especially in the shoebox and fan shaped rooms. As we saw before, hexagonal rooms have two separate group of solutions with high D_1 values. This is visible also in this figure. Shoebox rooms seem to have a more continuous Pareto Front.

If we had to select a room type based on D_1 and D_2 we would surely not choose the hexagonal type, but between the shoebox and the fan, the choice is not clear. There is a group of shoebox and hexagonal rooms that are non-dominated between each other.

Reverberation vs. Sound Strength

Figure 13.19b shows the Pareto fronts of all 3 room types for the EDT and G parameters. All 3 room types have individuals with a perfect D_3 value ($D_3 = 0$. This means that those rooms have G values that are all above Barron's minimum G curve. However, the highest values for D_1 are found in rooms that do not have a perfect D_3 score. Hence there is some contrast in this objective space, especially so for shoebox and fan shaped rooms.

As seen in the previous comparison, D_1 values are highest among shoebox rooms, but some of them are dominated by fan and hexagonal rooms on the basis of D_2. There seems to be more solutions with a good compromise between D_1 and D_2 in fan shaped rooms, some of these are inverse fans.

Reverberation vs. Lateral Fraction

Figure 13.20a shows the Pareto fronts of all 3 room types for the EDT and LF_{early} parameters. Most shoebox rooms have considerably lower D_4 values than hexagonal and fan shaped rooms. This seems to suggests that good lateral energy distribution is most present when sidewalls are not parallel. Only very narrow shoeboxes have good D_4 values, but they do not have high D_1 values.

Pareto front shapes in this case are almost orthogonal for hexagonal and fan shaped rooms, indicating little contrast, especially when compared to shoeboxes, they have a linear and almost 45 degree Pareto front. This shows that there is little possibility of achieving a good compromise between D_1 and D_4 in the shoebox type.

Clarity vs. Sound Strength

Figure 13.20b shows the Pareto fronts of all 3 room types for the C_{80} and G parameters. This graph shows very little contrast between these two

functions. This behavior was already seen in the fitness landscapes for all three room types. Many rooms in all types have a perfect D_3 score. In this case, those rooms with perfect D_3 scores also include the best D_2 rooms (especially in the fan shaped and hexagonal rooms), further confirming the lack of contrast between these two functions.

This direct Pareto comparison shows that fan shaped rooms dominate all other types. Fans that are open towards the room and have a short height seem to have better distributions of C_{80} and G optimal values than all other rooms in this case study.

Clarity vs. Lateral Fraction

Figure 13.21a shows the Pareto fronts of all 3 room types for the C_{80} and LF parameters. This graph shows that in all room types these two functions are contrasting, perhaps more evidently so in fan shaped rooms. The fan type contains both the best performing room for D_2 and for D_4. Moreover, the Pareto optimal rooms in the fan type dominate all other rooms from the shoebox and fan shaped types. In other words, a Pareto Analysis of all types for D_2 and D_4would place only fan shaped rooms in the front.

Shoebox rooms are generally better at distributing optimal C_{80} values that the hexagonal shaped rooms and they have very similar values when it comes to lateral energy. Therefore we see shoebox rooms dominate hexagonal rooms in this pair of functions.

Sound Strength vs. Lateral Fraction

Figure 13.21b shows the Pareto fronts of all 3 room types for the G and LF_{early} parameters. These two functions show contrast, in all three types there seems to be an almost orthogonal relationship between D_3 and D_4. We can see that the rooms that best distribute lateral energy, and therefore have higher D_4 values, are not the ones with a perfect D_3 score.

Reverberation, Clarity, Sound Strength and Lateral Fraction

The previous comparisons and Pareto fronts were made by considering pairs of objective functions. They help us understand the relationship between these functions and the three room types in great detail. But we are also interested in considering all fitness functions at the same time. While it is not possible to visually represent the Pareto front shape resulting from considering all functions in a four function problem such a this one, we can

still study the Pareto front in other ways. Figure 13.22 shows the shape of the rooms in the Pareto front for the four fitness functions. The following table shows the room type and x_1 and x_2 values for the Pareto rooms:

Shoebox		**Fan**		**Hexagonal**	
X_1	X_2	X_1	X_2	X_1	X_2
15.0	21.0	13.0	16.0	27.0	15.0
25.0	15.0	14.0	16.0	28.0	15.0
25.0	16.0	15.0	15.0	48.0	15.0
26.0	15.0	18.0	16.0	49.0	15.0
26.0	16.0	19.0	15.0		
27.0	15.0	19.0	16.0		
28.0	15.0	21.0	15.0		
28.0	16.0	27.0	15.0		
29.0	15.0	29.0	16.0		
30.0	15.0	30.0	15.0		
31.0	15.0	30.0	16.0		
31.0	16.0	31.0	16.0		
32.0	15.0	32.0	16.0		
33.0	15.0	33.0	16.0		
34.0	15.0	33.0	17.0		
35.0	15.0	33.0	18.0		
13.0	15.0	34.0	15.0		
		34.0	16.0		
		35.0	15.0		
		35.0	16.0		
		35.0	17.0		
		35.0	18.0		
		35.0	19.0		
		35.0	20.0		
		35.0	22.0		
		35.0	23.0		

We can see that the Pareto front is comprised of rooms in all three types. Fan shaped rooms are the most present in the Pareto front, followed by shoebox, and then we see just a few hexagonal rooms. We can also see that most Pareto rooms are low. Rooms from 15 to 17 meters in height comprise about 90% for the front, rooms higher than 17 meters are mostly dominated rooms.

The shoebox rooms that belong to the Pareto front are mostly wide and mid width rooms, there are only two very narrow and long rooms. We saw

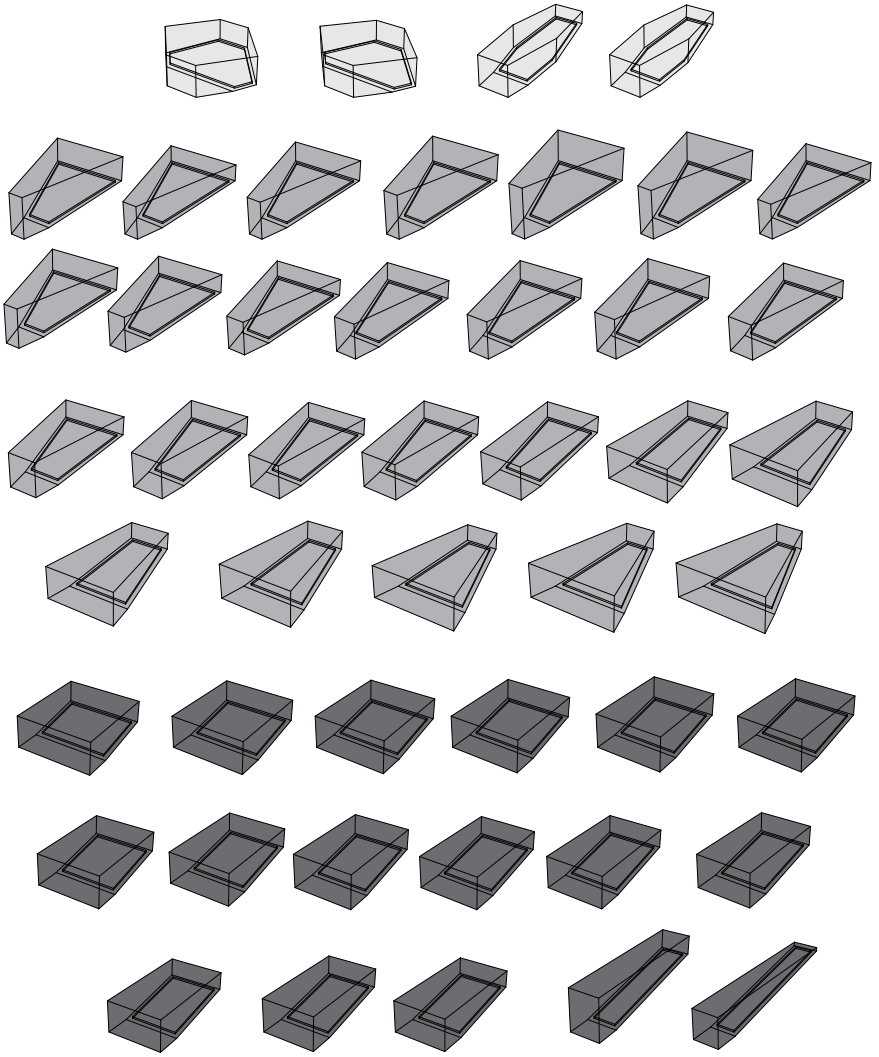


Figure 13.22: Pareto front rooms for the four fitness functions D_1, D_2, D_3 and D_4. Shoebox rooms in dark gray, fan shaped rooms in medium gray and hexagonal rooms in light gray.

above that wide shoeboxes have high D_1 values, and have the fairly high D_2 values as well. Most of them range from 25 to 35 meters in width, meaning a variation also in length (from 50 to 40 meters). These are quite significant differences in plan between rooms.

Fan shaped rooms also vary greatly in plan. Most rooms in this type are regular fans, but there are seven inverse fans in the Pareto front. The most significant difference between the fan rooms in the front is the angle between their sidewalls. Inverse fans have a place in the front thanks to their high D_1 values, while regular fans have very high values in all other functions.

The are four hexagonal rooms the front, two of them are very wide with very open sidewalls, the other two are very narrow, with nearly parallel sidewalls. These two groups correspond to the two groups seen in the fitness landscapes, they have high D_1 and D_4 values.

13.5 Conclusions

Case study 6 presents a parametric study of three concert hall types by means of their distribution of acoustic objective parameters. Distribution of acoustic quality is studied with the method described in chapter 13. The most significant geometrical aspects of three concert hall types are parametrized into three parametric models. These parametric models do not comprise all of the possible variations within these concert hall types. Since only two parameters were used to describe each room type, not all variations of the concert hall geometry were included.

Fitness landscapes and Pareto fronts were plotted from the distribution analysis performed on all three concert hall types. A comparative analysis of room types is made from these figures, and some specific geometrical attributes important to sound quality are derived.

Results show that a particular set shoebox, fan and hexagonal shaped rooms have distinct advantages over all other rooms in the study. The information provided to us by this search process would be of fundamental importance if we were to proceed with these study in a design process. The information obtained by this study could be interpreted in four ways in order to proceed:

- A shoebox from the Pareto set is selected and its acoustic distribution is further improved by including other features in the room.
- A fan shaped room from the set is selected and further developed.

- An hexagonal room from the set can be chosen and further developed.
- A more detailed study of all three types is carried out including other variables or different parametric models.

The first possibility is the selection and further development of the shoebox room. If we look at the values shown in the objective spaces comparisons for the shoebox, we can see that the shoeboxes are most deficient when it comes to D_4 and D_2. It could be argued that by selecting a narrow shoebox, and studying sidewall modifications (such as side balconies or reflectors) improvements to D_4 and D_2 could generate an optimal room.

The second possibility is the selection and improvement of a fan shaped room. In this case the most important improvement necessary for the fan rooms is their D_1 values. Reverberation can be improved by slight modifications in the room materials, especially in the audience seats. Also, a furtter refinement of the room dimensions, leading to higher volumes can also improve reverberation without dramatic changes to other distribution values.

The third possibility involves the selection of an hexagonal room. Narrow hexagonal rooms are the ones with the highest values in their type for all functions. They also rank very highly when compared to all types when it comes to D_3 and D_4. Interestingly, this type needs improvement both in reverberation and clarity. It is therefore difficult to visualize a clear step for improvement, but a study in greater detail in the sidewall angles around the stage might be good start.

The fourth possibility is that of creating a new search process with different, perhaps more detailed parametric models, using the information obtained in this study. For example, there is a way of significantly improving the performance of hexagonal rooms by including more control over sidewall angles and room width. As we saw in the fitness landscapes, hexagonal rooms were too wide or too open in some instances. A parametric model that works on these variables can probably obtain better results, giving designers more options. Different angles for stage and audience sidewalls are a great possibility of hexagonal rooms. This was not included in the present parametric study.

14

Acoustic simulation of complex shapes in concert halls

Ray tracing acoustic simulation (Krokstad et al. 1968) has been used for predicting the room impulse response for a number of years and represents an important method that is present in most of acoustic simulation packages.*

When it comes to describing the geometry of curved surfaces commercial acoustic simulation software depends on their discretization into small planar segments. This is partly because planar segments are needed to use the Image Source Method, partly because other more precise geometrical models are not easily implemented. These kind of segmented surfaces are sometimes called meshes, and they are the kind of surface geometry contained in DWG and DXF files, popular formats used for importing of geometry into commercial acoustic simulation software.

The advent of advanced computational geometry, in particular the use of NURBS, give us the possibility to better represent free-form complex and curved geometry. This presents the opportunity to significantly improve raytracing simulation models by accurately representing curved geometry, as has been recommended by researchers in the past (Kuttruff 1993, Vercammen 2010, Mommertz 1995). This chapter conducts a comparison of simulations done with and without the use of NURBS for a series of curved surfaces in order to understand the potential of this new possibility. In so doing we can demonstrate that with NURBS geometry we can correctly cal-

*The contents of this chapter were published in the proceedings for the AIA-DAGA symposium on acoustics in Merano in 2013 (Méndez Echenagucia, Astolfi, Shtrepi, van der Harten & Sassone 2013*c*)

culate reflection angles for curved surfaces of any kind, without resorting to discretizations. We also show, as previously seen in (Kuttruff 1993, Vercammen 2010), that with raytracing algorithms, using mesh geometry can that give out erroneous results, especially with regards to concave surfaces, resulting in the failure to detect sound concentrations.

14.1 Sound reflection from convex surfaces

In order to study the potential of NURBS surfaces in acoustic raytracing simulation, we turned to theoretical studies (Kuttruff 1993) and (Vercammen 2010) of sound fields of curved surfaces and compare their findings to results obtained with a NURBS raytracing acoustic simulation algorithm developed for this PhD thesis. A cylinder, a sphere and an ellipsoid serve as examples of concave surfaces studied in these research papers, and they are reproduced here as described bellow. Figure 14.1 shows the geometric and acoustic setup for all three geometry cases.

14.1.1 The Image Sources Method

In the case of Image sources, a number of flat segments that gives a good approximation to the correct sound pressure in the centre can be estimated (Kuttruff 1993, Vercammen 2010, 2012). This number of segments is frequency dependent and can be rather large, especially in the case of double curvature surfaces. In this chapter the image source method will not be considered, since this method is not possible with NURBS, due to the fact that there are an infinite number of image sources, and no flat segments.

14.1.2 The raytracing NURBS simulator

The studies described in (Kuttruff 1993, Vercammen 2010, 2012) call for a first order raytracing simulation of concave perfectly reflective surfaces. Such an algorithm was written with the use of NURBS geometry, inside Rhinoceros.

As explained above, Rhinoceros is a commercial Computer Aided Design package that is used in various fields, particularly those who employ complex curved surfaces, as it is capable of representing and operating on NURBS geometry. Rhinoceros also allows the user to customize its functionality by creating commands by calling a series of Rhinoceros functions. This is done through scripts that can be written in Visual Basic or Python programming

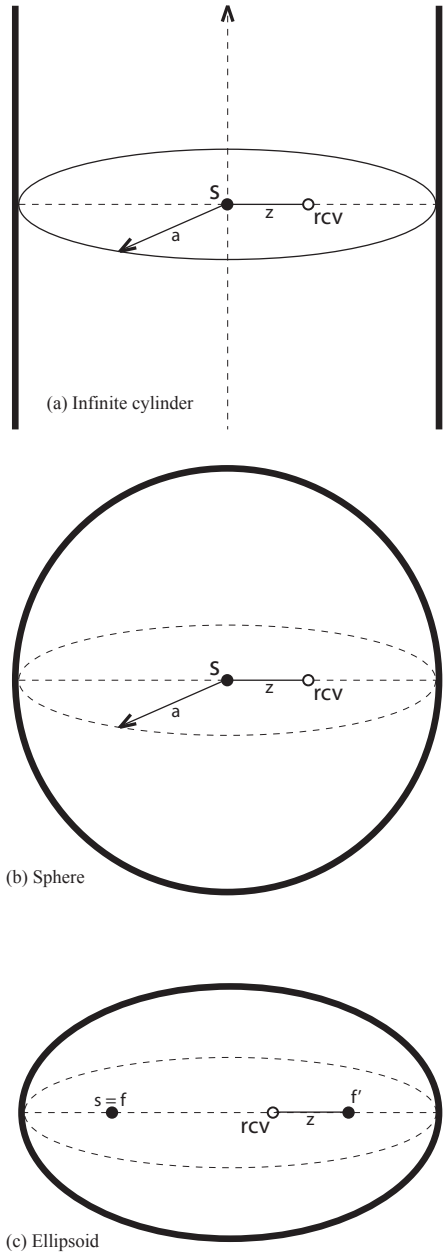


Figure 14.1: Cylinder (a), Sphere (b) and Ellipsoid (c) Geometry and acoustic setup for NURBS and Mesh raytracing analysis.

Languages. In this case a Python class containing a series of functions was written.The class of functions is able to cast rays into NURBS or mesh geometry and to plot their path, as well as create spherical receivers that can detect the energy inside each ray. The sound power in each ray can be described by:

$$P_i = P/N \qquad \text{(W)} \tag{14.1}$$

where P_i is the power of the i_{th} ray, P is the source power and N is the number of rays (Cingolani & Spagnolo 2005). The developed ray tracer does not take air attenuation or absorption from reflections into account, since this is not a part of this study, and the source is always considered to be omnidirectional. In (Cingolani & Spagnolo 2005) we find that the ray power detected by the source can be normalized in relation to the ray length inside the receiver volume. In this case the sound power in each ray will be:

$$P_{i,m} = \frac{l_{i,m}}{D_m} \cdot \left(\frac{P}{N}\right) \qquad \text{(W)} \tag{14.2}$$

where $P_{i,m}$ is the power of the i_{th} ray in the m_{th} receiver, $l_{i,m}$ is the i_{th} ray segment inside the m_{th} receiver volume, D_m is the receiver diameter. The intensity at the receiver will be:

$$I_i = \frac{P_{i,m} \cdot \Delta t \cdot c}{V_m} = \frac{P_{i,m} \cdot l_{i,m}}{V_m} \qquad \text{(W/m}^2\text{)} \tag{14.3}$$

where V_m is the m_{th} receiver volume, c is the sound speed and Δt is travel time of the ray inside the receiver volume (Xiangyang et al. 2003, Vercammen 2012). From the sound intensity we can calculate the mean squared sound pressure with:

$$p^2_{rms} = \rho c \sum_{i=1}^{N} I_i = \rho c \sum_{i=1}^{N} \frac{P_{i,m} \cdot l_i}{V_m} \tag{14.4}$$

where ρ is the air density. Since no absorption or attenuation is considered in the concave surfaces examples in (Kuttruff 1993, Vercammen 2010, Mommertz 1995) the ray tracer developed for this PhD thesis makes no considerations regarding sound frequencies. All of the simulations in this chapter assume conditions in which the rules of geometrical acoustics can be applied.

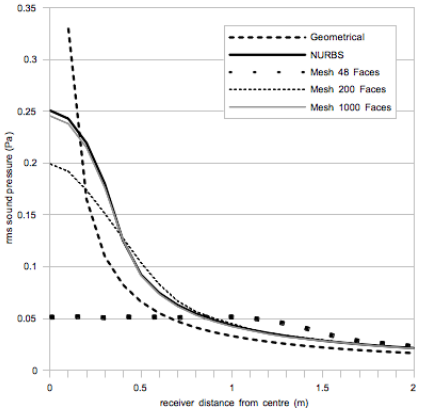


Figure 14.2: Infinite Cylinder with 10 meter radius - mean squared sound pressure in relation to receiver distance from cylinder centre. All receivers with a 0.8 m diameter.

14.2 Cylinder Study

Kuttruff describes in (Kuttruff 1993) a perfectly reflective infinite cylinder with a source at its centre axis. Considering only the first reflection and discarding direct sound, he then calculates the root mean squared sound pressure in various receiver positions near the focusing area of the cylinder. In the present chapter we consider receivers in the same plane as the source (figure 14.1(a)). We first carry out simulations with the NURBS ray tracer using this cylinder as a case study. Figure 14.2 shows the mean squared sound pressure in a series of receivers close to the sound source as estimated by the raytracer for a NURBS cylinder as well as cylinder polygons of increasing number of faces using equations 14.3 and 14.4. The results are also compared with a geometrical approximation of the correct results near the source as described in (Kuttruff 1993). This approximation is defined by the following equation:

$$p_{rms}^2 = \frac{\rho c \cdot P}{4\pi \cdot a \cdot z} \tag{14.5}$$

where a is the cylinder radius and z is the receiver distance from the source.

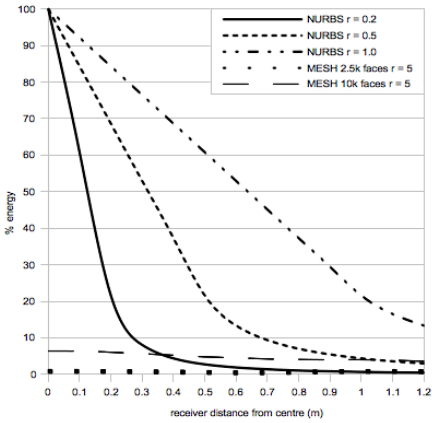


Figure 14.3: Sphere study - Percentile energy in relation to receiver distance from centre for NURBS and mesh geometry, with various receiver radii.

The geometrical approximation is not valid at the centre point (source position) because it results in infinite pressure. We can see that the NURBS results are close to the geometrical approximation, and we can also see that the number of faces necessary to approximate the NURBS result is over 1000 faces. This number is considerably lower if we use the Image Sources method (Kuttruff 1993). Like in any ray tracing simulation, the results are dependent on the receiver radius. A larger receiver volume will be able to detect a larger number of rays. Different acoustical simulators use different approaches from the one employed in this chapter, and their accuracy varies with changing rooms, receiver positions and ray paths (Xiangyang et al. 2003). This issue is not addressed in this PhD thesis, it is merely noted as part of the observations taken in the simulations.

14.3 Sphere Study

In (Vercammen 2010, 2012) Vercammen illustrates another case study of sound concentration from concave surfaces, this time involving a perfectly reflective sphere with a sound source at its centre (figure 14.1 (b)). Like in the cylinder example, only the first reflection is taken into account, attenuation, absorption from reflections and direct sound energy are neglected. In

this case we can expect 100% energy to return to the sphere centre, which provides us the opportunity to test the different geometrical models. Figure 14.3 shows the results of the NURBS and Mesh simulations with different receiver radius using:

$$P_{tot} = \sum_{i=1}^{N} P_{i,m} = \sum_{i=1}^{N} \frac{l_{i,m}}{D_m} \left(\frac{P}{N} \right) \qquad \text{(W)} \tag{14.6}$$

Where P_{tot} is the total energy detected by the receiver. The y axis in Figure 14.3 reports a the percentage of emitted energy P that is detected by the receiver P_{tot}. Results outside the sphere centre vary with receiver size as bigger receivers will detect more rays. However, if all reflections are calculated correctly, in the exact sphere centre all rays should go through the receiver perpendicularly and $l_{i,m}$ must be equal to D_m making $l_{i,m}/D_m = 1$. Thus with the correct ray reflection angles, the receiver size in the centre of the sphere is not relevant to the result. Figure 14.3 shows that at the sphere centre all NURBS calculations converge into 100% energy, proving a correct reflections in this geometrical model. Mesh calculations on the other hand, differ from receiver size and are far from 100% energy at the centre.

The root mean squared sound pressure in the focal point is explained in (Vercammen 2012) and calculated as follows:

$$p_{rms}^2 = \frac{1}{2} \hat{p} k^2 (1 - cos\theta_m)^2 \tag{14.7}$$

where $\hat{p}$ (N/m) is the amplitude descriptor that represents the pressure at 1 meter from the monopole source, k is the acoustic wave number and θ_m is the opening angle of the spheric segment, in the case of the full sphere $\theta_m = \pi$.

Formula 14.7 takes into account sound frequency by introducing the wave number k. Considering a frequency of 1000Hz, for a sphere and $\hat{p} = 1$(N/m) we get a ms pressure value of 73.27 (Pa^2), corresponding to an SPL of 112.6 (dB). Vercammen suggests a frequency dependent receiver diameter of $\lambda/2$. If we carry out this experiment in the center with a receiver of $\lambda/2 = 17$ cm an SPL of 120 dB is obtained. The receiver diameter to best approach the correct value has a diameter of 40 cm with an SPL of 112.7dB. As it is pointed out in (Vercammen 2012) this calibration of receiver size is impossible in cases where the exact value is unknown beforehand. However with a $\lambda/2$ receiver diameter a good approximation can be made.

Figure 14.4 shows the influence of receiver sizes in the sphere case study by plotting SPL in various receivers near the centre. NURBS geometry

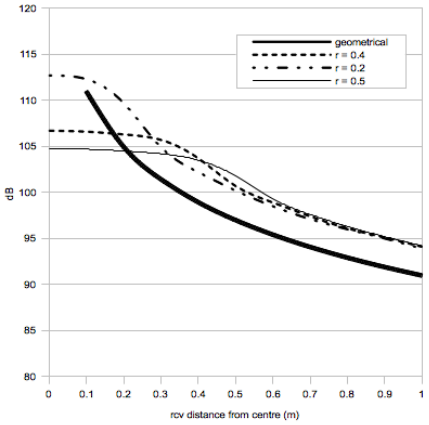


Figure 14.4: Sphere with a 10 m radius SPL (dB) in relation to receiver distance from centre with various receiver radius.

results are compared to a geometrical approximation described in (Vercammen 2012) and then converted to SPL. As in the cylinder example, this geometrical approximation is not valid in the centre, for the centre values we use:

$$p_{(z)^2 rms} = \frac{1}{2}\hat{p}/z^2 \qquad \text{(Pa)} \tag{14.8}$$

14.4 Ellipsoid Study

A similar series of simulations is done using an ellipsoid surface, the object in this case being to study the number of flat surfaces necessary to approximate the result obtained at the focusing area by the NURBS surface in a double curvature case. The sound source was placed in one of the ellipsoid's foci, and the receivers in the axis of the second focus (figure 14.1 (c)). Like in the case of the sphere, here too we can conclude that all rays should pass through a single point, in this case the focus opposite the source, therefore the receiver at this focus should detect 100% of the energy emitted by the source. Figure 14.5 shows that a mesh model with 10.000 faces has a result that is still far from correct, not reaching 20%. Mesh geometry with more

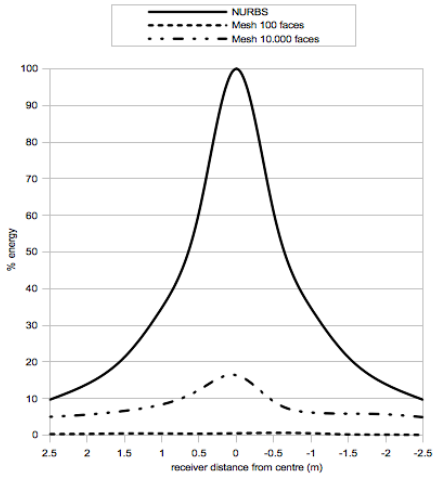


Figure 14.5: Ellipsoid study resutls Percentile energy in relation to receiver distance from ellipsoid focus, for NURBS and Mesh geometry with various number of mesh faces. All simulations done with a receiver of radius 0.5.

than 10.000 faces were not investigated for computational complexly reasons. NURBS geometry on the other hand shows a perfect 100% result when the receiver is placed in the second focus point.

14.5 Concave Surface Study Conclusions

The results of the concave examples show that NURBS representation of curved surfaces significantly improve the capability of acoustic simulation to predict sound concentration, because ray reflections are exact in curved surfaces without need for discretization.

The amount of planar sections required for accurate results in a ray-tracing model is not frequency dependent (Vercammen 2010), and it is also dramatically larger for double curvature surfaces when compared to single curvature.

Comparing results from the NURBS simulations to various mesh simulations, it's evident that increasing the number of segments improves the results, but in order to achieve reasonable results, the number of segments becomes unfeasibly large for current computational tools.

15

Early Sound Analysis of concert halls

When it comes to the design of concert halls and auditoria, the collaboration between architects and acoustic consultants can be difficult if one considers that they are both trying to define the form of the room but with completely different and sometimes contrasting priorities*.

As shown in previous chapters, computational acoustic simulation software is capable of estimating the impulse response and room acoustics parameters at receiver positions for diverse rooms with a good degree of accuracy. However these tools on their own do not give designers an idea as to the direction to take to improve results, nor do they help explore new room shapes.

Recent designs for concert halls around the world by renowned designers have prominently featured complex curved geometry. Concert hall designers have long employed convex surfaces in concert halls and concave surfaces in outdoor concert spaces, their potential to distribute energy where is needed has been proven by many examples (Vercammen 2012). Like any concave surface, complex double curvature surfaces need to be carefully studied and optimized to obtain desirable sound reflections and avoid sound concentrations.

In this chapter a form exploration tool is proposed, that can help designers and acoustic consultants interact with complex geometry and generate forms that can potentially distribute early sound energy in an optimal way.

*The contents of this chapter were published in the proceedings of the International Symposium on Room Acoustics 2013 in Toronto (Méndez Echenagucia, Astolfi, Shtrepi, van der Harten & Sassone 2013*b*)

15.1 Room Shape and Early Sound

Early reflections in concert spaces have been determined to be crucial to the overall acoustic quality by many researchers in the past (Barron 1971, Marshall 1994, Jurkiewicz et al. 2012, Patynen et al. 2013), and the characteristics of early sound are very dependent on the room shape, while the late sound energy is more dependent on average room absorption (Bradley 2011).

15.1.1 Insufficiency of Room Acoustics Parameters

ISO 3382-1 acoustic parameters are commonly used for the design of concert halls, but they are perhaps not detailed enough to describe the early portion of the sound that arrives at the listeners. Researchers have said that these indexes are insufficient to describe in detail the early sound, and that there are rooms with identical index values at various positions that have quite different subjective perceptions (Patynen et al. 2013, Bradley 2011, Bassuet 2011).

15.1.2 Studies and visualization methods of early sound

Research on the perception of early sound showed the importance of the first milliseconds in the overall perception of sound. Pioneering work by Haas determined echo audibility thresholds and the precedence effect (Haas 1951). He emphasized the importance of the first 30 milliseconds and the subjective perception of loudness. Barron later studied the effects of a single first reflection from side walls and ceilings (Barron 1971). He was able to study specific thresholds, source broadening and subjective impressions for side walls and ceilings. This study emphasized the importance of early lateral reflections.

Hidaka et al. published a comparison between shoebox and vineyard rooms that highlighted the importance of energy arriving in the first 80 milliseconds (Hidaka et al. 2008). They state that successful rooms show high sound strength in this first 80 milliseconds, corresponding a good amount of early reflections, and a high G_{early}, this is also mentioned by Bradley in (Bradley 2011). Hidaka et al. go on to say that before the first 80 milliseconds (for the 500 and 1000 Hz frequency bands) reflections angles and localization are perceived more accurately, reflections after 80 milliseconds are perceived by listeners as enveloping energy.

Different methods have been developed to study and visualize early sound. They show different levels of detail and have different specific purposes, but they all give us information about the early sound that we do not find in ISO 3382-1 acoustical indexes (Marshall 1994, Patynen et al. 2013, Bassuet 2011, Krokstad et al. 1968, Oguchi et al. 1988). Some of these studies have been done using measured impulse responses in existing concert halls to better understand their objective qualities while others are especially developed for design purposes and employ computational acoustic simulations. These methods help us understand differences in acoustical qualities, but they do not have ideal or optimal values for design purposes, as they have not all been compared to ideal existing conditions.

15.1.3 Uniform distribution of sound energy in time and space: A multi-objective problem

Acoustics parameters and measurements can be quite detailed when it comes to the distribution of energy in time. Early to late ratios (C_{50},C_{80}) for example are widely used and optimal values have been prescribed, but all of these indexes take single receiver positions into account. However, as previously stated and studied in the distribution chapters, the overall quality of a room cannot be summarized by index values in one single position, or a few advantaged positions close to the sound source. A good room considers the quality in all listening positions.

The distribution of sound energy in space and time poses a complex geometrical problem, reflections have to be directed in such a way as to provide sound energy uniformly over space and with time intervals in such a way as to satisfy subjective preferences. With a limited amount of energy being emitted by a given source, our goal is to generate room shapes that produce plenty of reflections in desirable time intervals over the whole listening area. We can say that room shapes that direct energy to the audience in the first milliseconds are in direct contrast to shapes that direct energy later in time. Hence, generating forms that evenly distribute energy in time and space is a multi-objective problem.

15.1.4 Time-Windows

In their pioneering paper in raytracing acoustical simulation, Krokstad et al. proposed the subdivision of sound reflections in rooms into "time-windows" that contain the reflections inside given time intervals (figure 15.1)(Krokstad et al. 1968). We can also consider receiver dependent time-windows in which

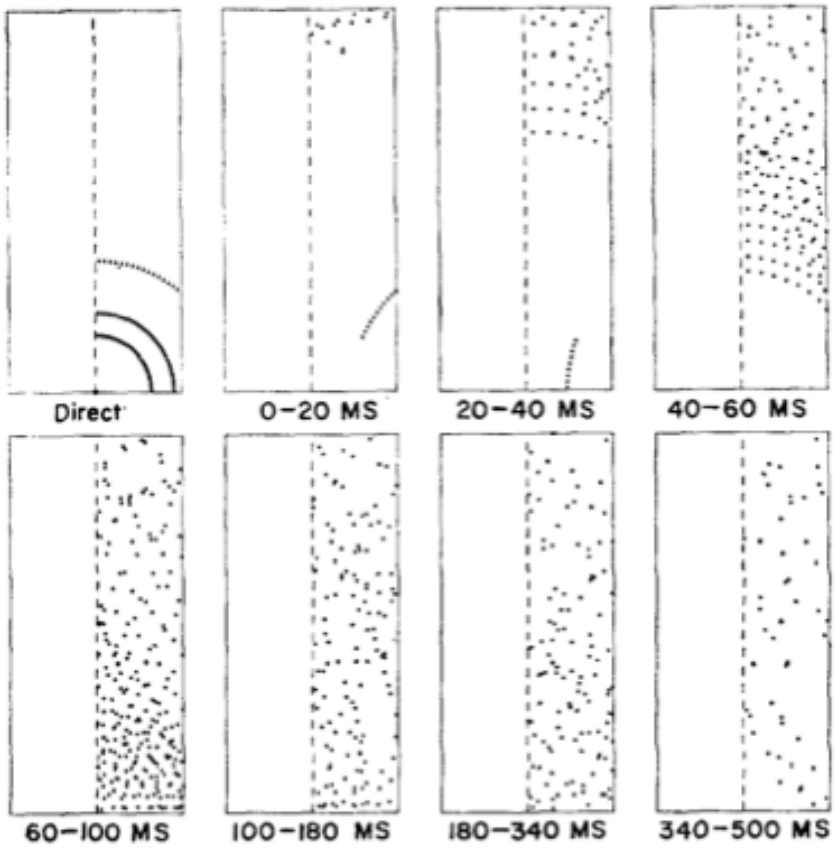


Figure 15.1: Krokstad's Time-Windows - An example of a spatio-temporal analysis.

time intervals are calculated from the arrival of the direct sound to each particular receiver. Time is counted by the algorithm from the arrival time of the direct sound of each receiver. By dividing the energy that arrives at the listening positions into time window we can study how rooms distribute energy in time, and evaluate the uniformity of this energy in the space of the audience area.

15.2 Tool for the uniform distribution of early sound in concert spaces

In this chapter a multi-objective genetic algorithm capable of finding room shapes that evenly distribute sound reflections in time and space using a NURBS based raytracing acoustic simulator is presented. This tool follows the same basic method described above in the structural case studies, with the exception that the fitness functions is calculated by a raytracing simulator and an acoustic study both developed for this PhD thesis by the author. A MOGA coupled with a parametric model and the acoustic simulator are employed to search for high-performing solutions.

15.2.1 The Ray tracing NURBS simulator

In the previous chapter a NURBS ray-tracing acoustic simulator was presented for the study of concave surfaces. The raytracer developed for this PhD thesis was also employed in the study of early reflections and complex curved geometry presented in this chapter. As was previously stated a class of python functions in combination with Rhinoceros comprises the raytracer. This class is able to cast rays into complex and free-form NURBS geometry and to plot their path, as well as create flat receivers that can detect ray reflections inside their area. The algorithm separates the reflected sound rays into time-windows that are calculated after the arrival time of the direct sound for each receiver. Since this study is about sound reflections, the direct sound rays are deliberately taken out of the simulation. Rays are reflected from the room surfaces until the travel time of the ray exceeds the time windows considered in the study, or until the ray reaches the audience. This means the rays are only allowed to reach the audience area once.

15.2.2 Acoustical fitness functions

MOGAs are capable of optimizing several objectives or fitness functions, through the combination of design variables. Fitness functions and design variables need to be well formulated by the designer in order for the MOGA to properly search for optimal solutions. In this chapter we are studying the distribution of sound energy in time and space, so we formulated 3 fitness functions based on the time-windows explained above.

The first fitness function considered a time-window from t_0 to t_1 in milliseconds (where instant t_0 is considered to be after the arrival of direct sound), the second from t_1 to t_2, and the third from t_2 to t_3.

The specific values for t_1, t_2 and t_3 should be chosen in the basis of the specific case being studied, for example, cases involving ceilings, side-walls, canopy reflectors or balconies can have different windows into account. The actual number of time-windows can also be specifically taylored for each design problem.

The uniform distribution of energy in space is the objective of each Fitness Function. For each time window the following equation was used to estimate the ideal amount of reflections for each receiver:

$$R_r = \frac{R_{tot}}{N_w \cdot N_r} \tag{15.1}$$

where R_r is the ideal number of reflections for each receiver inside a time-window, R_{tot} is the total number of rays emitted by the source, N_w is the number of time-windows and N_r is the number of receivers. With this ideal value of reflections in each receiver, we can calculate an error function that tells us how far is the room from an ideal reflection pattern for that time window:

$$E_{tot} = \sum_{i=1}^{N_r} |R_r - R_i| \tag{15.2}$$

where E_{tot} is the total error for the time-window and R_i is the number of rays in the i_{th} receiver. Equation 15.2 describes the fitness function that was used for each time window. The use of the absolute value of the difference $R_r - R_i$ means that the fitness function evaluates if the solution given errs by giving each receiver too many or too few ray reflections.

The minimization of these 3 fitness functions guarantees that the room shape found in the process delivers an equal amount of ray reflections in each time-window and also in each receiver. Thus ensures that the room

will provide sound reflections in all time windows, and that there will be no sound concentrations or shadowed areas anywhere in the audience area.

15.3 Case Study 7: Complex curved ceiling for a concert hall

The potential of the proposed design tool is explored by studying a shoebox concert hall with a curved reflective ceiling. The object of the study is to generate ceiling shapes that evenly distribute early sound energy oven the audience space and over 3 separate time-windows. The above mentioned ray-tracing algorithms for NURBS surfaces and the early sound analysis tools were employed, in combination with NSGA-II and a parametric model.

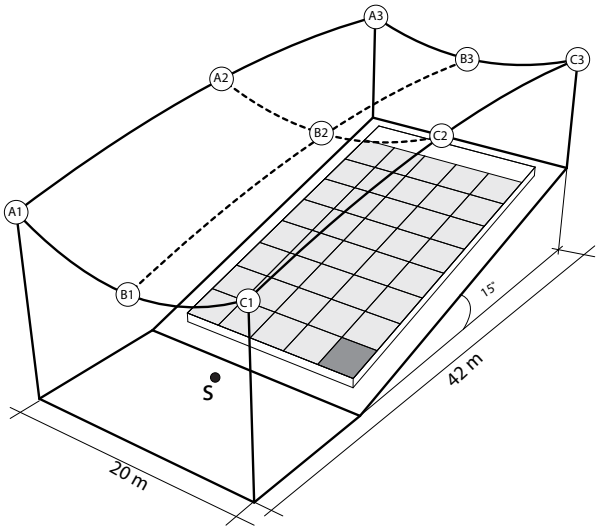


Figure 15.2: Parametric model for the acoustic ceiling case study.

15.3.1 Parametric Model

In any MOGA a set of design variables is needed to search for optimal solutions. In the case of acoustic geometry, design variables should define

Variable	Control Points	Movement Axis	Range of Movement
x_1	A1 and C1	Z	5 to 20m
x_2	B1	Z	5 to 20m
x_3	A2 and C2	Z	10 to 20m
x_4	B1	Z	10 to 20m
x_5	A3 and C3	Z	10 to 20m
x_6	B3	Z	10 to 20m

Table 15.1: Design variables, axis of movement and range of movement.

the changing geometry we want to evaluate, (for example we can use height, width and length dimensions for shoebox optimization). The case study presented in this chapter works with NURBS geometry to generate complex curve reflecting surfaces. These surfaces are specified by control points, and each control point is defined by its X, Y and Z coordinates. The case study described bellow employs fixed values of X and Y for all control points, and the Z values of these points are our design variables. Figure 15.2 shows the shoebox with a curved reflective ceiling defined by 9 control points. Design variables in geometric problems are always confined to a domain or range of movement, in order to limit the search possibilities, saving calculation times, and also to avoid undesired geometry.

Table 15.1 shows the design variables for the present study, their movement coordinates and range of movement. This parametrization of the shoebox ceiling ensures that the MOGA is capable of generating single and double curvature surfaces, both concave and convex towards the audience. Surfaces will always be symmetrical in the longitudinal axis of the room, and the area of ceiling above the stage can reach a height of just 5 meters if desired. If all design variables are equal, a flat ceiling is obtained, thus this study also considers the possibility that curved surfaces are not better than flat ones.

A single omnidirectional spherical source was employed. The audience has an inclination of 15 degrees from the stage plane. The audience area was subdivided into 40 flat segments of equal width and length.

15.3.2 Fitness functions

The total error function described in equation 15.2 was used as a fitness function for 3 different time-windows. The time-windows used in this case

study are described in the following table:

Time-window	t_{start} (ms)	t_{end} (ms)
1	0	80
2	80	120
3	120	200

Since there are 3 time-windows in this problem, there are 3 fitness functions. As we have seen above, NSGA-II can handle a high number of fitness functions and is capable of finding a Pareto set for the problem. The multi-objective problem studied in this case study can therefore be described with the following set of equations:

$$Case\ Study\ 7 \begin{cases} Minimize & f_{1(x)} = E_{tot,0-80}, \\ Minimize & f_{2(x)} = E_{tot,8-120}, \\ Minimize & f_{3(x)} = E_{tot,120-200}, \\ subject\ to & 5 \leq x_{1,2} \leq 20. \\ & 10 \leq x_{3-6} \leq 20. \end{cases} \tag{15.3}$$

15.3.3 Genetic algorithm inputs

NSGA-II explores 100 generations with 10 individuals in each generation. The overall genetic inputs for this case study is as follows:

Case Study 7		
Population Size (N)	10	
Number of Variables	6	
Number of binary digits	8	
Variable Domains	$x_{1,2} \in [5, 20]$	$x_{3-6} \in [10, 20]$
Mutation Probability (p_m)	0.2	
End Condition	End after 100 generations	

15.3.4 Results

Figures 15.3 and 15.4 show the ray distribution in the audience area for the 3 time windows of a series of room shapes, all belonging to the Pareto Front of solutions found by our design tool. We can see that the shapes obtained by the MOGA are varied and leave plenty of options for the architect and acoustical consultant to consider. Fitness values for each window are better in some individuals that in others, this is due to the fact that optimizing

rays for one window means there are fewer rays left over for the others, hence this is a true multi-objective problem with contrasting functions.

Solution 8 has the minimum value for window 1 and has good values for the other time windows, however it does neglect the first rows in the second window. Solution 5 has the minimum value for window 2, but it causes sound concentration in the first rows for window 3. Solution 3 has the minimum value for window 3 and has also good values for the other two windows. Solutions 0 and 2 are good intermediate solutions that have low E_{tot} values for all 3 windows.

15.3.5 Conclusions

An interactive acoustic search design tool is presented. It is capable of generating surface shapes that evenly distribute reflected rays in given time-windows and receiver areas.

Different shapes are better at distributing energy in particular time-windows than others, geometrical characteristics benefit certain areas of the room, and designers can consider the advantages of all suggested geometry.

Results show that the fitness functions formulated in the case study were in fact contrasting functions, hence a multi-objective approach is necessary to minimize all of them.

The effects of such distribution of sound energy towards the generally absorptive audience area has on later sound energy and reverberation needs to be further investigated.

Future research in this method should consider side-walls to study the 0-30 time-window (precedence effect, see (Haas 1951)).

More complex geometries and other concert hall typologies could be explored using the above described method, in order to study innovative room configurations. The possibility of incorporating the spatial aspects of early sound into this approach is also an interesting topic for further studies.

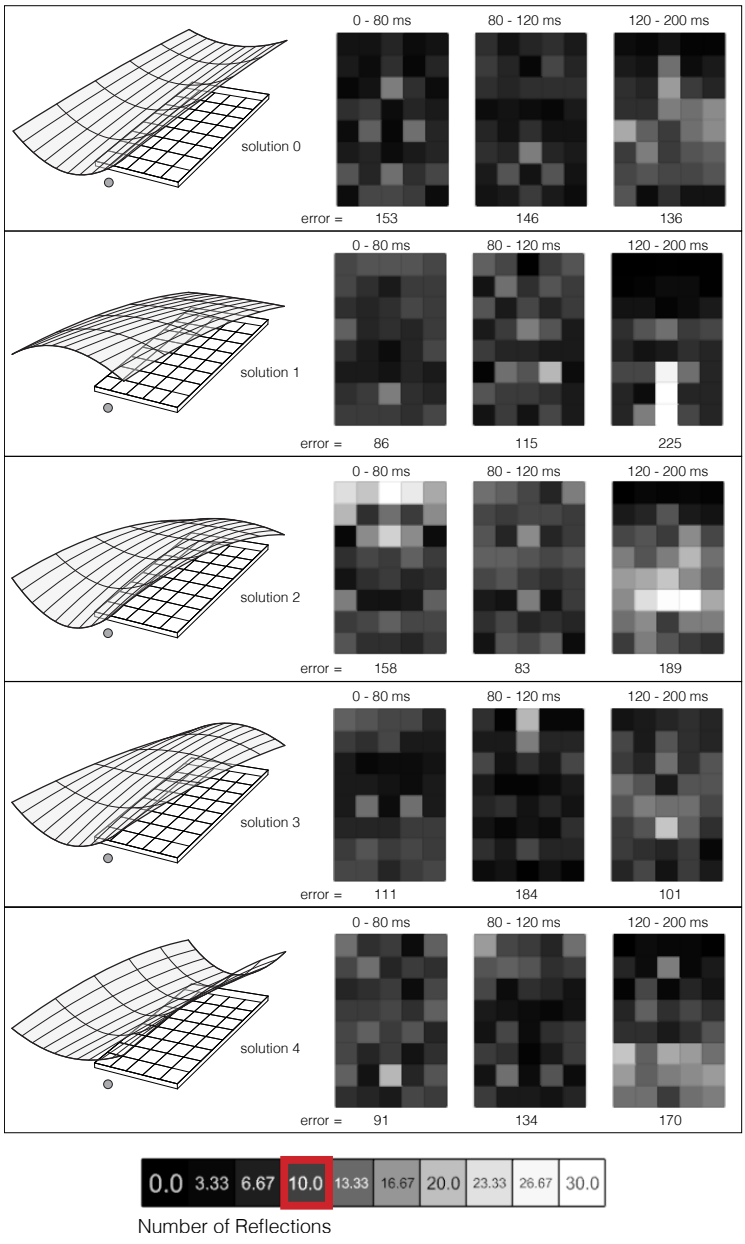


Figure 15.3: Distribution of reflected rays inside time-windows of a Pareto set of solutions for a Shoebox room with a reflective curved ceiling - Generated Curved Ceiling Shapes.

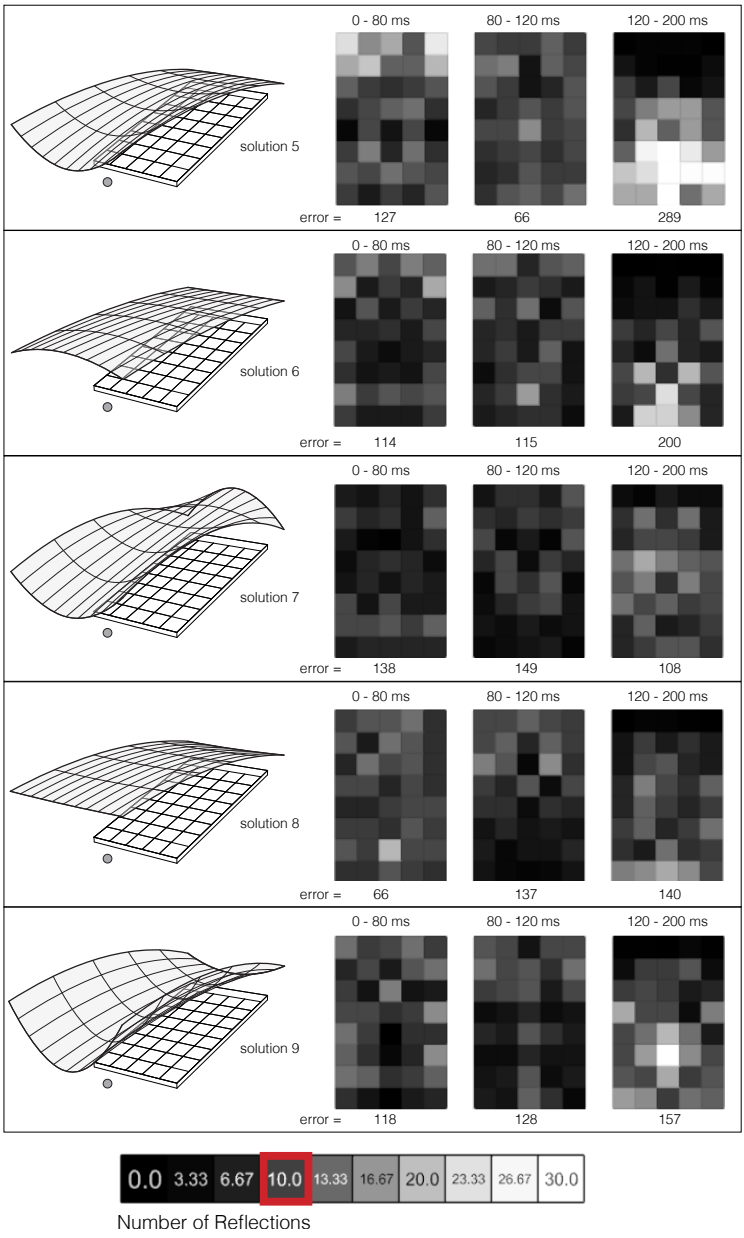


Figure 15.4: Distribution of reflected rays inside time-windows of a Pareto set of solutions for a Shoebox room with a reflective curved ceiling - Generated Curved Ceiling Shapes.

16

Energy design of building shape and envelope

It is now generally accepted that end cost, energy efficiency and general performance of buildings are strongly determined in the early stages of design (Miles et al. 2001, Wang et al. 2006, Turrin et al. 2011). It is therefore necessary for designers to be able to gather pertinent building performance information in this stage of the design process.

In order for designers to gather pertinent information for the early stages of design, a good comprehension of this phase is necessary. In this phase, design decisions have consequences for different aspects of the building, for example, structural integrity, indoor environmental quality, energy efficiency and costs. Many of these aspects are often in contrast with one another. A good example of these contrasts is natural illumination vs. solar shading in warm climates. Pertinent information in the early phase of design therefore needs to include multiple disciplines and deal with contrasting objectives.

A buildings general shape, fenestration, orientation and implantation on the building site are some of the first and most critical decisions made by architects in the design process. They have far reaching consequences and should be taken with a great many variables in mind. Most importantly for this work, a buildings shape, orientation and fenestration will greatly determine its exposure to the sun, and therefore have a great incidence in its indoor environmental quality and energy efficiency.

The building envelope is perhaps one of the most interesting subjects when we think about multidisciplinary design. Envelopes are responsible for most of the building's exposure to the elements, carrying with this a

good part of the responsibility for indoor environmental quality and energy efficiency. Envelopes are also regularly an important component it the buildings structure and are big part of their budget.

The energy related applications in this PhD thesis use search algorithms and parametric models to study the shape, orientation and envelope of an office building, in order to generate energy efficient solutions and information that will help designers make sound decisions in the early stages of design.

Fanger defines thermal comfort as a state of mind in which occupants desire no modifications to the air temperature of a room (Fanger 1970). Moreover, when humans are in an optimal temperature range, they do not need the use of their body's thermal control systems, they do not sweat or shiver. The thermal sensation we perceive is most importantly determined by six factors: Air temperature, mean radiant temperature, air velocity, humidity, metabolic rate and clothing. Out of these six factors metabolic rate and clothing do not depend on the building environment, but on the person. Air temperature, mean radiant temperature, air velocity and humidity depend greatly on external conditions, but also on building characteristics (Attia 2012).

Most countries have legislation and building codes that require fixed indoor air temperature ranges and relative humidity in order to guarantee the thermal comfort of their occupants. As previously mentioned, external conditions play a great role in these indoor characteristics, is therefore not always possible to maintain comfort ranges with the use of only passive building design. It is often the case in very common and populated climates to use indoor climate control systems such as air conditioning, heat radiators and humidity control systems. These systems consume a great deal of energy, and this consumption can be greatly reduced by efficient building design. In Italy, 45% of the total national energy consumption is due to building energy consumption, 83% of which is due to energy consumption during the buildings operation, the rest to their construction (Corrado & Paduos 2010).

In order to increase the energy efficiency of new buildings, we need to have accurate models of calculation the buildings future energy needs with a great deal of sensitivity on the buildings features. In this chapter we present the models used for the energy requirements calculations for heating, cooling and lighting of internal spaces.

16.1 Total Energy Requirements

The total energy requirements of a building's heating and cooling system can be summed up in the following equation:

$$Q_{tot} = Q_{H,nd} \cdot 1/e_H + Q_{C,nd} \cdot 1/e_C + Q_{E,nd} \cdot 1/e_E \tag{16.1}$$

where Q_{tot} is the total energy requirements for heating and cooling, $Q_{H,nd}$, $Q_{C,nd}$ and $Q_{E,nd}$ are the ideal energy requirement for heating, cooling and electricity respectively, and e_H, e_C and e_H are efficiency coefficient related to the heating, cooling electricity systems respectively.

Ideal energy refers to the amount of energy that is effectively needed to guarantee a given air temperature or a given luminance value. It is called ideal because it does not take into account the energy losses incurred from the primary energy source to the emission of this energy in the indoor environment. All building energy systems have energy losses in their procedures. Losses can be related to energy generation, distribution or emission. A perfect system would be one that emits the same amount of energy as the amount of primary energy it receivers.

This PhD thesis regards only the design of the building, this work does not consider building energy generation, distribution and emission systems, nor their efficiency in different building types. Therefore in this work we will only discuss ideal energy, the amount of energy required to achieve indoor environmental goals, mainly $Q_{H,nd}$, $Q_{C,nd}$ and $Q_{E,nd}$. We can thus rewrite the total energy requirement equation as follows:

$$Q_{tot} = Q_{H,nd} + Q_{C,nd} + Q_{E,nd} \tag{16.2}$$

16.2 Heating and Cooling Requirements calculation

The calculation of the required heating loads for any particular space involves many variables, and it is therefore a complex endeavor. Perhaps most significantly it involves the time factor, as the transmission of heat is subject to many time-related issues, such as thermal conduction, accumulation and release in building materials with big masses. Computational simulations of the heating and cooling loads are increasingly becoming of standard use in architectural practices. They have been incorporated into

many commercial CAD applications, either directly (as in the case of Autodesk Revit) or by means of plugins (as in the case of the Open Studio plugin for SketchUp). Attia and De Herde provide a review of 10 building energy performance tools in (Attia & De Herde 2011).

Among these computational energy calculation applications, Energy Plus has been signaled out by researchers for its accuracy (Attia & De Herde 2011). Energy plus is an application developed by the department of energy of the U.S. and is freely available.

> "EnergyPlus is an energy analysis and thermal load simulation program. Based on a users description of a building from the perspective of the buildings physical make-up, associated mechanical systems, etc., EnergyPlus will calculate the heating and cooling loads necessary to maintain thermal control set points."

(US Department of Energy 2013)

The Energy plus documentation lists among the many capabilities of the software:

- Sub-hourly, user-definable time steps for the interaction between the thermal zones and the environment; variable time steps for interactions between the thermal zones.
- Heat balance based solution technique for building thermal loads that allows for simultaneous calculation of radiant and convective effects at both in the interior and exterior surface during each time step.
- Transient heat conduction through building elements such as walls, roofs, floors, etc. using conduction transfer functions.

Figure 16.1 shows a diagram of the simulation method employed by energy plus. In it we can see how each thermal zone is studied. Thermal zones are subdivisions of the building model that should reflect not the internal partition of the building, but the thermal subdivisions of the building. Adjacent areas that are climatized at the same temperature should be added into a single thermal zone. Areas that are climatized at different temperatures or not climatized should be modeled as separate thermal zones.

The Air Heat Balance is calculated by a system of equations in which the unknowns are the superficial temperatures of all of the internal surfaces in each zone and the zone's air temperature. Several assumptions are made

in this calculation: Air and surface temperatures are perfectly uniform, surfaces are perfectly diffusive and internal air is transparent to thermal and solar radiation. The Air Heat Balance determines the heating and cooling loads for each thermal zone and, in turn, the entire building.

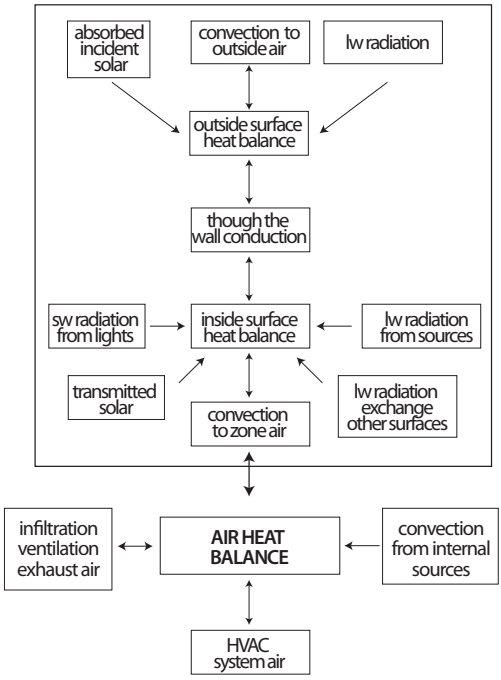


Figure 16.1: Energy Plus simulator diagram - Image taken from the software documentation.

In order for energy plus to calculate thermal loads, the building has to be described adequately. Buildings are described in three main areas:

- Their physical characteristics (geometry, construction, materials, orientation, etc.).
- Their HVAC system characteristics.
- Their functional characteristics (occupation hours, people activity level, internal electric equipment, etc.).

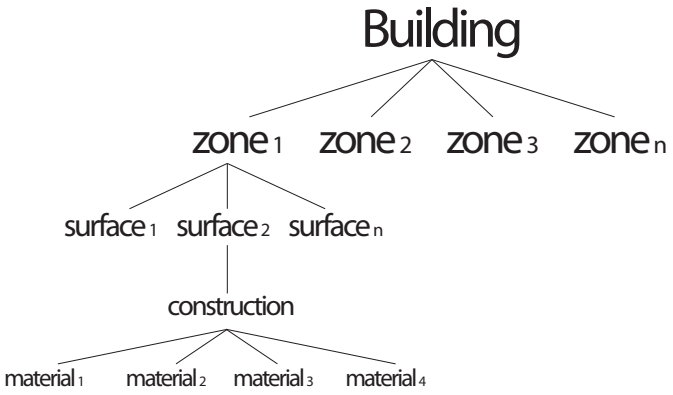


Figure 16.2: Energy Plus Building description diagram.

Figure 16.2 shows a diagram of the characterization of building components in energy plus (as well as the majority of energy simulations software). As previously mentioned, the building is subdivided into several thermal zones. Each zone is described in terms of its geometry, volume and internal surfaces. Each internal surface of the zone is in turn described in its adjacency to other thermal zones or to the external environment, and its construction; the way the surface is materially composed. Surfaces in a zone can include internal and external walls, ceilings, floors, doors or windows. External surfaces can also be described, such as external shading devices or other buildings. The surface construction is determined by a stratigraphy of materials. Each material also has to be defined, its thickness (s), conductivity (λ), density (ρ) and specific heat (c_p) values are all selected.

The HVAC systems are described in terms of their winter and summer temperature set points, function calendar (time of day and days of the week in which it functions).

Functional characteristics of the building include the number of persons per unit area, their activity level, generation of carbon dioxide, etc. Electrical equipment is also described for each zone, in therms of their energy consumption in (W/m^2) and their radiation.

The other important factor to describe for computational energy simulation is the outdoors climate conditions. This is done via a "weather file" that contains external air temperatures, solar radiation and wind conditions

for the entire year.

16.3 Lighting Energy Requirements

There are many computational methods for determining the quantity, quality and distribution of natural light in spaces, depending on desired luminance levels, internal surfaces and external conditions. This kind of calculation can then be used to determine the level of visual comfort of the building occupants. This kind of analysis is not included in this work, the present PhD thesis is interested in the energy requirements of lighting fixtures for internal spaces. Therefore, the object of this section is to describe how energy simulation software can determine illuminance levels in a few indoor reference points, and the amount of electrical energy required to compensate when natural light is not enough to meet a given level. Energy plus was also employed for this analysis.

> "The EnergyPlus daylighting model, in conjunction with the thermal analysis, determines the energy impact of daylighting strategies based on analysis of daylight availability, site conditions, window management in response to solar gain and glare, and various lighting control strategies."

(US Department of Energy 2013)

The documentation provided with energy plus describes three main steps in the daylighting calculation:

- "Daylight factors, which are ratios of interior illuminance or luminance to exterior horizontal illuminance, are calculated and stored. The user specifies the coordinates of one or two reference points in each daylit zone. EnergyPlus then integrates over the area of each exterior window in the zone to obtain the contribution of direct light from the window to the illuminance at the reference points, and the contribution of light that reflects from the walls, floor and ceiling before reaching the reference points. Window luminance and window background luminance, which are used to determine glare, are also calculated. Taken into account are such factors as sky luminance distribution, window size and orientation, glazing transmittance,

inside surface reflectances, sun control devices such as movable window shades, and external obstructions. Dividing daylight illuminance or luminance by exterior illuminance yields daylight factors. These factors are calculated for the hourly sun positions on sun-paths for representative days of the run period. To avoid the spikes of daylight and glare factors calculated during some sunrise and/or sunset hours when exterior horizontal illuminance is very low, the daylight and glare factors for those hours are reset to 0."

- "A daylighting calculation is performed each heat-balance time step when the sun is up. In this calculation the illuminance at the reference points in each zone is found by interpolating the stored daylight factors using the current time steps sun position and sky condition, then multiplying by the exterior horizontal illuminance. If glare control has been specified, the program will automatically deploy window shading, if available, to decrease glare below a specified comfort level. A similar option uses window shades to automatically control solar gain."
- "The electric lighting control system is simulated to determine the lighting energy needed to make up the difference between the daylighting illuminance level and the design illuminance. Finally, the zone lighting electric reduction factor is passed to the thermal calculation, which uses this factor to reduce the heat gain from lights."

(US Department of Energy 2013)

In order for energy plus to calculate the electric energy requirements for the artificial illumination of internal spaces, additional input is required. All of the geometrical and material considerations described above (like glass material properties) are considered in the daylighting calculation, but additionally, illuminance set-points in lux need to be specified for each zone reference point. Coordinates of the reference points, as well as their relative importance in the zone are also specified.

16.4 Climate Zones

> "Although now over 100 years old, the classification of climate originally formulated by Wladimir Köppen and modified by his collaborators and successors, is still in widespread use. It is widely used in teaching school and undergraduate courses on climate. It is also still in regular use by researchers across a range of disciplines as a basis for climatic regionalization of variables and for assessing the output of global climate models."

(Peel et al. 2007)

The Köppen - Geiger climate classification system is used in this PhD research to identify the most prominent climates in Europe. It is an attempt to select 4 climates that will represent the vast majority of the european continent that will then be used to study energy efficient buildings through multi-objective search algorithms. Table 16.1 shows the classification system, how it divides climates and the criteria for them to appertain to all categories.

Figure 16.3 shows the Köppen - Geiger climate map of europe. By looking at it we can see that the vast majority of the continents surface is covered by C and D climate types. Southernmost regions of Europe are described as Csa, Cfb and Cfa climates cover most of the mid-latitude European regions, and many of the most northern and eastern regions are described as Dfb climates.

For this PhD research, four cities were chosen to represent these climates. The Italian city of Plalermo is selected as a Temperate dry-hot summer (Csa), Torino in northern Italy is chosen as a Temperate wet-hot summer (Cfa), Frankfurt Germany as a Temperate wet-warm summer (Cfb) and Oslo Norway as a Cold wet-warm summer (Dfb). Figure 16.4 shows Tregenza Sky domes for all cities, showing solar radiation directions and energy in kWh/m^2. Table 16.2 shows the average monthly dry bulb temperatures (C°) in these locations.

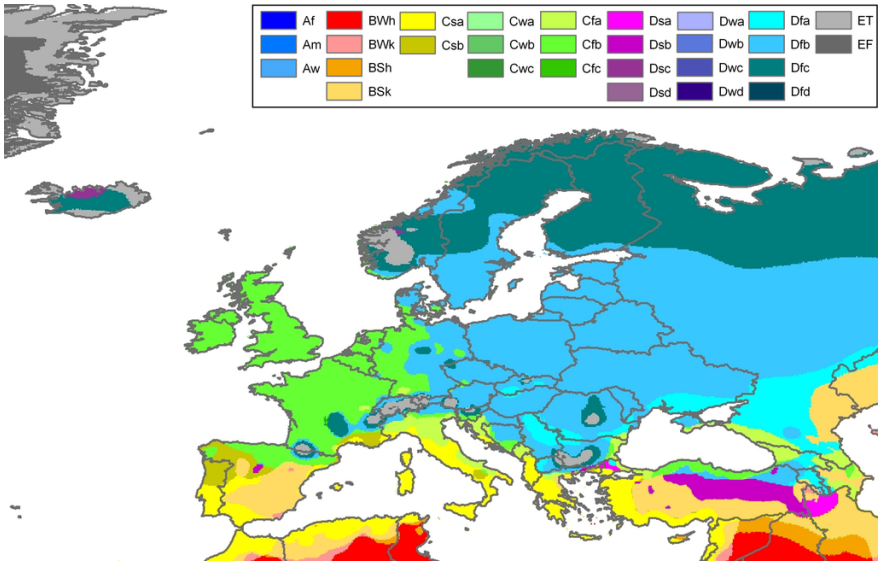


Figure 16.3: Köppen - Geiger Climate type map of Europe.

Table 16.1: Description of Köppen - Geiger climate symbols and defining criteria - Taken from (Peel et al. 2007).

1_{st}	2_{nd}	3_{rd}	Description	Criteria
A			**Tropical**	$T_{cold} \geq 18^{\circ}$
	f		-Rainforest	$P_{dry} \geq 60^{\circ}$
	m		-Monsoon	Not(A_f) & $P_{dry} \geq$ 100-MAP/25
	w		-Savannah	Not(A_f) & $P_{dry} <$ 100-MAP/25
B			**Arid**	MAP $< 10\times P_{threshold}$
	W		-Desert	MAP $< 5 \times P_{threshold}$
	S		-Steppe	MAP $\geq 5 \times P_{threshold}$
		h	-Hot	MAT $\geq 18^{\circ}$
		k	-Cold	MAT $< 18^{\circ}$
C			**Temperate**	$T_{hot} > 10^{\circ}$& $0 < T_{cold} < 18^{\circ}$
	s		-Dry Summer	$P_{sdry} < 40$ & $P_{sdry} < P_{wwet}/3$
	w		-Dry Winter	$P_{wdry} < P_{swet}/10$
	f		-Without dry season	Not (C_s) or (C_w)
		a	-Hot Summer	$T_{hot} \geq 22^{\circ}$
		b	-Warm Summer	Not (a) & $T_{mon10} \geq 4^{\circ}$
		c	-Cold Summer	Not (a or b) & $T_{mon10} < 4^{\circ}$
D			**Cold**	$T_{hot} >10^{\circ}$& $T_{cold} \leq 0^{\circ}$
	s		-Dry Summer	$P_{sdry} < 40$ & $P_{sdry} < P_{wwet}/3$
	w		-Dry Winter	$P_{wdry} < P_{swet}/10$
	f		-Without dry season	Not (D_s) or (D_w)
		a	-Hot Summer	$T_{hot} \geq 22^{\circ}$
		b	-Warm Summer	Not (a) & $T_{mon10} \geq 4^{\circ}$
		c	-Cold Summer	Not (a, b or d)
		d	-Very Cold Winter	Not (a or b) & $T_{cold} <$ -38°
E			**Polar**	$T_{hot} <10^{\circ}$
	T		-Tundra	$T_{hot} > 0$
	F		-Frost	$T_{hot} \leq 0$

MAP = mean annual precipitation, MAT = mean annual temperature, T_{hot} = temperature of the hottest month, T_{cold} = temperature of the coldest month, T_{mon10} = number of months where the temperature is above 10, P_{dry} = precipitation of the driest month, P_{sdry} = precipitation of the driest month in summer, P_{wdry} = precipitation of the driest month in winter, P_{swet} = precipitation of the wettest month in summer, P_{wwet} = precipitation of the wettest month in winter, $P_{threshold}$ = varies according to the following rules (if 70% of MAP occurs in winter then $P_{threshold}$ = 2 x MAT, if 70% of MAP occurs in summer then $P_{threshold}$ = 2 x MAT + 28, otherwise $P_{threshold}$ = 2 x MAT + 14). Summer (winter) is defined as the warmer (cooler) six month period of ONDJFM and AMJJAS.

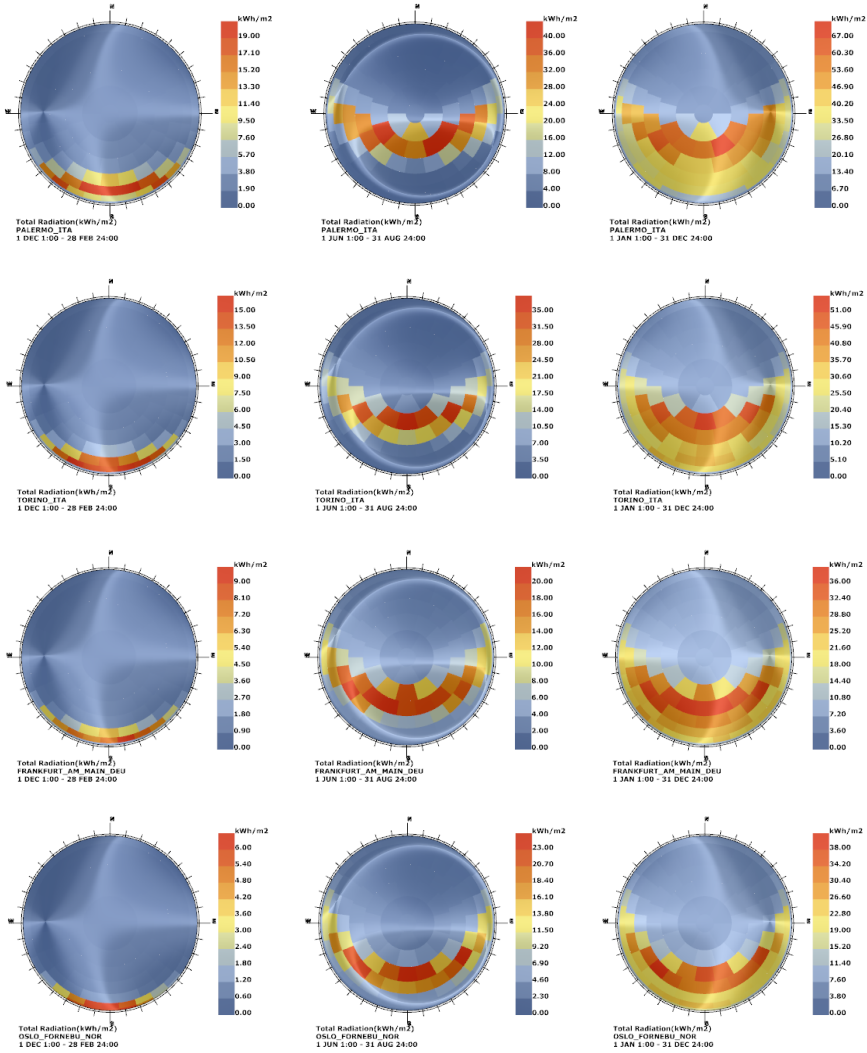


Figure 16.4: Solar radiation Tregenza Sky Dome diagrams in winter, summer and whole year for Palermo, Torino, Frankfurt and Oslo

Table 16.2: Average monthly dry bulb temperature (C°) for Palermo, Torino, Frankfurt and Oslo.

Month	Palermo	Torino	Frankfurt	Oslo
Jan	12.6	1.8	2.3	-3.7
Feb	11.8	3.8	1.7	-0.8
Mar	13.8	8.1	5.5	0.9
Apr	15.6	11.8	9.2	4.6
May	19.1	16.0	14.7	11.9
Jun	22.8	19.5	16.4	14.7
Jul	25.5	23.0	19.5	17.5
Aug	27.0	21.9	18.6	16.5
Sep	24.1	18.1	14.9	11.0
Oct	21.6	12.3	10.6	6.7
Nov	17.2	6.3	4.7	1.7
Dec	13.9	2.6	1.7	-1.6

17

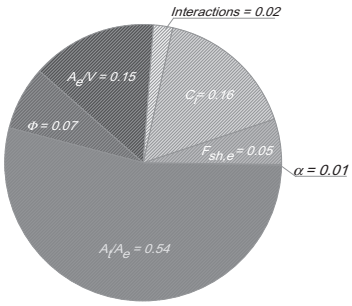
The Building Shape and Orientation

Mechri et al. (Mechri et al. 2010) performed a sensitivity analysis with the objective to determine the building variables that have the biggest incidence in heating and cooling energy needs in several italian climates. The variables they studied were the compactness ratio (Area of the envelope to volume ratio (A_e/V), envelope transparent to opaque ratio(A_t/A_e), absorptance or external color (α), building orientation (Φ), external shading ($F_{sh,e}$) and internal effective heat capacity (C_i). Figure 17.1 shows the decomposition of the total variance of the energy needs for heating and cooling for Palermo and Cuneo. The study found that the variable that had by far the most influence in the energy requirements was the transparent to opaque envelope ratio (A_t/A_e) both in cooling and heating and in all climates. Envelope to volume ratio (A_e/V) and internal heat capacity (C_i) are also influential in a lesser degree. Surprisingly building orientation did not have a big influence for the orientation ranges studied by Mechri et al.

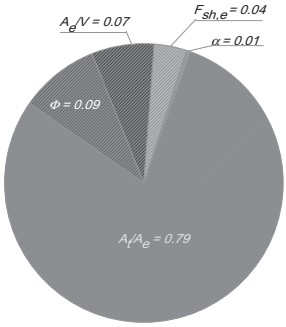
If we look at these variables in terms of early design stage, A_e/V, A_t/A_e and Φ are some of the first decisions to be made in the design process. A_e/V can be defined as the building shape since perimeter to area ratios vary with plan shape. A_t/A_e is the combination of the building shape and the overall fenestration scheme. Building orientation in most cases is very much related to the building site, but in the cases where sites allow for different implantations, this is commonly defined in the early stages as well.

This PhD thesis looks into these early stage design variables in energy efficiency. In this chapter the building shape and orientation are studied.

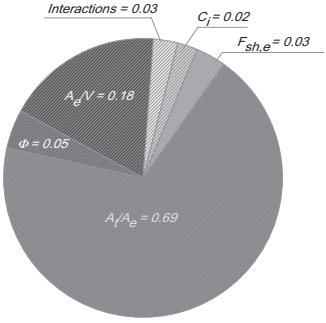
Kämph and Robinson studied the overall shape of the building in an urban context by means of evolutionary algorithm, looking to maximize solar



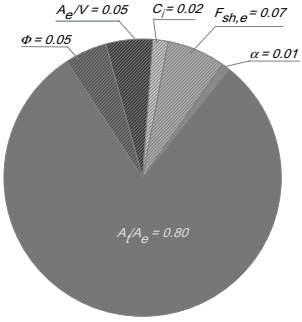
(a) Decomposition of the total variance of the heating energy needs for Palermo.



(b) Decomposition of the total variance of the cooling energy needs for Palermo.



(c) Decomposition of the total variance of the heating energy needs for Cuneo.



(d) Decomposition of the total variance of the cooling energy needs for Cuneo.

Figure 17.1: Decomposition of the total variance of the energy needs for cooling and heating for Cuneo and Palermo (Mechri et al. 2010).

irradiation (Kämpf & Robinson 2010). They investigated 3 different parametric models of different shapes and found "non-intuitive" geometries that had up to 20% increase in solar gains when compared to more traditional shapes.

Wang et al. studied the polygonal plan shape of a building by means of multi-objective genetic algorithm, with the purpose to minimize Life Cycle Costs (LCC) and Life Cycle Environmental Impact (LCEI) (Wang et al. 2006). These two contrasting functions were found to be well influenced by the shape of the polygonal building. Buildings that had low LCC values had more regular polygonal shapes, while LCEI low values were found when buildings had longer south-facing facades.

17.1 Case Study 8: Building Shape and Orientation

Case study 8 describes the use of multi-objective search algorithms to determine energy efficient rectangular building proportions and orientations. Buildings in this case study will be studied in terms of their heating, cooling and lighting energy needs for 4 european climates. Palermo, Torino, Frankfurt and Oslo were chosen to represent a high percentage of all of the climates present in the european continent.

The objective of this case study is to determine optimal orientation and building width/length ratios (w/l). The search process will examine combinations of orientations and (w/l) to minimize building energy consumption. Since only orientation and (w/l) are being examined, all other parameters will be kept fixed, most importantly, the A_t/A_e ratio. Masonry building envelopes are studied in this PhD thesis. As was the case for the structural case study, brick facades and their openings are the objective of this series of energy efficiency search process. In this case however, these values will be kept fixed. A fixed A_t/A_e value of 45 % is used, and a fixed 50cm thickness is used for all orientations.

The are no external shading devises, but the building envelope thickness itself is used as a shading surface, since windows are position at the internal edge of the envelope. We can make an analogy with traditional external shading devices, the lateral surfaces of the wall opening can be considered as shading fins, and the superior surfaces as an overhang.

17.1.1 Case Study Building

Energy requirements are studied in a case study building, or more accurately, on one of the building's floors. The building has the same characteristics of the building described in chapter 11. It is a 6 story high office building with a rectangular plan. The floor studied is a standard floor just above the ground floor. The floor has a fixed area of 280 m^2 and a height of 4m.

The building floor that is being studied is modeled by means of a singe thermal zone. In order to properly model the building as an office building, the following functional characteristics were given to Energy Plus:

- People activity level: 13.8 W/m^2
- Electric equipment: 6.454 W/m^2
- The ventilation rate was set to 1.7 air changes per hour during weekdays from 8.00 AM to 9.00 PM and to 0.25 h^{-1} during the rest of the day and during weekends.
- The heating and cooling set point temperatures were respectively set to 20 °C and 26 °C. The systems were active from 7.00 AM to 9.00 PM during weekdays only.
- Lighting control was performed with two control points and dimmed control option. For glare control, the occupants' seats were placed facing north. The maximum lighting level was set to 10 W/m^2 per zone floor area. The lighting schedule was set equal to the occupancy one.
- The solar absorption coefficient of the external opaque surfaces was set to 0.6, which corresponds to a medium color.

The occupancy schedule of the building during weekdays is shown in figure 17.2. The building was assumed to be unoccupied during the weekends.

17.1.2 Building envelope materials

We previously mentioned the way in which the building to be studied is detailed for Energy Plus, its materials, constructions and thermal characteristics. The materials and constructions used in case study presented in this chapter are described bellow.

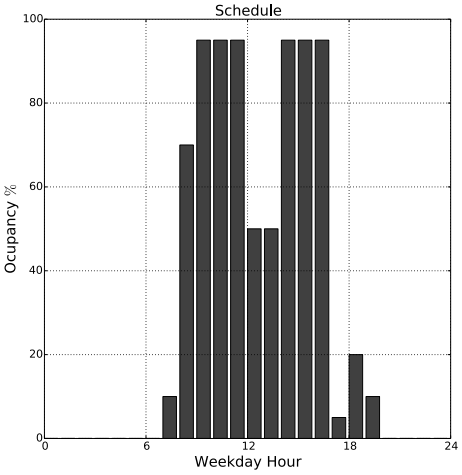


Figure 17.2: Office building occupancy schedule for weekdays.

Windows

Since case study 8 focuses on building orientation and proportions, window materials are kept fixed. There is one single window construction given to Energy plus to describe the windows. The construction used in described in the following table:

Composition mm	Position low-e coating	U_g $W/(m^2K)$	g_g -	τ_1 -
4glass; 12air ; 4glass	-	2.68	0.77	0.81

A double glazed window construction is used, with 4mm glasses and a 12 mm air gap.

Walls and Slabs

Wall constructions for case study 8 are the same for all orientations. Ceiling and floor slabs are modeled as adiabatic surfaces in order to consider only the building envelope as a design variable. They are however described in constructions on their own in order to consider their influence on the

study floor. Wall and slab materials and constructions for case study 8 are described in tables 17.1 and 17.2 respectively. EPS thickness was explicitly calculated to achieve a U-value of 0.33 W/(m^{2}K).

The floor of the office building in this case study is an open space plan. This choice was made to place greater emphasis on the building envelope. There are no other internal surfaces that can create shaded areas or add internal mass to the space.

Table 17.1: Characteristics of materials for case study 8.

Material	s	λ	ρ	c_p
	m	W/(mK)	kg/m^3	J/(kgK)
External gypsum	0.02	0.9	1800	840
EPS	0.07	0.031	112.1	1450
Bricks	0.5	0.5	1600	840
Internal gypsum	0.01	0.7	1400	840
Floor slab	0.25	0.678	1280	1000
Floor tiles	0.02	2.69	2700	984
Air gap	0.13	R:	0.18	m^{2}K/W

Table 17.2: Wall and slab Constructions for case study 8.

Component	Layer 1	Layer 2	Layer 3	Layer 4
Masonry wall	External gypsum	EPS	Bricks	Internal gypsum
Floor	Internal gypsum	Floor slab	Air gap	Floor tiles
Ceiling	Floor tiles	Air gap	Floor slab	Internal gypsum

17.1.3 Parametric Model

The parametric model used for case study 8 has only two variables. The first variable x_1 refers to the buildings width (w). The building's length is calculates from its width and the fixed floor area A_p of 280 m^2. Following this logic, all possible solutions have the same floor area A_p and the same internal volume. Hence, these two dimensions do not influence the results

of the search process in any way. In this sense length is also varies during the search process, but it is determined as $l = A_p/w$.

The second variable x_2 is the angle of rotation that determines the buildings orientation. The building is allowed to rotate 45°in a clockwise direction and 45°in a counterclockwise direction. The parametric model for case study 8 is shown in figure 17.3.

It is important to notice that the 50 cm masonry walls create shading overhangs and fins on the windows that are on the inside of the wall. Since these windows are long and not very high, the overhangs are in a better position to shade the windows, especially during the summer moths.

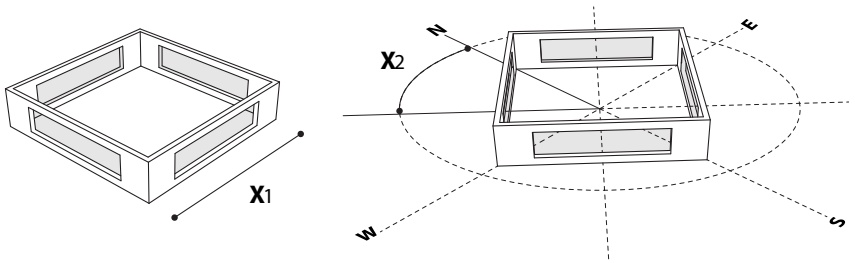


Figure 17.3: Parametric Model for Case Study 8.

17.1.4 Fitness functions

Case study 8 is a search process for energy efficiency considering heating, cooling and lighting (electric) energy requirements. It can be argued that in the end these energy requirements will all be translated into one single energy requirement value, much in the way that is shown in equation 16.2. However, these energy values would have to be calculated in terms of primary energy in order to be added together. Heating and cooling needs are estimated in therms of thermal loads, while lighting energy needs are calculated as electric energy, hence they cannot be added as they are. We would need to calculate them in terms of primary energy, taking into account the influence of the heating, cooling and lighting systems.

Different systems would produce different results in calculating primary energy. Thus, selecting an energy system for our case study would greatly influence the envelopes studied during the search process. Energy systems are not a part of this thesis' scope, and in addition, it is the aim of this

chapter to study the building's shape on its own. For this reason, it is best to keep heating, cooling and lighting needs in separate fitness functions.

Case study 8 can then be described by the following set of equations:

$$\textit{Case Study 8} \begin{cases} \textit{Minimize} & f_{1(x)} = Q_{H,nd}, \\ \textit{Minimize} & f_{2(x)} = Q_{C,nd}, \\ \textit{Minimize} & f_{3(x)} = Q_{E,nd}, \\ \textit{subject to} & 14 \leq x_1 \leq 20. \\ & -45 \leq x_2 \leq 45. \end{cases} \tag{17.1}$$

17.1.5 Genetic algorithm inputs

NSGA-II explores 50 generations with 50 individuals in each generation. The overall genetic inputs for this case study is as follows:

Case Study 8		
Population Size (N)	50	
Number of Variables	2	
Number of binary digits	8	
Variable Domains	$x_1 \in [14, 20]$	$x_2 \in [-45, 45]$
Mutation Probability (p_m)	0.2	
End Condition	End after 50 generations	

17.1.6 Results

Case study 8 has a large amount of results, there are 4 different climates and 3 fitness functions. This means that there are 3 two-dimensional Pareto fronts for each climate. We will start by looking at a general overview of all of the results. Figures 17.4, 17.5, 17.6 and 17.7 show the objective spaces for Palermo, Torino, Frankfurt and Oslo respectively.

The most striking result from this study is the fact that orientation and room proportions in their own do not have a large influence in energy needs. The largest change in heating needs from a single climate comes from Oslo, and it barely reaches 2 kWh_t/m^2 a year. The largest variation in cooling energy need is seen in Palermo (3.5 kWh_t/m^2 a year), and for lighting it comes from Oslo (0.14 kWh_t/m^2 a year). These are very low variations, especially when we consider how different the orientations and building proportions are. These results support the findings of Mechri et al. (Mechri et al. 2010), building orientation and proportions are not very influential on their own. If we were to vary the openings in these buildings

together with orientation and proportions, then we would expect to find much bigger relevances.

As is to be expected, heating needs are highest for Oslo and Frankfurt, while cooling needs are highest for Palermo. Lighting energy needs are quite similar for all climates, with a slight increase in need for Oslo.

Palermo

Heating needs for the office building studied in the Palermo climate are very small, they are negligible when compared to cooling needs. However, it is interesting to note that results for the Palermo study show very little contrast between heating and cooling needs. The best performing buildings for heating and cooling functions (A and B respectively) are quite similar, with only a slight variation in orientation angle. Both buildings A and B are 14×20 rectangles, and they are both almost perfectly oriented with the long facades due north-south. This orientation makes good sense for heating needs since it exposes the largest windows due south, and in so doing maximizes the solar gains for the winter. It is not so clear why this orientation is the best one for summer cooling needs (solution B). It can be argued that the solar paths for Palermo are high enough in the horizon, that the best way for the building to shield itself, is to depend on the shading overhangs created by the thick masonry walls.

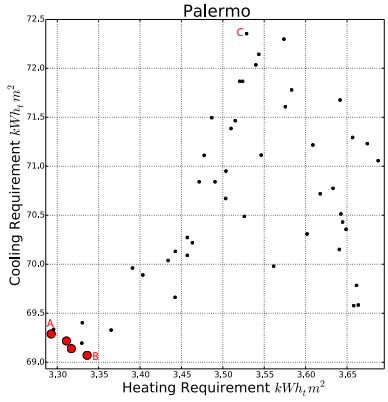
Lighting needs for the Palermo climate are best met by solution C, it is a 14×20 rectangle that is oriented in such a way as to expose its long facades due east-west. This orientation exposes the most sunlight to the light sensors in the model.

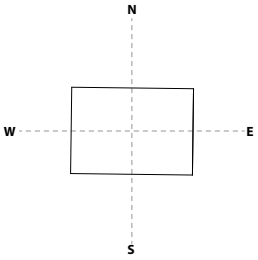
Torino

Heating and cooling energy needs for the Torino climate are almost of the same magnitude. Meaning that the office building in Torino needs to be optimized for both of them equally. When compared to the Palermo results, in Torino we see a great deal of contrast in the heating and cooling needs. Interestingly, the Pareto front for heating and cooling needs is convex until one point when it becomes concave, showing quite different levels of contrast.

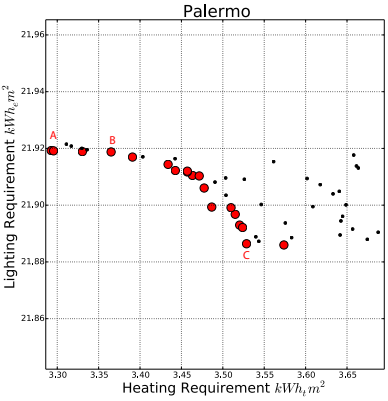
Solution A is the best performing one for heating in Torino. It is a 14×20 rectangle north-south exposed. This is congruent with the results obtained in Palermo, long south exposed windows increase winter solar gains.

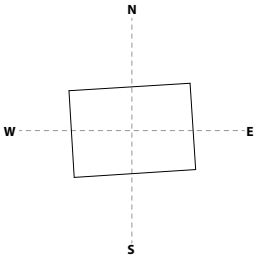
Solution B is the best performing solution for cooling needs. This solution is quite different from the previous ones, it is a 16.7×16.7 square



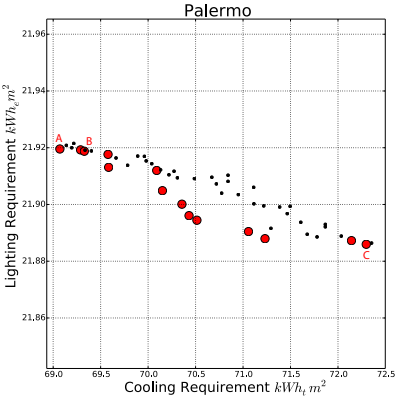


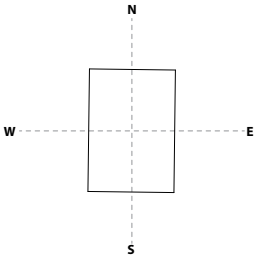
Palermo A





Palermo B





Palermo C

Figure 17.4: Objective spaces for Case Study 8 for Palermo.

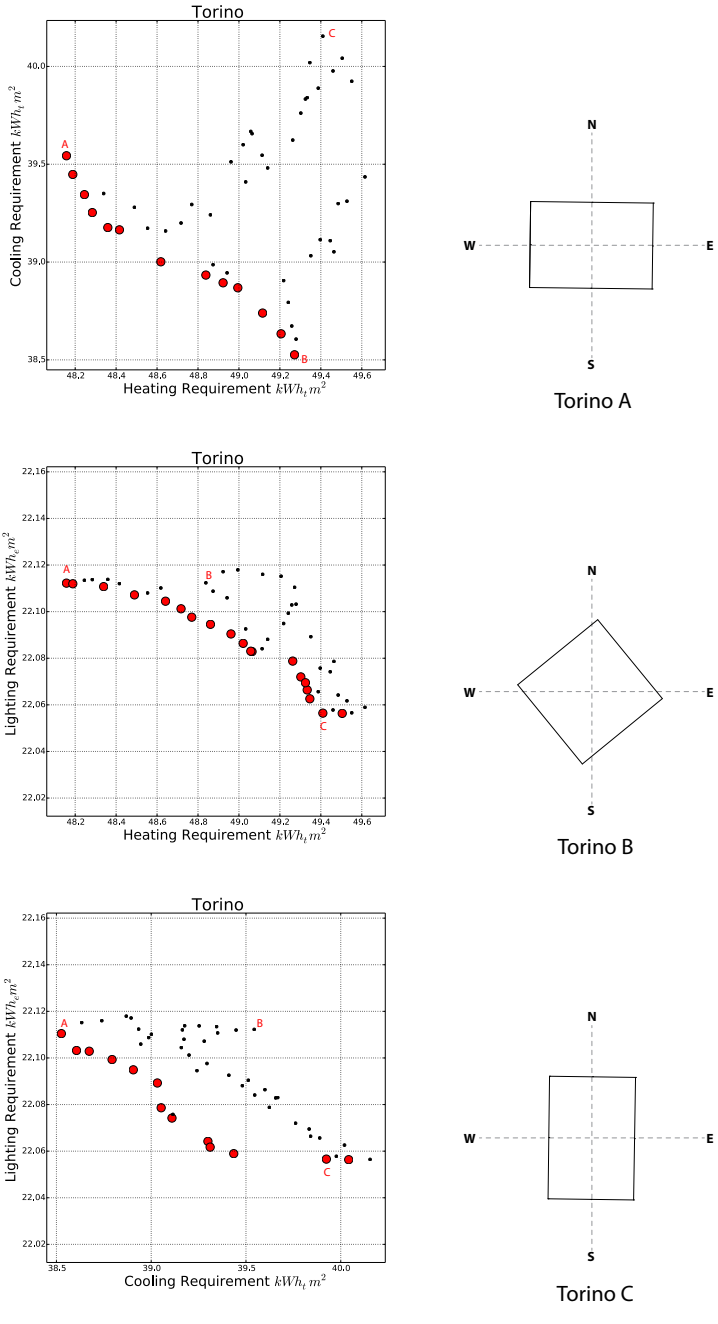


Figure 17.5: Objective spaces for Case Study 8 for Torino.

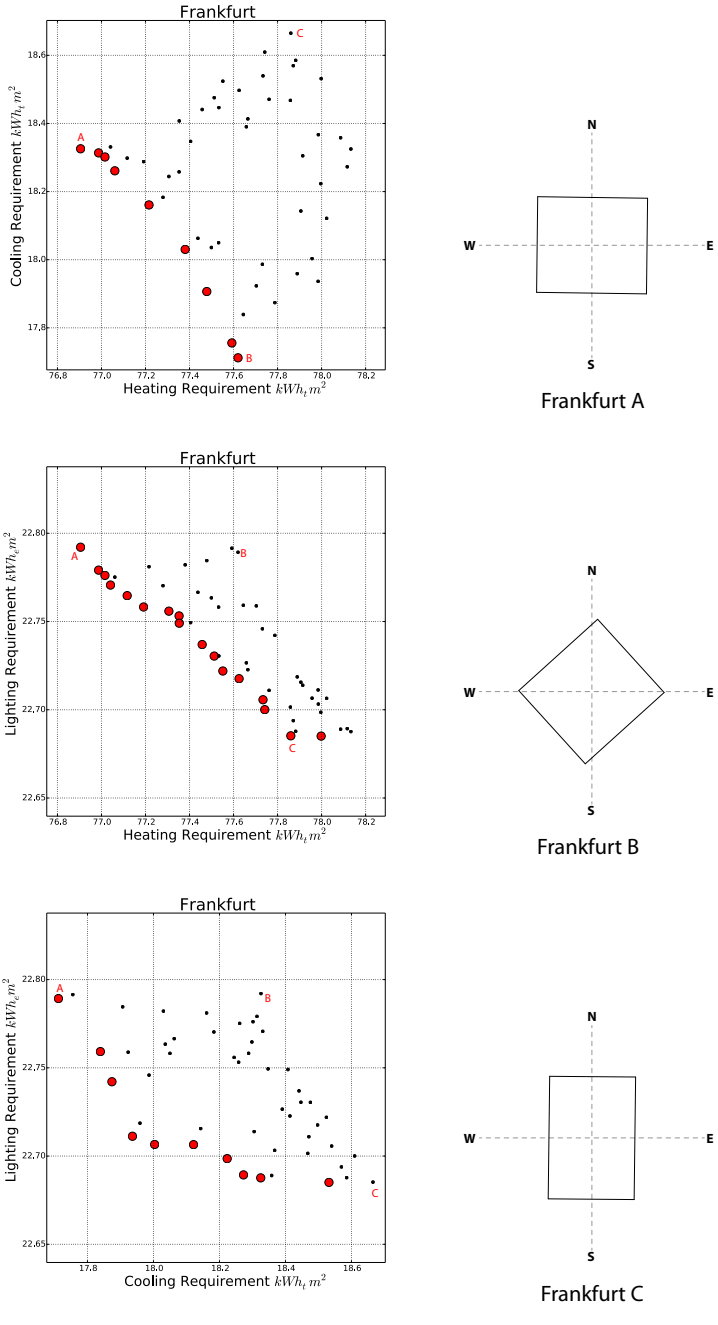


Figure 17.6: Objective spaces for Case Study 8 for Frankfurt.

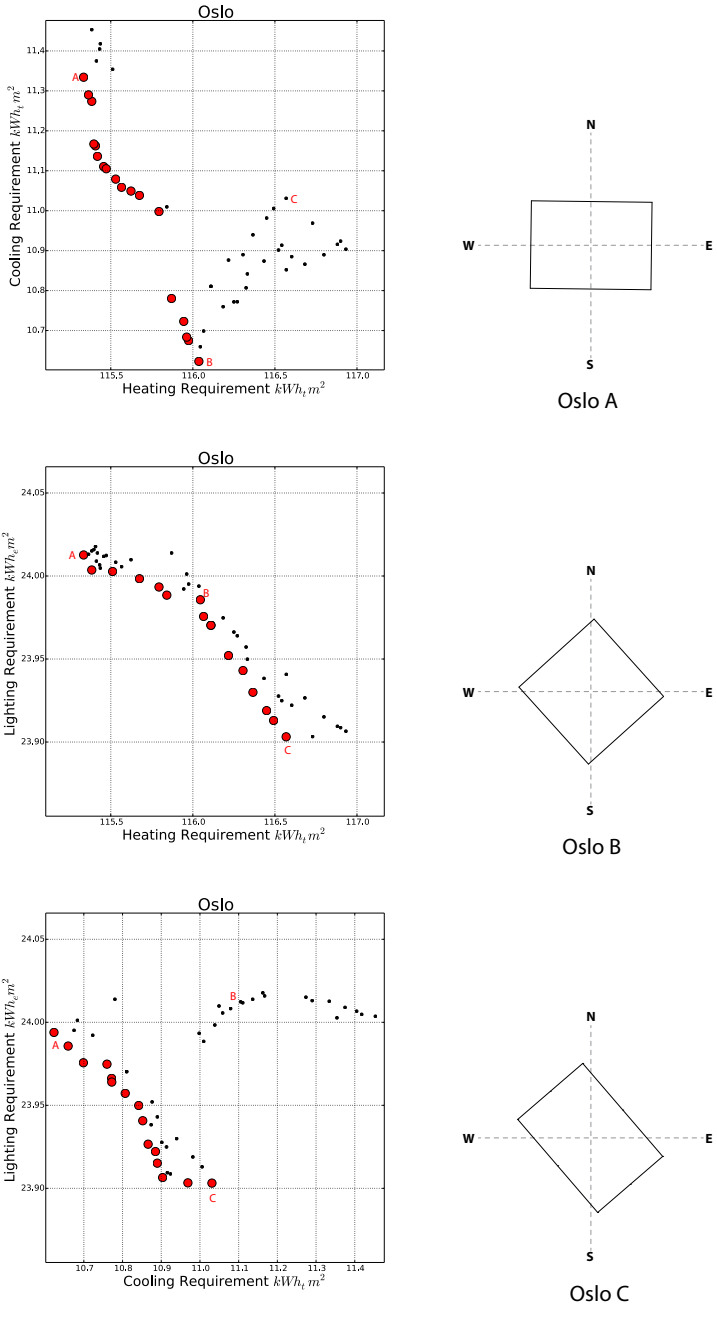


Figure 17.7: Objective spaces for Case Study 8 for Oslo.

that is oriented at an almost 45°angle from north. This solution is not very high performing in the Palermo study, but similar solutions will continue to appear in Frankfurt and Oslo. It is likely that sun paths for these cities in the summer are not high enough for the overhangs to shade during the summer. When this occurs, the only way the MOGA can find to shade internal spaces is to rotate the building, and place the opaque walls in the corners towards the east, west and south orientations. Another possible explanation for this 45°orientation is that the GA is trying at all costs to avoid having windows in the east and west orientations. The square form tends to make windows have better fin shading as well. These two possibilities are in contrast with one another, further investigation into this solution is required to fully understand the reason behind this orientation.

As was also the case for the Palermo study, the best performing solution for lighting needs (solution C) is a east-west exposed 14 × 20 rectangle. This solution is one of the worst performing solutions for heating, and the worst for cooling.

Frankfurt

In Frankfurt we see heating needs that are significantly larger that the cooling needs. Heating and cooling needs in this climate are also quite contrasted, the Pareto front having a concave shape.

Solution A is the best performing one for heating energy needs. It is a rectangle north-south exposed, but it is almost a square. This is to be expected as latitudes rise, sun paths become lower but they also reach farther into east and west that in lower latitudes. We can therefore assume that in oder to increase solar gains in Frankfurt, the MOGA is exploiting the east and west exposures, and not just southern one.

As was seen in the Torino climate, the best performing solution for cooling needs (solution B) is an oblique square. The best performing lighting solution is an east-west exposed rectangle (solution C).

Oslo

Oslo has the largest heating energy requirements out of all of the climates studied in this PhD thesis. Cooling requirements for Oslo are small but not negligible. Heating and cooling needs in this climate are also highly contrasted. Also in this case, the rectangular north-south exposed solution is the best one for heating (solution A) and the oblique square for cooling (solution B). Interestingly, for the Oslo latitude, the best performing solution

for lighting is not north-south exposed. It is a 14×20 rectangle oriented at 45°from north. It appears that light is more efficiently introduced into the building when there is a long facade facing southwest. Looking at the sun radiation for Oslo, this makes sense.

18

The Building Envelope

The sensitivity analysis of energy efficiency in office buildings published by Mechri et al. shows a very strong relevance of the envelope transparent to opaque ratio A_t/A_e and a much lower importance of the building orientation and proportions (Mechri et al. 2010). This finding was supported by the previous case study. In this chapter we will investigate the A_t/A_e ratio by means of multi-objective search algorithms. This study involves the masonry envelope thickness and the window materials and construction.

Goia et al. studied optimal A_t/A_e ratios for 4 european climates, Rome, Athens, Frankfurt and Oslo. They performed a parametric analysis, increasing the A_t/A_e ratio step by step, and calculating a total energy requirement E_{tot} (Goia et al. 2012). They performed this study in all four locations, in North, South, East and West orientations separately.

> "The results clearly show that each climate requires a dedicated optimized solution, being the minimum value of E_{tot} reached for different window to wall ratio, depending on the climate. The south exposed façade module is the one that has the highest variation, if located in a cold or in a hot climate. Furthermore, except the coldest climate, west and east-exposed façade modules are always those with the highest energy need."

(Goia et al. 2012)

Wright and Mourshed used genetic algorithms to optimize fenestration configurations for a large atrium located in the city of Chicago (USA)

(Wright & Mourshed 2009). The objective of their study is that of minimizing E_{tot} by using a window cell parametric model. This model is the same described in chapter 11.2.1. Also in this case, single orientations are studied at a time. The results of the study are described by the authors as follows:

> "Given that, each optimization experiment resulted in a different distribution of window cells, but that the optimized energy and window area was of the same order of magnitude in each case, it is concluded that, for the example building studied here, the position of the window cells has only a "second-order" effect on energy use. However, in the results from all experiments, the optimized position of the windows cells was biased towards the top-west corner of the façade. Locating the windows towards the top of the façade results in the penetration of daylight to a greater depth in the atrium; correspondingly, this reduces the energy use from artificial lighting, particularly when the windows are positioned towards the top-west quadrant of the façade. The position of the windows also has an impact on the distribution of the beam solar radiation on the internal surfaces of the atrium, which in turn affects heat loads through the different heat loss and storage effects of the various construction elements."

(Wright & Mourshed 2009)

The studies presented in this PhD thesis do not consider orientations separately, the search process is conducted with four orientations simultaneously. There are two important reasons for this:

- There is no reason to believe that by combining 4 optimized façades for North, South, East and West orientations into a single building would result in an optimal energy efficiency design. It is only by means of a search process that considers all façade simultaneously that we can be sure to obtain an optimal building design.
- During the early stages of design, architects are most likely to consider the buildings shape, fenestration and orientation as a whole, and not consider them in separate detailed orientations.

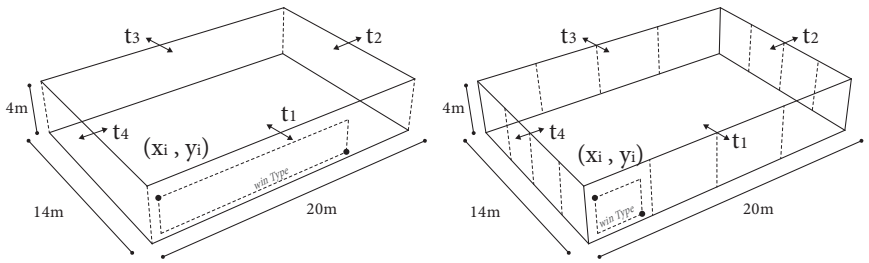
The efforts presented in this chapter consider the 4 orientations of an office building simultaneously.

18.1 Case Study 9: Masonry building envelope - Sub-urban context office building

The first case study in this chapter refers to the same office building used in case study 8. A rectangular office building in a sub-urban context with a masonry envelope. The important differences in this case study lie mostly in the parametric model, the problem variables. This case study keeps building size fixed at 20×14 m and the envelope thickness is variable. The buildings are studied in the same 4 climates as seen above: Palermo, Torino, Frankfurt and Oslo.

18.1.1 Parametric model

The parametric model used in this chapter follows the same geometric rules than the one described in chapter 11. As was discussed in chapter 11, this parametric model is dependent on the selection of the number and configuration of window influence areas. Hence, in oder to cover a large part of the solution space, more that one parametric model and search processes are needed. Figure 18.1 shows two parametric models used in case study 9. The first model (a) uses one single window area that covers the entire length and hight of each façade. The second model (b) uses 4 window areas distributed along the length of each façade. In this configuration, each window area covers 1/4 of the length of the façade and its full height.



(a) Parametric Model 1Window

(b) Parametric Model 4Windows

Figure 18.1: Parametric Models for Case Study 9 - (a) Model with one window area per façade - (b) Model with four window areas per façade.

Along with the variable widow configurations, this parametric model also changes the wall thicknesses in each façade, as was previously seen in the structural case in chapter 11. In this case however, thickness variation is used not for structural capacity, but for increased shading on the windows from the masonry overhangs and fins. Increased external wall thickness has also an influence on the thermal transmittance U of the wall. It is the purpose of this case study to consider the influence of the thickness as a shading device, and not on the thermal transmittance of the wall, therefore, for the search processes in this PhD thesis a fixed U-value of 0.33 W/(m^{2}K) was chosen. In order to maintain a fixed U value while still varying the thickness of the brick walls, the thermal insulant EPS was employed with a variable thickness. The EPS thickness for each wall is calculated in the parametric model, in such a way as to have the required thickness to keep U at 0.33 W/(m^{2}K). Table 18.1 shows the materials used in this case study.

Table 18.1: Characteristics of materials for case study 9.

Material	s	λ	ρ	c_p
	m	W/(mK)	kg/m^3	J/(kgK)
External gypsum	0.02	0.9	1800	840
EPS	Variable	0.031	112.1	1450
Bricks	Variable	0.5	1600	840
Internal gypsum	0.01	0.7	1400	840
Floor slab	0.25	0.678	1280	1000
Floor tiles	0.02	2.69	2700	984
Air gap	0.13	R:	0.18	m^{2}K/W

An important difference between these models and the one presented in chapter 11 in the fact that the window construction is also variable. This is the first case in this PhD thesis in which material properties are the subject of the search. The parametric model used has the ability to select a window construction from the ones shown in table 18.2. The model selects one window construction from the table, and uses it for all of the windows in the building. As shown in table 18.2, window constructions vary in the number of glass panels, the presence and position of low emissive coating, thermal transmittance etc.

The number of variables in this case study is a high one. There are 4 thicknesses and one window construction plus 4 variables for each window

Table 18.2: Window constructions for case study 9.

Number	Composition mm	Position low-e coating	U_g $W/(m^2K)$	g_g -	τ_1 -
0	$4g; 12air; 4g$	-	2.68	0.77	0.81
1	$4g; 12air; 4g$	3	1.31	0.60	0.80
2	$4g; 12air; 4g$	3	1.31	0.64	0.82
3	$4g; 12air; 4g; 12air; 4g$	3, 5	0.72	0.50	0.71
4	$4g; 12air; 4g; 12air; 4g$	2, 5	0.74	0.55	0.71
5	$4g; 12air; 4g; 12air; 4g$	3, 5	0.72	0.54	0.75
6	$4g; 12air; 4g;$	2	1.31	0.41	0.71
7	$6g; 16air; 4g;$	-	1.14	0.27	0.60

U_g is the thermal transmittance of the glass construction ; g_g is the solar energy transmittance of glass ; τ_1 is the spectral transmittance of the outer glass pane.

area. This amounts for 21 variables for model (a) and 69 for model (b).

18.1.2 Fitness functions

The object of this case study is to search for energy efficient solutions. We will be using the same 3 separate energy calculations for heating, cooling and lighting energy needs that we used in the previous case study. The fitness functions for case study 9 can be explained by the following expression:

$$Case\ Study\ 9 \begin{cases} Minimize & f_{1(x)} = Q_{H,nd}, \\ Minimize & f_{2(x)} = Q_{C,nd}, \\ Minimize & f_{3(x)} = Q_{E,nd}, \\ subject\ to & 0 \leq x_{winPoints} \leq 1. \\ & 0.05 \leq x_{thickness} \leq 1. \\ & 0 \leq x_{winType} \leq 7. \end{cases} \tag{18.1}$$

18.1.3 Genetic algorithm inputs

NSGA-II explores 100 generations with 50 individuals in each generation. The overall genetic inputs for this case study is as follows:

Case Study 9		
Population Size (N)	50	
Number of Variables	21 for model 1	69 for model 2
Number of binary digits	8 for win Points	6 for thickness
Variable Domains	$x_{winPoints} \in [0, 1]$	$x_{thickness} \in [0.05, 1]$
Mutation Probability (p_m)	0.2	
End Condition	End after 100 generations	

18.1.4 Results

Figures 18.2 and 18.3 show the objective spaces and Pareto fronts resulting form case study 9 and figure 18.4 shows a few relevant solutions from the Pareto fronts. The energy requirements for heating, cooling and lighting for all locations vary greatly in this case study. For example, cooling needs for Palermo vary almost 80 kWh/m^2 a year form the best to the worst performing solution in the Pareto front. When compared to the variations seen in the orientation and proportions study, we can note the great importance of the A_t/A_e ratio as explained in (Mechri et al. 2010).

Each objective space presents results for both parametric models (the one window per façade model, and the 4 window model). Looking at the results we can see that the one window model covers a wider area of the objective space. The best performing solutions in most cases are found by the one window model. There are two possible explanations for this result:

- The one window model contains the best performing solutions for all functions in all locations. The four window model is unable to propose solutions that outperform the single window model.
- Since the four window model has a greater number of variables that the single window model, and the GA ran the same number of generations for both models, the exploration on the first model is greater that the second one. In other words, the search space in the four window model is much larger, and in order for us to make a comparison between models, a greater number of generations need to be performed in the larger model.

The second reason is certainly true, exploration in the second model is inferior due to search space dimensions. The first reason is unlikely to be true. There is no evident reason to state that the second model contains inferior solutions to those in the first model. However, with the results

obtained in this study, there is not enough information to rule out this possibility. The purpose of this study is not to compare the two models, but to study the behavior of the proposed buildings with our fitness functions. Therefore we will study the results as they are, and consider solutions from both models, regardless of the exploration level they have achieved.

While heating, cooling and lighting energy need values vary greatly between climate locations, Pareto front shapes for all climates share overall similarities. There seems to be a high level of contrast between heating and cooling requirements, and between lighting and cooling requirements in all locations. There is a very low level of contrast between heating and lighting in all locations. This results are not surprising, but a close examination of the resulting shapes reveals interesting and more specific information found in this study. Figure 18.4 Pareto Solutions for all climate locations. Results will be discussed not by location, but by fitness function.

Cooling Requirements

Solutions A represent the best performing solutions for the cooling requirements for each location respectively. We can see that all A solutions are quite similar to each other, with the exception of Oslo. Solutions A for Palermo, Torino and Frankfurt all have a single wide and short window in the south façade that is positioned very high in the wall. These solutions have very thick walls in all orientations, but most especially in the south façade. High thicknesses means that these short windows are very well shaded by the masonry overhangs. The absence of windows in the east and west orientations is explainable, the GA is avoiding solar heat gains to keep cooling needs low. It could be argued that the best solution for some of these climates could be one without any windows at all. However this is not the case in these results, the solution without windows is outperformed by solutions with the high and wide south facing window. The fact that the window is positioned high in the wall insured an effective lighting strategy. And this is the reason why it outperforms the solution without windows. The lighting fixtures themselves are a significant internal heat source, and since the energy model used in this study uses a dimmer to reduce lighting when it is not needed, the more daylight is present in the room, the less internal lighting heat is introduced. Therefore, if the south facing window is well shaded, but still introduces indirect light, the cooling loads are lower.

Oslo has very small cooling energy needs, and thus the resulting solution in this fitness function is not very significant. However, results are interesting. The Oslo configuration has no window towards the south, it is

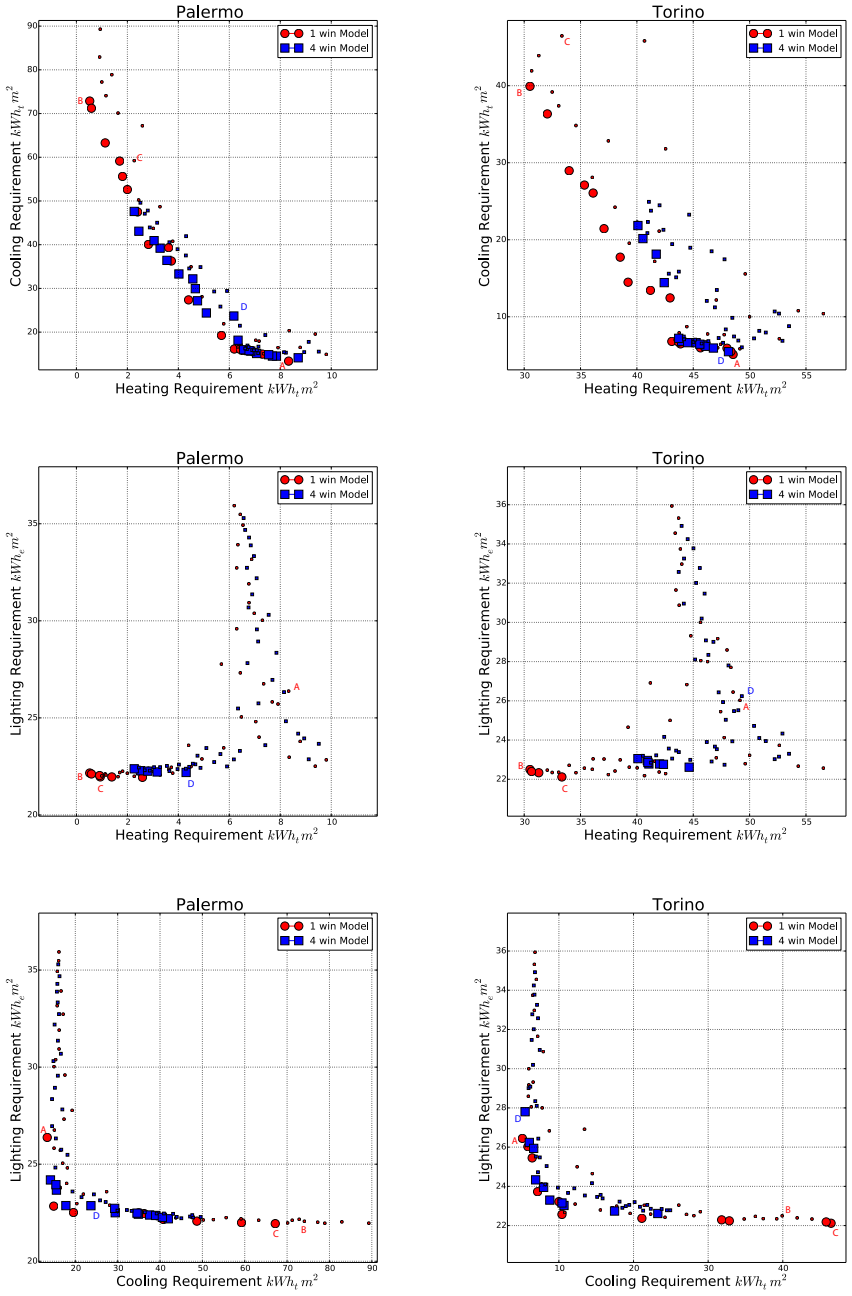


Figure 18.2: Objective spaces for Case Study 9 for Palermo and Torino.

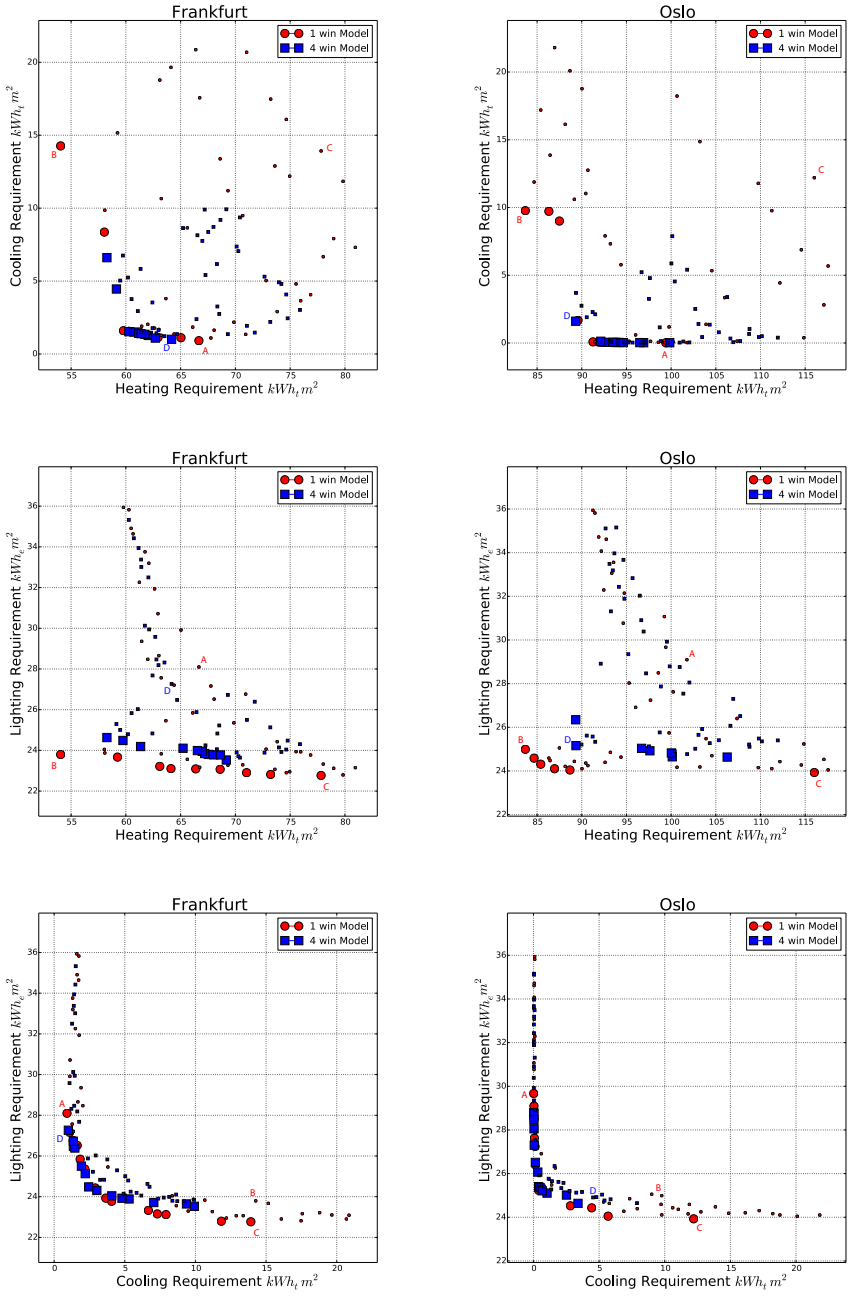


Figure 18.3: Objective spaces for Case Study 9 for Frankfurt and Oslo.

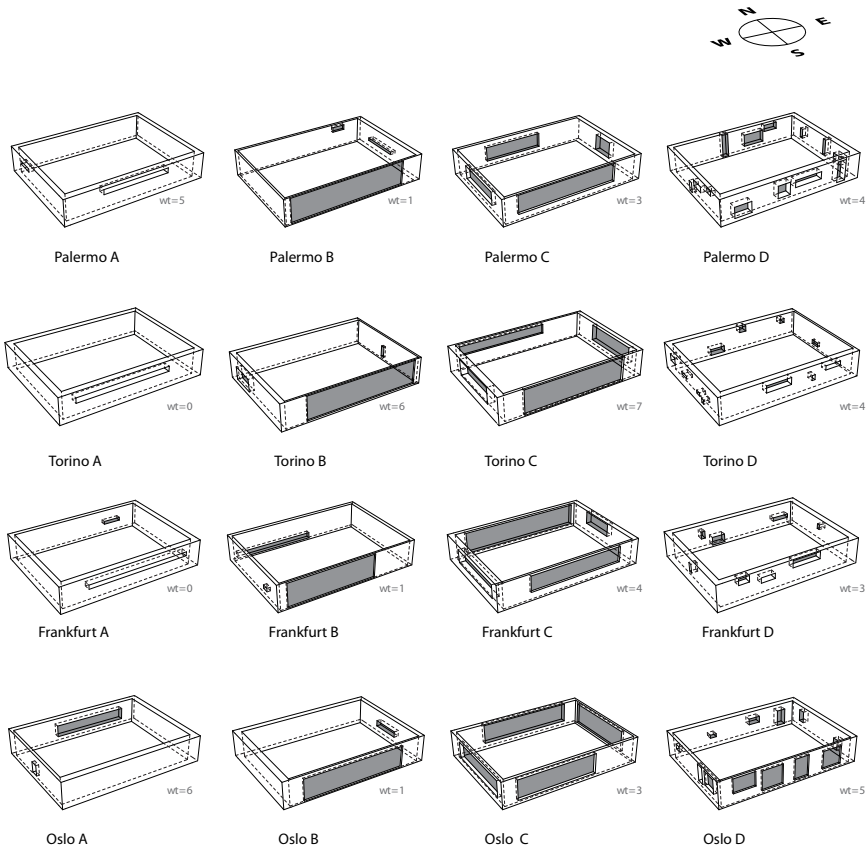


Figure 18.4: Pareto Front Solutions for case study 9.

characterized by a large and tall north-exposed window. This result could be explained by the fact that the sun-paths during the summer months in Oslo's Latitude is not as high as it is in the other climates (see figure 16.4). Lower sun-paths are harder to shade with overhangs, thus the GA opted to have no south-facing windows, and reduce lighting requirements by having a large north-facing window.

Another important aspect to look at in the best performing solutions for cooling requirements is the window type selected by the GA. Window type for Palermo is number 5 in table 18.2, Torino and Frankfurt have window type 0 and Oslo has type 6. Window Type 0 is characterized by having a high visible transmittance of the outer glass pane τ_1, insuring high transmittance of visible light. This selection makes sense because of the lighting internal gains considerations made above. Types 5 and 6 have low solar transmittance values g_g , thus reducing solar gains.

Solutions A tend to be among the worst performing ones for heating and lighting needs since they tend to avoid solar radiation, and introduce just enough light to keep internal gains low.

Heating Requirements

Solutions B are the best performing solutions for the heating energy requirements. Solutions in all locations have very large south facing windows that are not shaded by thickness and are as tall as the wall height. It is clear that the reason for them is to maximize solar gains during the winter months in order to reduce heating loads. Windows in other façades are very few and very small. Solar gains in the winter months are mostly significant only in the south façade, and very poor in the others. Since U values of the window constructions are significantly higher than the wall construction ($0.33\ W/(m^2K)$) the GA avoids windows in non-south façades. The small solar gains acquired by east-west windows during the winter months are not worth the loss of energy due to high U window surfaces.

Window types selected by the GA for the winter months are mostly characterized by having among the lowest U_g values, especially type 3, selected for the Palermo and Oslo locations.

Another interesting trend among results is the presence of high thicknesses among the walls with the exception of the south facing walls. South walls are kept thin to avoid shading, but other walls have much higher thicknesses. This is perhaps more evident in the Oslo B solution. The explanation for this finding can lie in the internal mass of the envelope. The office building model uses in this study contains no internal masses apart

from the ones introduced by the walls. Having higher internal masses seems to increase the energy efficiency for the winter months, the accumulation of heat in the mass could be responsible for better start-stop heating cycles.

B solutions are very poor performers in the cooling function, but are among the best in the lighting function since they introduce a good amount of direct sunlight.

Lighting Requirements

Solutions C represent the best performing for the lighting function. The lighting function leads the GA to produce solutions that have large windows in all façades. Solutions in this category tend to be poorly shaded, especially in the south façade.

Window types selected for this function (4 ,4 ,3 and 5) are among the ones with the highest τ_1 values, while curiously not selecting the highest (type 2).

A low level of contrast would be expected between lighting and heating functions, and this seems to be true for the Palermo and Torino climates. Heating energy needs vary a little between solutions B and C in these locations, but solutions C are never as optimal as solutions B. Frankfurt and Oslo show a large level of contrast between these two functions. Solutions B and C in these locations have very high differences in heating needs, reaching as much as a 32 kWh/m^2 difference in Oslo. The reason for this was already explained above, high window areas loose heat, and are not worth it in north, east or west orientations.

Four Window Model

Apart from the best performing solutions for each function, other solutions are singled out in this section. Solutions D represent interesting results belonging to the four window model.

Solution D for the Palermo location is an interesting compromise solution in the Pareto front. It has a series of mid-size windows in the north and south façades, and very small ones in the east and west ones. All windows seem to be very well shaded, meaning that they introduce very little direct solar radiation, but a good amount of indirect light. As a result, solution D for Palermo is among the best in the lighting function, and has a fair performance in the cooling function, having $10kWh/m^2$ difference from solution A. Solution D is not a very good performer in the heating function, but since heating need in Palermo are very low to begin with, this fact can

be overlooked. Solution D includes window type 4 that has a low g_g value, helping to contain the cooling needs.

Solution D for the Torino study has a series of very small windows distributed among all orientations. They are very well shaded and tend to be wide and short. Solution D has a type 4 window. Solution D is a very good solution for the summer months, significantly containing the solar gains. It is not a bad solution for lighting needs, but it is a poor performer in the heating function.

Solution D in the frankfurt location is quite similar to the Torino D. It has small windows well distributed. However, in this location, these well distributed and sized windows not only insure a good cooling and lighting performance, it also means that solution D is above average in heating needs, being close to $10kWh/m^2$ behind the best heating performer.

The Oslo location produced the best compromise in this study. Solution D for Oslo is well above average in all functions, having a less than $5kWh/m^2$ difference from solution B in the critical heating function. It has 4 large windows facing south, insuring a good solar gain in the critical winter months, and very small windows in the other façades. This solution has the best U_g value available, and also has good thicknesses, insuring high insulation and good internal mass.

It was previously stated that the results do not show definitively that the four window model contains superior solutions in any of the functions. However, there is good reason to suspect that with further exploration, this model can vastly improve its capabilities. It also shows very good compromise solutions come out of it, thus justifying further research into higher window area models.

18.2 Case Study 10: Masonry building envelope - Urban context office building

A second study of the window arrangements in an office building in four european climates is performed, this time having an urban context. The sub-urban study did not have any adjacent buildings casting shadows on its façades. As we have seen in previous studies, solar radiation plays a fundamental role in the energy efficiency of the buildings and by consequence the GA selects solutions that make best use of it. Urban context have adjacent buildings shading façades and therefore window arrangements generated by the GA should have significant differences. Figure 18.5 shows the character-

istics of the urban context used for case study 10. It shows a grid of 20×14 buildings with a 14 meter street and sidewalk between them.

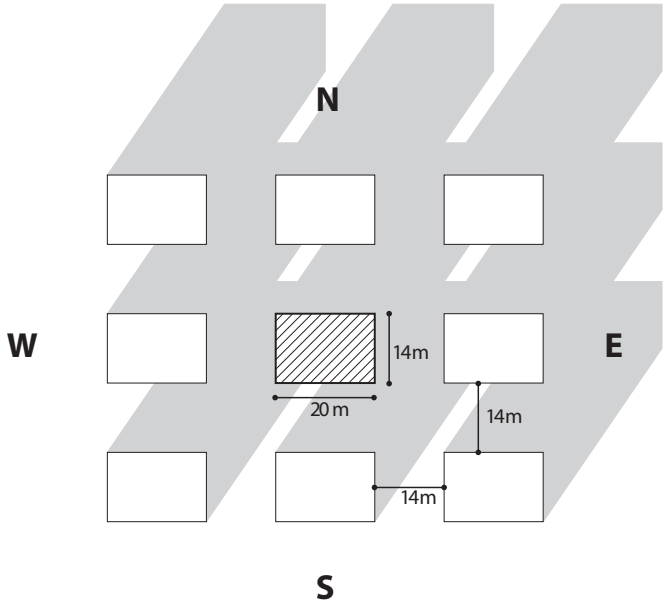


Figure 18.5: Urban context configuration used for case study 10.

18.2.1 Parametric model

This case study employs the same two parametric models used in the previous study, the one window and the four window area models shown in figure 18.1. The model contains the same variable window geometry, wall thickness and window construction types. The window construction types are selected from table 18.2. The building material characteristics are the same as in case study 9. Thermal transmittance of the walls is kept constant at $0.33W/(m^2K)$ as the wall thickness changes, by changing the EPS material thickness as well. As described for study 9, the number of variables is quite different for the single window model an the four window model.

The only difference between case study 10 and case study 9 is the presence of the urban context casting shadows on the building façades. Keeping

the same parametric model and the same building characteristics, allows us to properly compare the results obtained in the studies, and discern the influence of the context in the energy efficiency of the buildings. In order for this comparison to be possible, the fitness functions must also be the same.

18.2.2 Fitness functions

The object of this case study is to search for energy efficient solutions. We will be using the same 3 separate energy calculations for heating, cooling and lighting energy needs that we used in the previous energy case studies. The fitness functions for case study 10 can be explained by the following expression:

$$\textit{Case Study } 10 \begin{cases} \textit{Minimize} & f_{1(x)} = Q_{H,nd}, \\ \textit{Minimize} & f_{2(x)} = Q_{C,nd}, \\ \textit{Minimize} & f_{3(x)} = Q_{E,nd}, \\ \textit{subject to} & 0 \leq x_{winPoints} \leq 1. \\ & 0.05 \leq x_{thickness} \leq 1. \\ & 0 \leq x_{winType} \leq 7. \end{cases} \tag{18.2}$$

18.2.3 Genetic algorithm inputs

Genetic Inputs for case study 10 are also the same as for case study 9. NSGA-II is used for 100 generations with 50 individuals in each generation. The overall genetic inputs for this case study is as follows:

Case Study 10		
Population Size (N)	50	
Number of Variables	21 for model 1	69 for model 2
Number of binary digits	8 for win Points	6 for thickness
Variable Domains	$x_{winPoints} \in [0, 1]$	$x_{thickness} \in [0.05, 1]$
Mutation Probability (p_m)	0.2	
End Condition	End after 100 generations	

18.2.4 Results

The most evident result in this study is the fact that heating energy needs are significantly higher than those of study 9 for all locations, and cooling needs significantly lower. This is the most important influence of the Urban

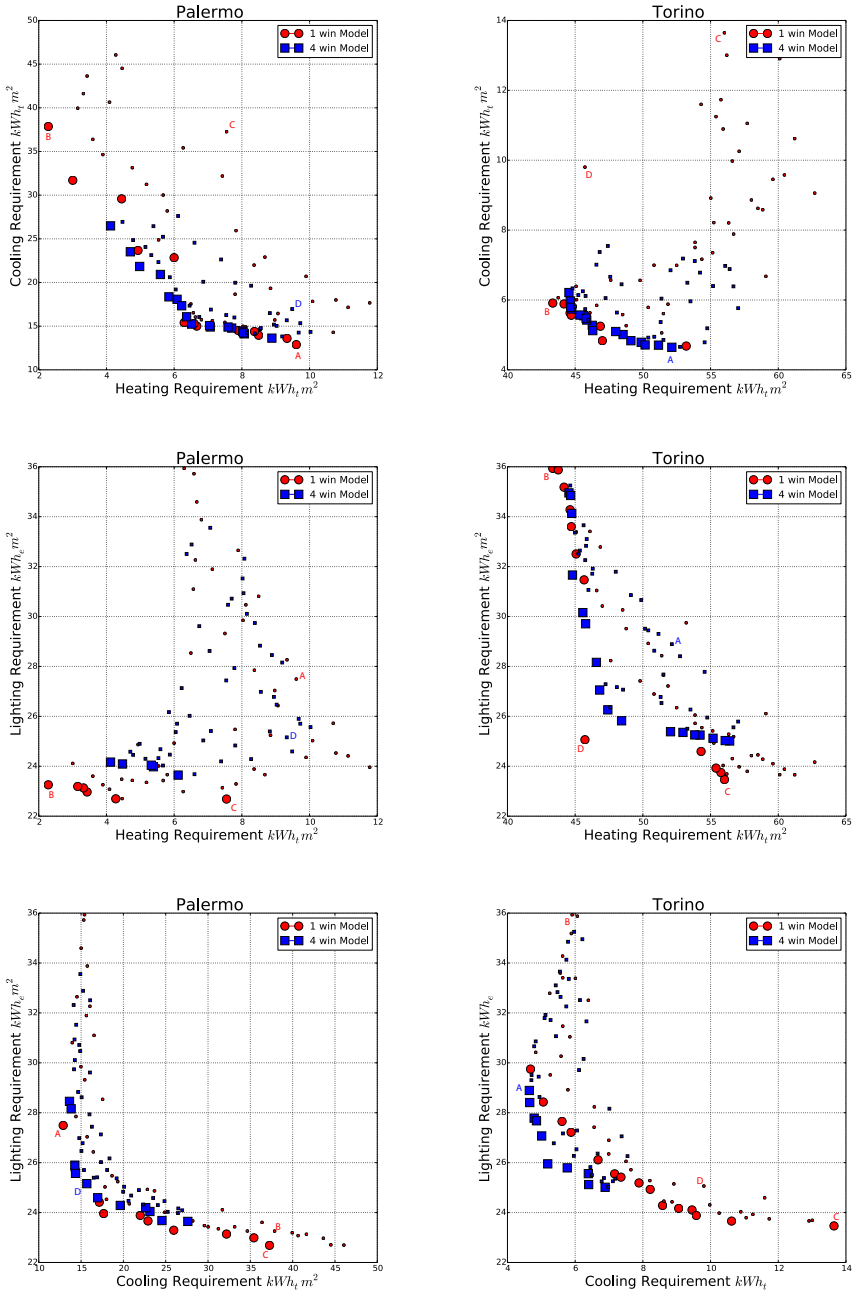


Figure 18.6: Objective spaces for Case Study 10 for Palermo and Torino.

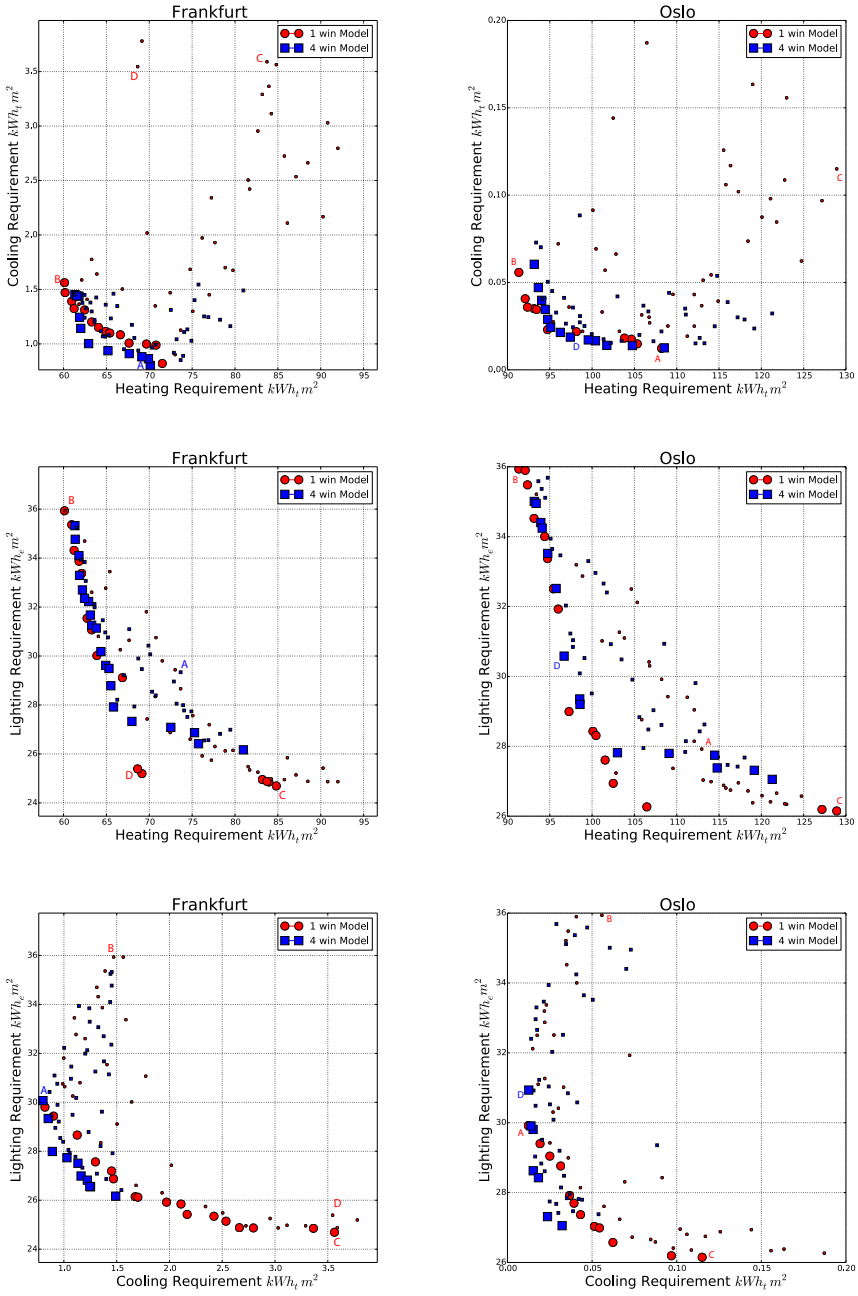


Figure 18.7: Objective spaces for Case Study 10 for Frankfurt and Oslo.

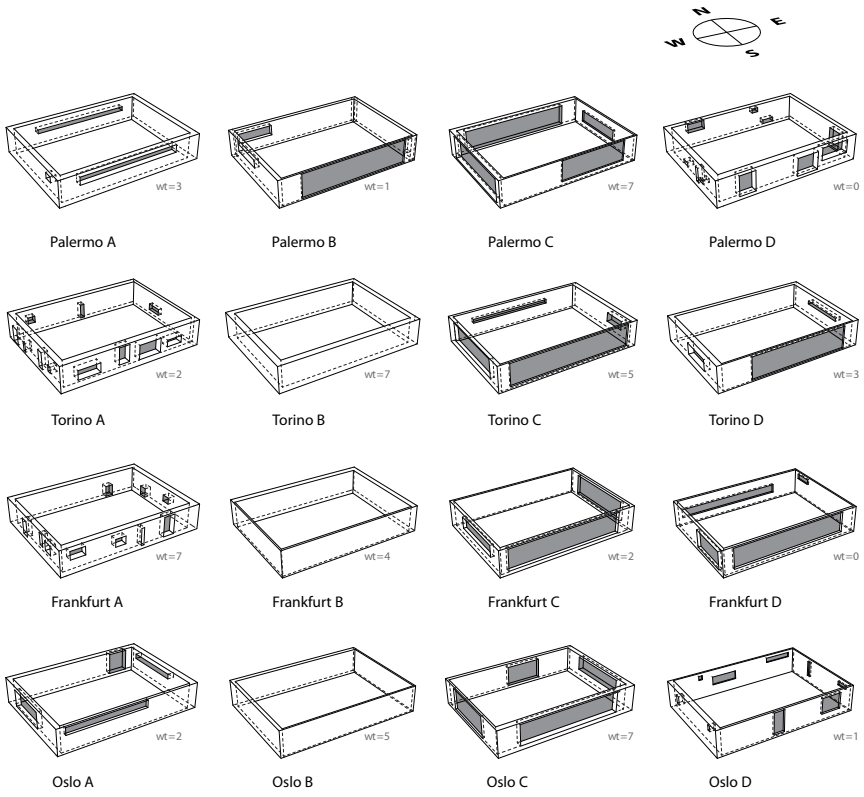


Figure 18.8: Pareto Front Solutions for case study 10.

context, the shading of the adjacent buildings. Variations between best and worst performing solutions for the cooling function are smaller in case 10 than in case 9. This is also a sign of influence of the shading, solar input is much smaller and this from a cooling point of view, solutions are more similar to each other. Lighting variations are similar between the two studies, and heating variations are similar as well, with a slight increase in the Oslo climate.

The overall shapes of the Pareto fronts are quite similar between studies 9 and 10. There is considerable contrast between the heating and cooling functions, and between the lighting and cooling functions. Case study 9 shows little contrast between heating and cooling functions for Palermo and Torino, and more significant contrast in Frankfurt and Oslo. In case study 10 this behavior is different, contrast between heating and lighting functions is very significant in Torino, Frankfurt and Oslo, and still present in Palermo. The reasons for this will be further detailed in the Pareto solutions analysis bellow.

A discussion on the exploration done by the GA in the single window and the four window models presents some difficulties in this study as well. Genetic inputs for case study 10 show that the same number of individuals and generations was used for both parametric models. As was the case in study 9, the single window model has a wider extension of solutions in the objective space than the four window model. This is a sign of higher exploration in this model, and it is explained by the lower number of variables. However, in case study 10 this difference in exploration seems to be less pronounced. Best performing solutions in the first model do not have a large difference from those in the second one. In fact, the best performing solution in the cooling function for Torino belongs to the second model, and in some cases the Pareto front from the second model dominates a good number of solutions in the first model front. These results suggest that, while having had less exploration during the GA run, the second model has an advantage in this case study.

Cooling Requirements

Best performing cooling requirement solutions in case study 10 (Solutions A) share some characteristics with those in case 9. Mainly the presence of the long and shaded window in the south façade (excepting in Torino where there are 4 mid-sized and shaded windows facing south). But the presence of shading adjacent buildings does have an influence. Since allowing daylight into the spaces requires bigger windows than in case study 9, artificial light-

ing heat gains are lowered by increasing window sizes. Other façades have more openings in case 10 than in case 9, increasing daylight and reducing cooling loads.

Solution A for Palermo in case study 10 is the best performing for cooling. It has a long and shaded window in the south-facing wall as is common in the previous study, but in this case, there is also a similar window in the north wall. This north window improves internal daylight and reduces internal lighting heat gains. Other high performing solutions for cooling have similar window arrangements, window areas in for Palermo are significantly higher in this study, suggesting that the lighting internal gains have an important effect. Window type selection for solution A is type 3, a low U_g construction with relatively high visible light transmittance, further highlighting the importance of daylight in cooling energy efficiency.

Solution A in Torino has four south-facing windows, four small and vertical west-facing ones and a few very small north and east-facing ones as well. All windows are small enough, and the walls are thick enough for them to be well shaded. It seems that the best way to shade the windows in the west façade is to have them be vertical, and shade with the fins, not the overhangs. This configuration found by the four window model outperforms any solution found by the single window model, including the single long and short window that had so far outperformed all others for cooling. This could be because this four window configuration is able to introduce more daylight in our office space without allowing solar radiation in more facçades. Window type 2 gives solution A a high amount of daylight as well.

Solution A for the Frankfurt climate is similar to the Torino A. It is also a four window solution with a series of shaded south-facing windows and some smaller ones in other façades. Shading in east and west façades in this case is also mostly done by the fins since they have vertical windows. Solution A for Frankfurt has window type 7, this is the type that allows the least amount of solar energy in the room.

Solution A in the Oslo climate has the usual long shaded south-facing window, with the addition of a larger and taller west window, a squared north one, and another long window due east. Solution A has the particularity if having some of its windows be placed asymmetrically in the wall, especially the north and south windows. It is unclear if this asymmetry is advantageous from a cooling point of view, if this positions maximize daylight, or if they are better shaded in the urban context. Also in this climate, window type 2 gives solution A a high amount of daylight as well.

Heating Requirements

Heating requirements in case study 10 are significantly higher in comparison with case 9. Solar gains are harder to come by with the presence of the adjacent buildings in the winter months when the sun is low. Solutions B represent the best performing solutions for the heating function.

Solution B in Palermo is quite similar from case 10 to case 9. A very large and unshaded south-facing window maximized solar gains that are beneficial to heating loads. Window type 1 has does not have the lowest U_g value, but it does allow a good amount of solar radiation, having a 0.60 g_g value. Sun-paths in Palermo are still high enough during winter for the sun to find its way over the adjacent buildings into the office space.

Solution B in the Torino climate is quite an interesting result. It has no windows in any façade, not even due south. We saw in the previous study that large unshaded windows on the south wall improve heating considerably by allowing solar radiation indoors. However, in this case study this is not so, because the sun-paths in the Torino latitude (and upwards) are not high enough during winter to irradiate over the adjacent buildings. Large windows therefore provide no solar heat to the internal spaces, on the contrary, they represent a heat loss because of their higher U values. When there is no solar radiation be be had, having a continuous wall with a U value of 0.33 $W/(m^2K)$ is better than having windows with a U_g value of 0.72 $W/(m^2K)$ at best. Window types in this result are irrelevant since there are no windows. We must consider that this result is the valid only when direct solar radiation is very low, and this is true in our case because we are studying a very low floor of our building (the first one above the ground floor).

Having no windows at all clearly represents a problem from the lighting point of view. We can see that solution B is the worst performing solution for the lighting function. Consequently we can note a good amount of contrast between the heating and lighting functions. We assume that internal heat gains due to lighting fixtures are beneficial to the heating loads, but a good amount of contrast is present nonetheless.

Solution B in Frankfurt and Oslo have the same result as the Torino climate, no windows are present. The same reasoning applies to Oslo. The results for Frankfurt show two very similar solutions to be the best performing for heating, one of them is solution B and the other one is also a solution containing no windows, but with the difference of having a much higher thickness of the walls. Solution B outperforms by a very small difference the other no-window solution. Since there are no windows and U

values are fixed (they do not change with the variation of thicknesses), the only influence of the thickness is the internal mass. We see a very small difference in heating energy needs, this means that internal mass has a very small influence in heating requirements for this climate as was shown in (Mechri et al. 2010).

Lighting Requirements

Lighting energy needs do not increase significantly in case study 10. Adjacent buildings have shown to noticeably decrease direct solar radiation, but this is not the case for daylight. Variations in lighting needs are quite similar in both case studies.

The best performing solutions in the lighting function are shown as solutions C. As is to be expected, best performing solutions have large windows in all façades. This is also true in the results for case study 10. Windows are generally unshaded, especially in the south-facing façade. There are also some asymmetrically positioned windows in these results, but the reason for them is unclear with the present study.

Compromises

Solution D in the palermo climate is an interesting compromise solution generated by the four window model. It is a very good performer in both cooling and lighting functions (the most critical functions for Palermo). Solution D is not a very good performer in the heating function, but as we have mentioned above, this is not a big problem in Palermo. Solution D has 3 almost square windows in its south wall. These windows seem to be shaded enough, either by the thick south wall, or by the adjacent buildings. The east wall contains one such similar window that is positioned very close to the south edge of the wall. The northern and western façades contain a series of very small windows. Solution D has a less than 5 kWh/m^2 a year difference with the best performing cooling solution, and less than 3 kWh/m^2 difference with the best lighting solution.

Solution D for the Torino climate is an excellent compromise for the heating and cooling needs generated by the single window model. Since in the urban context the cooling needs are significantly reduced, and heating ones increased, we can say that it is heating and cooling that require the most attention in this case. Solution D is a mere 3 kWh/m^2 behind the best performing solutions for both heating and cooling functions. Solution D has a large and unshaded window in the south wall and a couple of mid-sized

windows in the east and west facades. Solution D uses the window type with the lowest U_g value, making the best use out of the window areas. We have seen above that windows in the Torino climate represent heat loss and very little solar gain, but with this window type this balance is perhaps not completely unfavorable. This high window area also improves significantly the lighting energy requirements.

Solution D in Frankfurt is a very good compromise for heating and lighting functions. It optimizes lighting needs by having very large windows, and solar gains for heating by a large and unshaded south-facing windows. Heating requirements are only above average but lighting needs are near optimal. Cooling requirements are not optimal for this climate, but they are still very low due to the fact that Frankfurt has low cooling requirements in the urban study.

Solution D for the Oslo climate is a very good heating performer with an above average cooling performance as well. Lighting is not very well solved in this solution. Solution D has a series of small windows in all of its façades, and window type 1 that allows a good amount of solar radiation to enter the room.

18.3 Conclusion

The case studies presented in this chapter show the fundamental role that the window arrangements have in the energy efficiency of the office building in question. Solar radiation seems to be the key aspect in all functions and climates. Hence there is a big difference between the urban and sub-urban contexts, both in the energy requirements and the resulting solutions. The window to wall ratio was determined to be the more important aspect to study when compared to the orientation and building shape.

The MOGA was able to provide us with detailed and useful information on the configurations that best dealt with the fitness functions, climates and contexts studied. Optimal configurations for all functions and climates were found and the important relationships between the functions were deduced from the Pareto fronts. Good compromises, solutions that are good performers in at least two important functions were also presented in each climate, but most importantly, the characteristics that made these solutions work were noted in the search process. The information provided by the MOGA was site and context specific, making it quite useful in the early stages of the design process.

Not all of the aspects that need to be considered in the design of the

building envelope were subject of study in this chapter. For example, the visibility from the interior to the interior, the visual connection of the people inside to the external environment. Not only was is not considered, some solutions proposed by the GA had no windows at all, and some have windows so high that visibility is only possible with the sky. Visibility, as many other important considerations can be determined by the designer during the search process in the following ways:

- Designers may use parametric models that have a minimum window area as a constraint. Meaning that all solutions generated would have at least some percentage of windows. Windows may also be constrained in space, allowing the GA to move them only in certain areas where designers consider them to have the most visibility, or for them to have some aesthetic value.

- Designers may let the GA generate any kind of window arrangement or no windows at all (as was the case in this chapter) and then choose a final solution considering not only their fitness values, but also considering visibility, aesthetics, etc*.

- Designers may chose to interact with the GA during its search process, keeping visibility as an implicit goal not present in the fitness functions†.

*This issue is also discussed in section 7.7.
†This possibility is discussed in section 1.7.

19

Multi-Disciplinary Search

Section 1.1 of this thesis gives a description of the early design phase of architectural design. Particular attention is given to the multi-disciplinarity and the contrast of the design problems faced in this stage of the design process. This PhD thesis presents two series of multi-disciplinary problems based on the studies presented above.

In a few words, Turrin et al. describe the reasoning behind the use of performance based search processes in the early phase of architectural design:

> "Despite the fact that conceptual design is well known to be initiated based on a set of design requirements, traditionally the conceptual phase of architectural design addresses only a rather limited selection of requirements (in most cases, functional and esthetic aspects prevail), while key disciplines tend to be entirely omitted in this phase and postponed. In contrast with this tendency, the concept of performance oriented (also called performative) architecture has recently emerged, as a design approach in which building performance, broadly understood, becomes a guiding criteria."

(Turrin et al. 2011)

Building performance based search processes are proposed by Turrin et al. among many for the early design stages. They employ structural FEM simulations coupled with energy simulations in long span roof case studies (Turrin et al. 2009, 2011). Their parametric models include two types of

variables, a first group determines the overall shape of the structure, and a second group determines the shape of a louver system on top of the structure. The structural performance of roof would depend solely on the overall shape, while the energy performance would depend on both, arguably mostly on the second. They use GA's to search for high performing solutions to both functions and discuss the importance of such tools in early architectural design.

A previous study developed by the author also discussed the relevance of multi-disciplinarity (Méndez Echenagucia et al. 2008*a*). In this case the attempt was that of embedding the multi-disciplinary efficiency onto the architectural shape. While Turrin et al. take an approach that suggests that different components can address different issues, the case studies presented in this PhD work try to generate single, continuous and homogeneous shapes that are advantageous for multiple performance metrics.

The previous chapters presented search processes that involve performance analysis of different architectural shapes. Multi-disciplinary search processes are carried out in two kinds of shapes, and for two kinds of performances:

- Complex curved surfaces are studied for their acoustical *and* structural capabilities. As we have seen in the previous chapters, curved surfaces present a great deal of opportunities in their ability to evenly distribute sound energy inside concert auditoria, as well as carry structural loads efficiently, with very little material and with very interesting shapes. We have discussed in this thesis different methods for studying both kinds of performances, and to parametrize and discretize (when needed) complex curved surfaces. This gives us all of the tools we need to study these shapes multi-disciplinarily.

- Masonry building envelopes are studied for their energy *and* structural capacities. The study of a rectangular building with a masonry envelope form both structural and energy points of view has been shown above. In both cases the GA was able to generate geometries that optimize different functions, drawing Pareto fronts that reveal important information about these buildings.

The use of the multi-objective approach described in this PhD thesis on multi-disciplinary problems is a fundamental tool in the study of these geometries. Knowledge on the contrast (or lack there of) in these functions can be used by designers to effectively define more efficient and informed geometries, early in their design process.

19.1 Structural and Energy Search

Many traditional and contemporary buildings have used the envelope as the main structural component. From historical masonry buildings, balloon frame houses to steel façade skyscrapers, the envelope is an habitual and logical part to place structural supports, but it is also inevitably the most important environmental filter, and as previous studies in this PhD thesis as well as other research has shown, it bears a big responsibility in the energy consumption of the building. This chapter studies the structural and environmental capabilities of building envelopes.

19.2 Case Study 11: Masonry building envelope - Urban context office building

Case study 11 follows the work presented above on masonry building envelopes. It combines the work developed in case study 5 on structure and studies 9 and 10 on energy efficiency.

19.2.1 Parametric model

The Parametric model for case study 11 is the same used in case studies 9 and 10 and shown in figure 18.1. It contains the same variables (window configurations, wall thicknesses and window construction). Wall material construction is the same used in cases 9 and 10 and described in table 18.1. The window constructions available to the GA are also the same ones used in the previous case studies and shown in table 18.2. Case study 11 does not use the single window model, it used only the four window model described in figure 18.1b.

Case study 11 is set in an urban context, the same urban context used in case study 10 and described in figure 18.5. The only climate chosen for this case study is the Palermo climate. Average monthly temperatures for Palermo are shown in table 16.2 and solar radiation diagrams are shown in figure 16.4.

19.2.2 Fitness functions

Case study 11 proposes the structural and energy study of masonry envelopes for rectangular office buildings. Fitness functions for this study are

directly taken from studies 5 for the structural part, and 9 and 10 for the energy aspects.

Case 5 employs equation 11.1 to determine the structural adequacy of load bearing masonry wall envelopes. This equation is also chosen to study envelopes in case study 11. Case 5 also studied the weight of the walls as a contrasting function to equation 11.1, but in case 11 weight is not used. The contrast between structural and energy functions is the only subject of this study.

Cases 9 and 10 use the energy needs for heating, cooling and lighting separately as 3 functions that study the total energy needs of the buildings. Case study 11 uses these 3 functions as well, along with the structural function. The multi-objective problem studied in case 11 can be summed up in the following equation:

$$\textit{Case Study 11} \begin{cases} \textit{Minimize} & f_{1(x)} = Q_{H,nd}, \\ \textit{Minimize} & f_{2(x)} = Q_{C,nd}, \\ \textit{Minimize} & f_{3(x)} = Q_{E,nd}, \\ \textit{Minimize} & max(f_{case1}; f_{case2}) \\ \textit{subject to} & 0 \leq x_{winPoints} \leq 1. \\ & 0.05 \leq x_{thickness} \leq 1. \\ & 0 \leq x_{winType} \leq 7. \end{cases} \tag{19.1}$$

19.2.3 Genetic algorithm inputs

Genetic inputs for case study 11 are also the same as for case studies 9 and 10 with the consideration that only the four window model is used. NSGA-II is used for 100 generations with 50 individuals in each generation. The overall genetic inputs for this case study is as follows:

Case Study 11		
Population Size (N)	50	
Number of Variables	69	
Number of binary digits	8 for win Points	6 for thickness
Variable Domains	$x_{winPoints} \in [0, 1]$	$x_{thickness} \in [0.05, 1]$
Mutation Probability (p_m)	0.2	
End Condition	End after 100 generations	

19.2.4 Results

Figures 19.2 and 19.1 show the objective spaces and best performing solutions for case study 11. The results for case study 11 show similar patterns to those seen in case 10. With the exact same GA inputs, exploration in this case seems to be lower when compared to case 10. This can be explained by the presence of a fourth fitness function. Contrast between energy functions is analogous to the ones found in study 10. For this reason, comments on results for case 11 will be concentrating on contrast between the structural and energy functions.

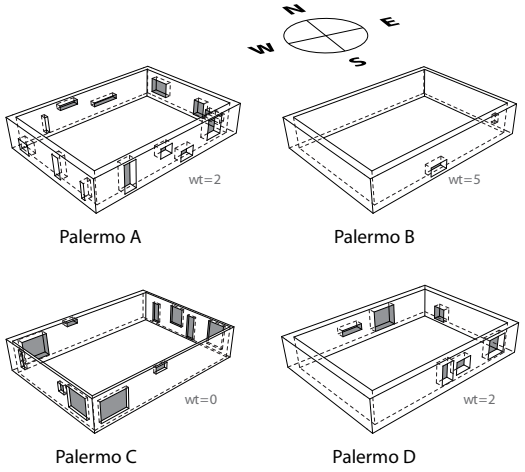


Figure 19.1: Best performing solutions for Case study 11.

Structure

Solution D represents the best performer for the Structural function, its most important characteristic is not its window arrangements, but its thicknesses. Solution D has the thicker walls in all façades than all other solutions in the Pareto front.

Structural fitness functions, material conditions and parametric models for cases 5 and 11 are identical. Therefore comparisons for structure can be made between these two studies. Case study 5 achieved a best performing structural fitness value of 1086, while case 11 was able to surpass this result

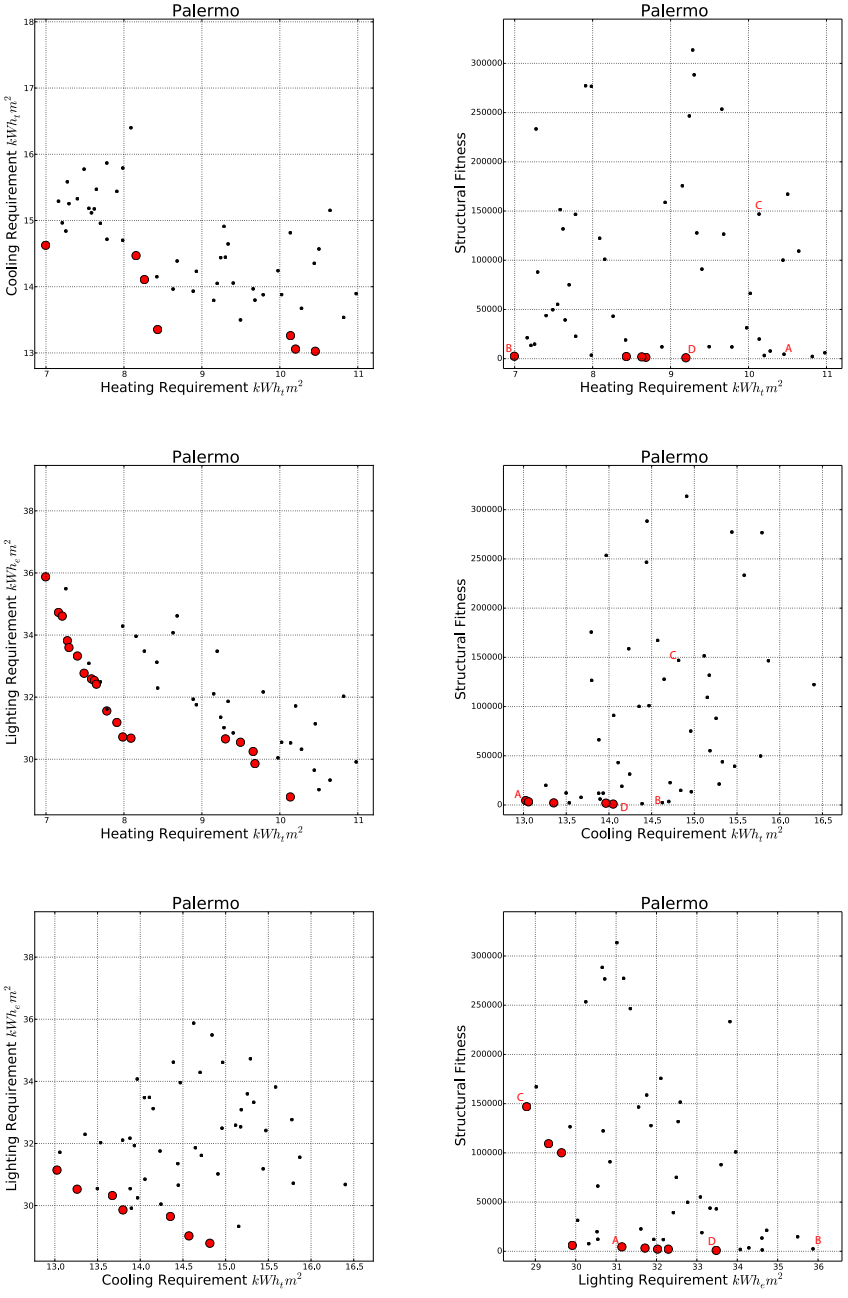


Figure 19.2: Objective spaces for Case study 11.

with solution D, achieving a value of 930*. GA inputs were also the same between case 5 and 11, so the better result in case 11 cannot be directly attributable to better exploration. In fact, there are more fitness functions in case 11, so we could argue that exploration for structural attributes should be diminished. The improved results can therefore be only explained in two ways:

- There is an important random component in the search process done by the GA. Therefore, two identical search processes do not generate identical results. Inevitably one of the two will have better results that the other.
- The weight function is more contrasted to structural efficiency than the energy functions. As it is explained above, results in case 11 show that the wall thickness is the most important variable when it comes to structure. The higher the thickness the better the result. The weight function is the opposite, the lower the thickness the lower the weight. Contrast between the structural function and the energy ones are explained bellow, but they seem to be less pronounced than the weight function. It can be argued that the higher the contrast, the more trouble the MOGA will have in finding results for each single function. Hence, the weight function makes it harder for the GA to find thick walls and optimal structures.

Cooling vs. Structure

Solution A is the best performing solution for the cooling function. It has thick walls that shade its small windows. Shading and thick windows were also shown to be important in cooling functions in the previous case studies. For this reason, contrast between the cooling and the structural functions is not very big. The structural fitness difference between solutions A and D is very small when compared to most of the other solutions found in the population. Other high performing solutions in the cooling function are also high performing in the structural function.

We have established that high performing cooling solutions are also high ranked structurally, but the opposite is not true. Solution B for example is high performing structurally since it has almost no windows and has thick walls, but it is not a good cooling solution. It was established in cases 9 and 10 that solutions with no windows are not very good for cooling since they require high lighting internal gains.

*The structural fitness function is a minimization function, lower values are better

Heating vs. Structure

Heating solutions for the Palermo climate in the previous urban study were shown to have large south facing windows that increased solar gains. These solutions have a heating energy requirement almost as low as 2 kWh^2 a year. Solution B is the best performing solution for case 11, it has a heating energy requirement of 7 kWh^2 a year, not as low as the one in case 10. Solution B does not have large and unshaded south-facing windows. It has only 2 very small and shaded windows, it has a low heating value not because of high solar gains, but because of low U values. The absence of windows keeps the thermal transmittance to a minimum. The high thicknesses and lack of windows give solution B a high structural performance.

The exploration in this case was not enough for it to find high performing heating solutions such as the ones found in case 10. This is also probably due to the fact that the GA is looking for high thicknesses for structural reasons. This gives us a low degree of contrast between these two functions, but looking at the results in case 10 is easy to assume that better performing heating solutions would not be as good performing in the structural functions. However, the results found show that the contrast is not very high, the GA is able to find good compromises between heating and structure, such as solution B.

Lighting vs. Structure

Lighting is the function that shows the most contrast with the structural function. Solution C is the best performing solution for the lighting function, it has larger windows than all other solutions shown, and these windows are not very well shaded (low thicknesses). This allows sunlight to enter the room freely. As was the case with heating, solution C is not as good a performer as the solutions found in case 10.

Large windows and low thicknesses give solution C a very bad structural performance, and this is true for *all* other high performing lighting solutions. This explains the high contrast between these two functions.

19.3 Structural and Acoustic Search

The acoustic case studies shown above make use of various simulation tools that obtain sound quality descriptions inside the rooms, and in turn shape those rooms to better distribute sound quality. The overall shape of the room was shown to be important in determining the distribution of quality,

but also singular surfaces can have a great impact as well, particularly the roof surface.

Complex surfaces have also been shown to be very interesting in structural problems. Compression only surfaces for example are presented in this PhD thesis show low weights and high structural performance. Concrete shells are also studied above in terms of maximum displacements and weight.

Complex surfaces are becoming more and more present in contemporary concert spaces, and shell structures are also being employed by architects as expressive and efficient structures. This chapter considers the possibility of combining the study presented on shell structures and acoustic surfaces, in order to study the relationship between these two types of functions.

19.4 Case Study 12: Concrete shell roof for a concert hall

Case study 12 is the result of the combination of the studies on concrete shells (cases 1 and 2) and the study of complex acoustic reflectors (case 7). A concrete shell roof for a 20×42 shoebox concert hall is studied for both its acoustic and structural capabilities. Case study 12 involves free-form curved surfaces, hence it involves the use of the acoustical study of early sound developed for this PhD thesis and discussed in chapter 15.

Since reinforced concrete is capable of resisting tension forces and not only compression forces, concrete shells are much less constrained form a shape point of view than compression only shells. They are able to take concave and convex shapes without loosing structural capacity. We can say that FEM calculations of the maximum displacement in specular shells (identical shapes, one concave and one convex[†]) under the same loading conditions, would not show any difference. For this reason, previous structural results are found to have no preference for concave or convex shapes. The acoustic study detailed above, on the other hand, shows that there is some preference for convex surfaces that avoid sound concentrations in the audience area. While some concave curves are shown in the study, especially in the longitudinal section, most solutions exhibit convex shapes. For this reason it is interesting to se the resulting level of contrast between structural

[†]Since this is also an acoustic study, shapes are referred to as concave or convex from the point of view of the audience, hence from the bottom of the shells. Shapes that are convex towards the audience tend to avoid sound concentrations.

and acoustic fitness functions.

19.4.1 Parametric Model

The parametric model used for case study 12 is the same one used in case 7 and shown in figure 15.2. The model presents a 20×42 shoebox concert hall, with a 15° inclination of the audience area. The shell roof is supported all along the perimeter of the room, meaning that the roof structure has a span that is 20 meters in its transversal section. No special reflectors, balcony fronts or overhangs are present in the room. The sound source is placed in the center axis of the room, four meters behind the stage front edge. An aisle of 2 meters in width was left all around the audience area, and this area was subdivided into flat segments of 3.2×3.2 meters.

The variables in this case are also the same ones used in case study 7 and detailed in table 15.1. This variable settings imply that symmetry is imposed in the shell surface, thus reducing the number of possible solutions by excluding asymmetrical configurations that would most likely be low performing in acoustical fitness functions.

19.4.2 Fitness functions

The fitness functions used for case study 12 are taken directly from the structural and acoustical case studies of shell surfaces.

The structural performance of of the concrete shell is studied by means of the maximum displacement of the structure in the Z axis. A FEM simulation of the shells behavior is performed for each individual solution. A NURBS surface is discretized into small triangular shell elements of the FEM study. Gravity loading is applied in this case study, meaning that only the weight of the shell itself is considered.

A weight function was also used in case studies 1 and 2 as a contrasting function to the maximum displacement. In case study 12 only the contrast between the acoustic and structural function is object of study, therefore the weight function is not included in this study.

The acoustic fitness functions are the same ones used in case 7 and described in equation 17.1. Three separate time-windows are used in this study of acoustical quality of the early sound inside the room. The first time-window starts at 0 ms from the arrival of direct sound, until 80 ms after, the second window goes from 80 to 120 ms and the third one from 120 to 200 ms.

The fitness functions used in case study 12 can be expressed in the following equation:

$$Case\ Study\ 12 \begin{cases} Minimize & f_1(x) = max(\Delta_{Z_i}), \\ Minimize & f_{2(x)} = E_{tot,0-80}, \\ Minimize & f_{3(x)} = E_{tot,8-120}, \\ Minimize & f_{4(x)} = E_{tot,120-200}, \\ subject\ to & 5 \leq x_{1,2} \leq 20. \\ & 10 \leq x_{3-6} \leq 20. \end{cases} \tag{19.2}$$

19.4.3 Genetic algorithm inputs

Genetic Inputs for case study 12 are described bellow. NSGA-II is used for 100 generations with 50 individuals in each generation. The overall genetic inputs for this case study is as follows:

Case Study 12		
Population Size (N)	50	
Number of Variables	6	
Number of binary digits	8	
Variable Domains	$x_{1,2} \in [5, 20]$	$x_{3-6} \in [10, 20]$
Mutation Probability (p_m)	0.2	
End Condition	End after 100 generations	

19.4.4 Results

Figure 19.3 shows the objective spaces for all 6 combinations of fitness functions used in case study 12. The objective spaces shown in the left column are all regarding the structural function in combination with the first second and third time-windows. Figure 19.4 shows a few signaled out solutions resulting from the study.

The objective spaces found in case study 12 show a moderate level of contrast between structural and acoustic functions. The highest level of contrast being present in the first time-window f_2, as evidence by the fact that the best performing solution in this time-window is one of the worst performing in the structural function f_1. In the other acoustical functions the contrast is fairly low, solutions that are high performing acoustically are very high performing structurally as well, but there is still a small level of contrast between these functions.

While the level of contrast is moderate between structural and acoustic functions, acoustic functions among themselves are very contrasted. The

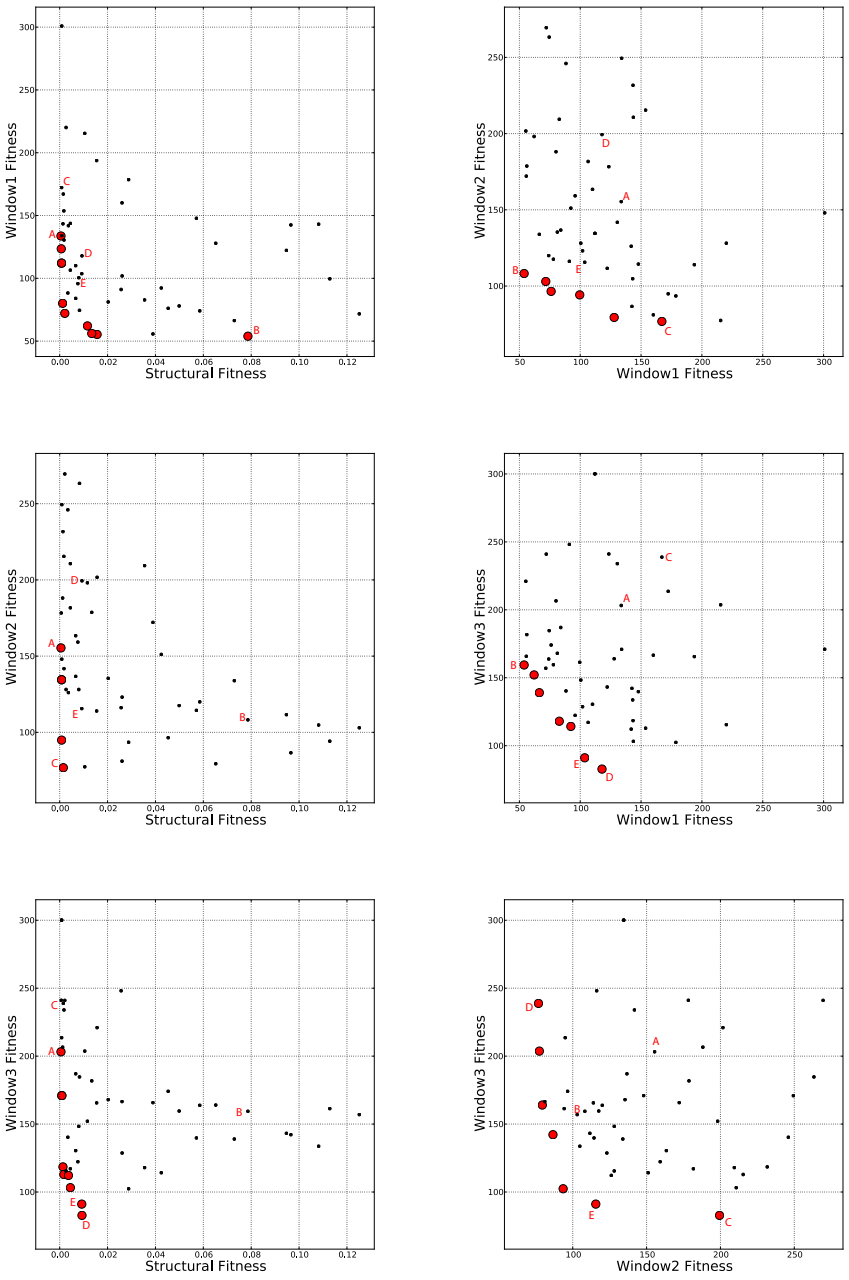


Figure 19.3: Objective spaces for Case study 12.

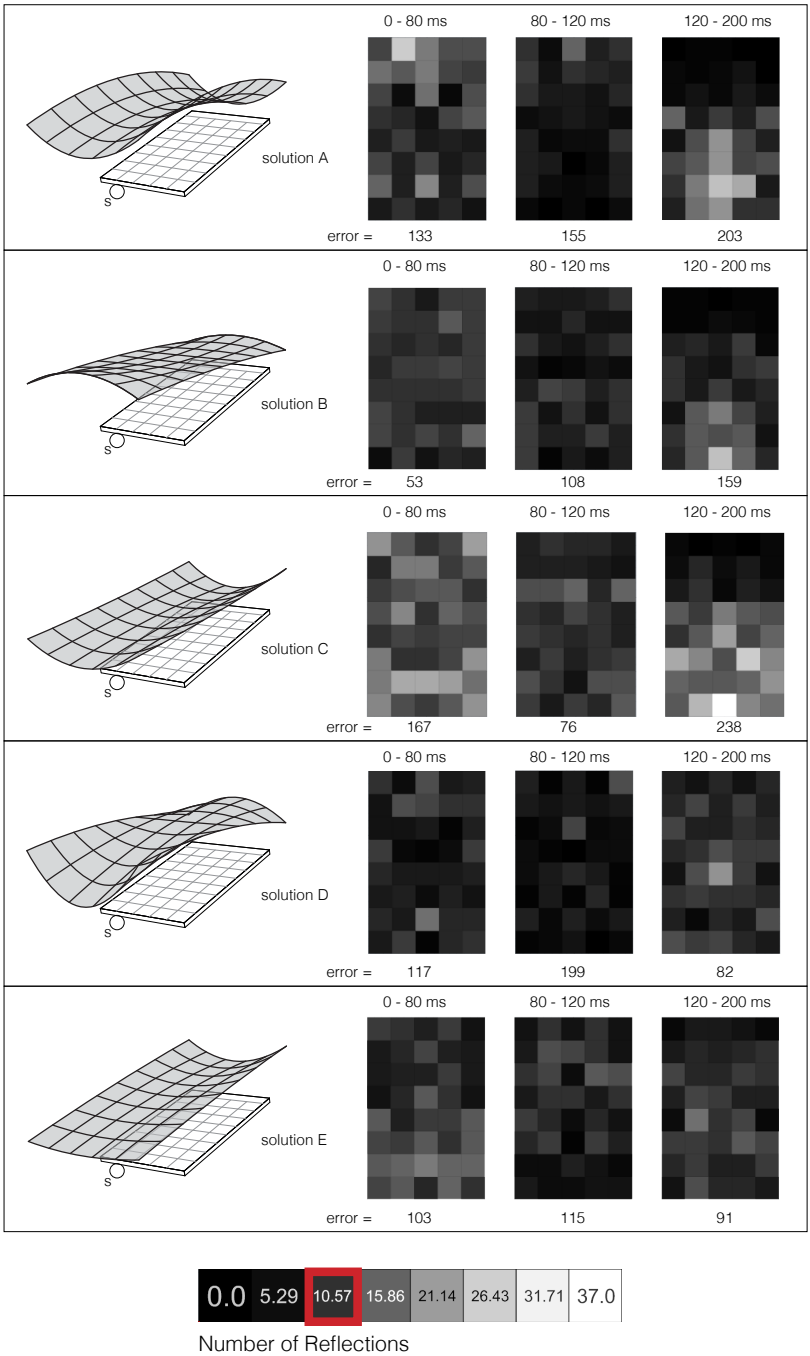


Figure 19.4: Pareto Front Solutions for case study 12.

Pareto fronts describing the combination of f_2 vs. f_3, and especially f_2 vs. f_4 denote high contrast. Compromises between these functions are very hard to achieve.

Solution A is the best performing individual for the structural function f_1. It has a maximum displacement in the Z axis ($max\Delta_Z$) of 0.5 millimeters in a 20 meter span. It has a hypar shape with a slightly concave longitudinal section and a convex transversal section. From an acoustic point of view it is a poor performer in all time-windows. It causes important focusing in the third time-window towards the stage and has very few reflections falling in the back of the room, as well as few reflections in the second window.

Solution B is the best performer in the first time-window f_2. Its shape is more complex that the one shown in solution A, it has concave cross-sections over the stage and in the back of the room, but has flat and slightly convex cross-sections towards the center of the room. This interesting shape creates a very uniform sound field in the first time-window, and a pretty uniform one in the second one. The third time window shows some sound concentration near the stage due to the concave section over the sound source. From a structural point of view it is not a good performer, having a $max\Delta_Z$ of 78 millimeters, much higher than solution A. High displacements (and lack of rigidity) in the shape are possibly due to the presence of flat portions in the shape of the shell.

Solution C is the best performing solution for the second time-window f_3. It is an almost cylindrical convex surface, inclined towards the audience. It has a flat longitudinal section and only convex cross-sections. Fitness values for f_2 and f_3 are not very good, they are in fact bellow average. Too many reflections fall into the first window, and there are both focusing and dead areas in the third window. From a structural point of view, solution C is very much above average, having a $max\Delta_Z$ value of 1.5 millimeters.

Solution D has the highest fitness value in the third time-window f_3. Solution D has perhaps the most complex and pronounced shape i among the Pareto solutions. It has flat, concave and convex cross-sections, and a slightly concave longitudinal section. It has a fairly good sound distribution on the third time-window, a moderate one in the first one. It lacks a good number of reflections in the second time window, resulting in a low f_3 value. Form a structural point of view solution D is above average with a $max\Delta_Z$ of 9 millimeters.

So far we have only looked at 2D objective spaces and considered contrast between pairs of functions. If we consider all four functions, it is much more difficult to visualize the results. The Pareto front is made up of a large number of the solutions in the population, but it is very hard to come

upon good compromises for all four solutions. Solution E is perhaps the best compromise that can be found for all solutions. It has above average fitness values for all functions. Its curvature is not very pronounced but always convex, with an almost flat part towards the back of the stage. It has uniform sound distributions in all time-windows, with a slight lack of reflections close to the stage in the second window. Structurally it performs well, with a $max\Delta_Z$ value of 9 millimeters. Solution E is either on or very close to all of the 2D Pareto fronts, making it a good compromise between all functions in the problem.

19.5 Case Study 13: Masonry shell roof for a religious building.

Case study 13 is a search process based on the previous case studies on shells. Case studies 3 and 4 search for optimal compression only shapes, freeform masonry vaults with optimal structural capacities while still being as light as possible. Case study 7 on the other hand studied freeform shells form an acoustical point of view, selecting shapes that evenly distribute sound energy in time-windows and spatial subdivisions inside the room.

Case study 13 involves a religious building, not a concert hall. Many religious traditions of different faiths involve musical performances during the ceremonies, and all of them involve the listening of the spoken word. Traditional european religious buildings have very large volumes and very few absorptive materials. This results in very long reverberation times, in some cases this is used in the favor of the musical performances. Some choral and organ Christian music appears to have been conceived for this spaces, having very long pauses and slow tempo and making good use of RTs as that go well beyond 3 or 4 seconds. However, this kind of reverberant spaces result in very poor speech intelligibility. The spoken word is not easily understood in such spaces.

This case study is designed to search for shell shapes that have a good structural capacity while also distributing sound energy in such a way as to optimize speech transmission inside the room.

19.5.1 Parametric model

Figure 19.5 shows the parametric model, acoustic setup and FEM model for case study 13. The parametric model is very similar to the one used in case study 3, the shell has a 40×20 plan projection, has 5 variable control points

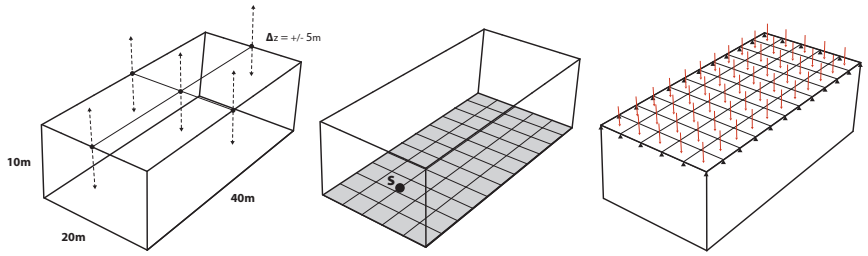


Figure 19.5: Case Study 13 Parametric Model - 40×20 Masonry Shell roof for a religious building.

and has a continuous and fixed thickness of 30cm. The base surface is 10m above the audience, but the 5 control points can move 5 meters above or bellow the base surface. These control points are the only variables in this parametric model.

The acoustic setup for case study 13 shows a single sound source placed close to one of the room's ends and in the center line. The audience area is subdivided into square segments that cover the entire room, and it does not have a pitched rake (the audience surface is flat). Sidewalls are of course parallel, and they are also perpendicular to the audience area.

The FEM model used is also the same one used for case study 3. It shows a meshed surface that is loaded in each shell element, loading is related to the surface weight. The surface is structurally constrained in all of its edges since the shell is supported by the four walls.

19.5.2 Fitness functions

Case study 13 proposes the structural and acoustic study of masonry shells for a religious building. Fitness functions for this study are taken from studies 3 for the structural shells, and 7 for the acoustic aspects. The first fitness function is the structural function developed for the study of masonry shells and described by equation 10.8 in page 163. The second and third functions are reserved for the acoustic study of the space.

Case study 7 used three time-windows to study the early reflections of a concert hall. The time-windows went from 0 to 80 ms, 80 to 120 ms and 120 to 200 ms. These time-windows were selected for the study of a shell roof meant for the enjoyment of music, while the building in this case is a

religious building. The present case study used two time-windows designed to describe early reflections for the listening of the spoken word. The first window goes from 0 to 50 ms, the 50 ms barrier has been used in acoustical parameters meant for the study spoken word (most importantly C_{50} and D_{50}). The second time-window goes from 50 to 300 ms.

The problem put forth in case study 13 can be expressed in the following way:

$$\textit{Case Study } 13 \begin{cases} \textit{Minimize} & f_1(x) = \frac{U_e}{U_{e,0}} + \left(\frac{max(\tau^+)^2}{max(\tau_0^+)^2} \cdot w \right), \\ \textit{Minimize} & f_{2(x)} = E_{tot,0-50}, \\ \textit{Minimize} & f_{3(x)} = E_{tot,50-300}, \\ \textit{subject to} & -5 \leq x_i \leq 5. \end{cases} \tag{19.3}$$

19.5.3 Genetic algorithm inputs

Genetic Inputs for case study 13 are described bellow. NSGA-II is used for 100 generations with 50 individuals in each generation. The overall genetic inputs for this case study is as follows:

Case Study 13	
Population Size (N)	50
Number of Variables	5
Number of binary digits	8
Variable Domains	$x_i \in [-5, 5]$
Mutation Probability (p_m)	0.2
End Condition	End after 100 generations

19.5.4 Results

Figure 19.6 shows the objective spaces found in case study 13 for all fitness functions as well as some significant resulting solutions. Since there are 3 fitness functions there are 3 possible combinations of them, hence we see 3 two-dimensional objective spaces. We can see that there is a some contrast in all 3 combinations, but the most contrast is found between the 2 acoustic functions. We can see that the first window function has much smaller E_{tot} values. The second window seems to be much harder to solve, this time-window has too many reflections when compared to the first one, which is to be expected.

Contrast between the acoustic and structural functions seems to be lower than expected, but if we look at the numbers carefully we can see that this is not so. The scales in the objective space can be misleading in this case, the structural function has very low values, and the differences between similar solutions is very significant. Good acoustic performers are poor structurally. Concave shells that minimize our structural function cause sound concentrations that create high error values in both acoustic functions.

Solution A is the best performing solution for the structural fitness function. Interestingly it has a negative double curvature shell. Structurally solutions found in case study 13 were inferior to those found in case studied 3 and 4. Case 4 ran for many more generations, but this is not the case in case 3. It can be argued that contrast between structural and acoustical functions is higher that the weight functions used in case 3, thus explaining the better results. In fact, none of the solutions found in case 13 resemble the high performing sail vaults found in cases 3 and 4.

Solution B has the best performance in the first acoustical function (the 0 to 50 ms time-window). It does so by having as low a roof possible while not being convex (probably for structural reasons). The low roof sends a high amount of reflections into the audience in the first 50 milliseconds. But it does have trouble reaching receivers at the end of the room within those early milliseconds. Its structural performance is not as bad as other solutions in the rest of the population, but it is not very good when compared to other case studies. It has one of the worst performances in the second acoustic function (50 to 300 ms). This is due to large sound concentrations in the center of the room in the second time-window.

Solution C is the best performer in the second acoustic function. It is a tall and double curvature surface that is concave in the longitudinal section and slightly convex in the transversal section. It has a fairly good distribution of sound energy in the second time-window in most receivers, but has some concentration near the source and near the back of the room. It has the best performance, but it is however not a perfect solution for this function. It has one of the worst distributions in the first time-window, mainly because of its height. It is too high to have a good number of reflections reach receivers within 50 ms. Its convex cross-section makes it not a good structural performer.

Solution D is a good compromise between the two acoustic functions. It is also a double curvature surface, but most importantly it has a low roof near the source and a high one near the end of the room. This seems to give it a good number of reflections in both time-windows. Spatially however it is not as good as it could be, there are some sound concentrations due to

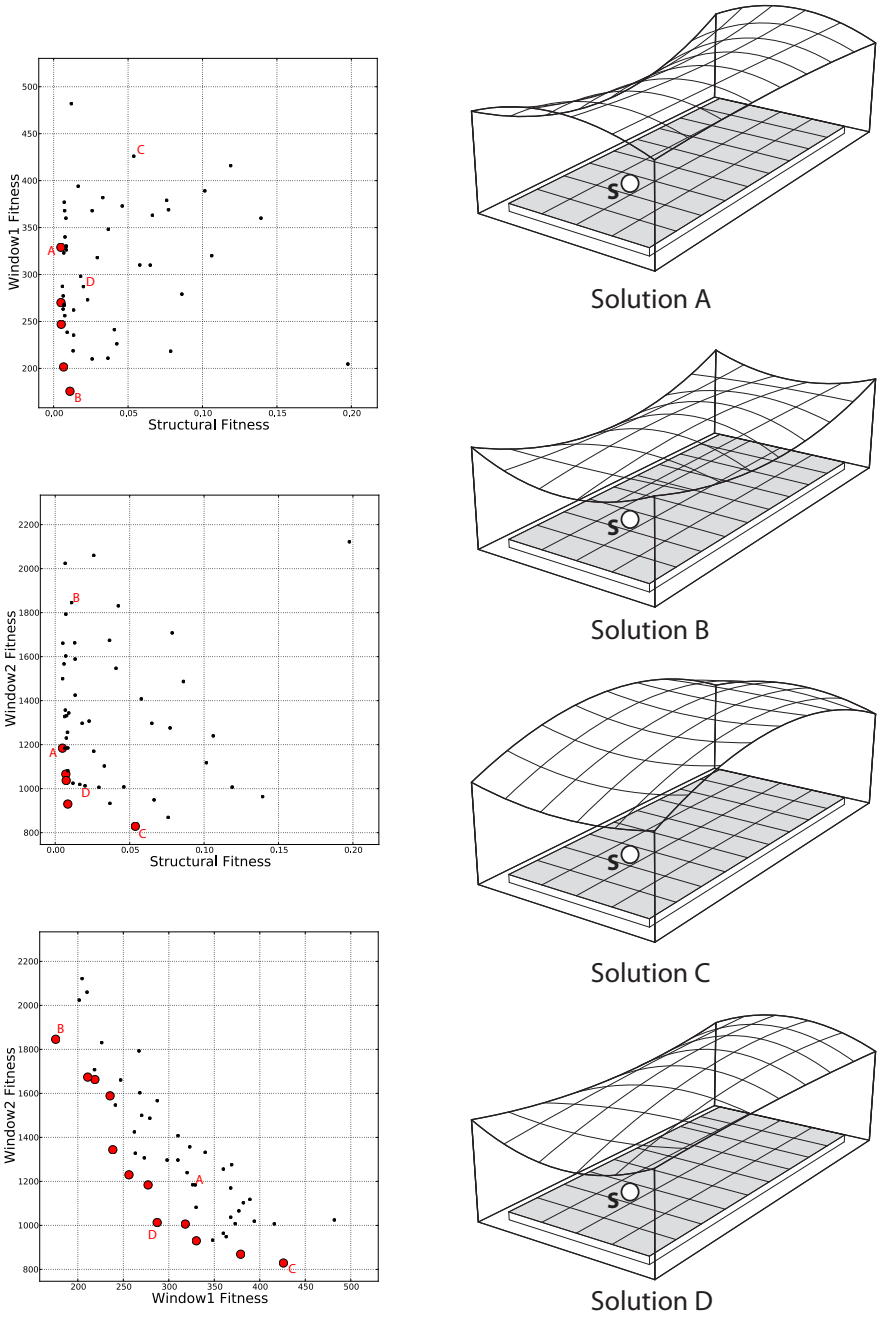


Figure 19.6: Objective spaces and Results for Case study 13.

very concave cross-sections towards the end of the room. This solution is not a very high performer in the structural function.

Conclusions

This PhD research proposes the use of computational search in the early design phase as an exploration and information gathering tool. Search algorithms in combination with parametric models and building performance simulation software are implemented and employed in problems related to structure, acoustics and energy. A theoretical framework on the use of these methods in the early phases of architectural design is presented, opportunities and limitations are outlined as well as a view in how these methods relate to existing design practices. The search algorithms employed in this research are described in detail, as well as the study of multi-objective search itself, contrast between functions is also discussed. Parametric models are introduced from many points of view, theoretical discussions on their use as well as mathematical implementations are shown. Many different parametric models containing geometries of varied nature are implemented in all of the case studies presented. Performance simulation and the physical phenomena involved are made very clear in each case study, presenting in some cases different approaches to the problem and proposing their use in search.

A clear idea on the usefulness and potential of search in architectural design is given through a good number of case studies. Ten studies are devoted to structure, acoustics and energy problems in architectural design, and three multi-disciplinary studies combine these disciplines to increase our understanding of the relationship between them. The information gained in these studies shows that search processes have much to offer designers during the early design phase, they can help designers significantly improve building performance without limiting their creativity or imposing particular solutions.

An important issue was discussed in section 1.6.3 on the dichotomy of design and instrumental knowledge and their encapsulation in software tools, as presented by Andrew Witt.

Search algorithms, parametric models and performance evaluation are instruments that do not originally come from the architecture discipline, they do not fall into the category of what Witt calls the "intrinsic" knowledge of the architect, hence they can be considered to be instrumental knowledge. The use given in this PhD thesis to such instruments is evidence of their usefulness in design processes, their ability to (i) generate practical and project specific information, as well as (ii) general knowledge on the relationship between shape and performance. This second category is not project specific, this information can be generalized and applied in other related problems. Both of this categories however constitute design knowledge, the information found during search processes is design knowledge.

One of the studies presented is intended to study architectural types, it is the parametric study of concert hall types described in section 13.3 (case study 6), where the object is to test the acoustic performance of three different types, while maintaining the same number of audience members. The intent in this case study is that of expressing the defining characteristics of each room type into their respective parametric model. Even more ambitiously, the intent is to include in the study as many instantiations of each concert hall type as possible, with the purpose of having a comprehension of the general type, not just a particular detail. This is done by using very general descriptions of each type, and not including detailed elements in the search process. No balconies, canopy reflectors or ceiling configurations were included in the search process, thus keeping the focus on general room shape, size and proportions. It would not be possible to include all of the possibilities in details pertaining to each type in such a study.

However, this does not mean that the study gives us any less of a clue as to the potential of each concert hall type. On the contrary, because of the generality of the parametric model used, the overall information found should remain largely unvaried for more detailed studies further in the design process. Not only is it possible to compare the performance of the best solutions for each type, but also strengths and weaknesses are found in each type. Contrast between acoustical objectives inside each typology is studied, as well as the geometrical characteristics that are involved in the in the contrast for each type. Good compromising solutions to contrasting objectives were found in some cases, in other cases where contrast is too strong, the new knowledge on the problem hints at possibilities of new formulations and geometric families that could reduce contrast and help us find acceptable compromising solutions.

The knowledge gained in this study is very detailed and specific to the specified requirements, it serves to improve the general knowledge architects

have of the concert hall types. We are able to quantify when a shoebox room becomes too wide to include sufficient side reflections, or what angle becomes too open for fan or hexagonal rooms. Weaknesses that are attributed to entire types are sometimes kept constrained to some versions of the type and not all of them, allowing us to consider them and not to ignore them. A clear example of this is the finding in shoebox and fan shaped rooms, not all of the fan rooms are poor distributors of early reflections, and not all shoebox rooms are good providers of early reflections.

If we consider the Pareto fronts found in case study 6 as information regarding the strengths and weaknesses of each type, the room acoustics parameters that are problematic and those which are best achieved by the room shape, a clear picture on kind of sound that is achievable by each type emerges. The information contained in the resulting data, expressed through objective spaces and solution shapes, can be of fundamental help to designers in selecting a room type very early in the design process, and provide them with specific knowledge on the challenges they face further in the process.

The parametric models used in other case studies did not have the ambition to contain or represent easily identifiable architectural types. Many of the shell studies can jump from traditional shapes such as arches or sail vaults, into hypars and free-form shapes. In these cases, the information found is not so easily attributed to known types, but comparisons between the solutions found are still very much possible. Comparisons can be made outright when the search process opted not to chose solutions from a given type, or selected solutions of only one type. In these cases we get information not only on specific solutions, but on the types involved.

Case studies 7 and 12 reveal that not all concave surfaces are detrimental to the acoustic quality of concert spaces, and that in fact some of them help better distribute early sound reflections more uniformly over the audience area. Moreover, case 12 shows that the combination of concave and convex section present in double curvature surfaces can be mutually beneficial to structural and acoustic performances.

The case studies involving the fenestration arrangements of rectangular buildings are also interesting from a typology point of view, in the sense that they can be considered to cover only a specific part of a known building type. The window configurations in study in this case studies do not have a big impact on the general type of the building, but as results show, they have a big impact on their structural and energy efficiencies. In this cases we can safely say that the information found serves as a a guide of the possibilities of the type, while not describing *all* possibilities. Findings in this case study

illustrate only a small part of the design possibilities in the type, when compared to results found in other studies.

An interesting result is given by case study 8, in which the proportions and orientation of an office building are investigated for their energy efficiency. Because this study uses a parametric models that changes the general shape of the building, strong changes in energy efficiency were expected. But only very minor changes in energy use are shown by the performance evaluation, due to the importance of the fenestration. This result is both a warning and an opportunity for designers. The selection of shape and orientation is shown to be insufficient to guarantee a good energy performance, fenestration also needs to be considered together with the shape and orientation. Therefore there are two possibilities as to how to interpret this information: (i) designers can consider both shape and fenestration in a more detailed search process, or (ii) designers can select shape considering other performance values or implicit design goals, and leave the design of the fenestration for a later stage of design.

The first alternative presents the opportunity to create solutions that perform much better when compared to solutions that can be generated by the second alternative. If we think of this in terms of search spaces we can say that the first alternative contains a much larger search space than the second. Because in the second alternative energy efficiency is not studied until the general shape is fixed, the search space is confined to the fenestration possibilities that can be generated with that shape. While in the first one, the search space considers the fenestration solutions that can be generated with that shape as well as many others. Since the energy efficiency is strongly determined by the *combination* the shape and the fenestration, it is very likely that the larger search space contains solutions that a far more efficient than those contained by the second.

Future Work

The parametric models employed in this thesis show a small part of the wide range of geometric possibilities that can be achieved with the use of parametric models. The parametrization of geometry is in no way a limit in the exploration of shapes for architectural design. However, the creation of parametric models for search processes during the early phase of design does presents some challenges that could be subject of future research.

The design process is most commonly subjected to time constraints, designers need to make decisions in short periods of time. The use of software certainly helps speed up the design process with more efficient representation

methods, and to help increase the quality of the building by means of search processes such as those proposed in this research. However, the creation of parametric models is time consuming, and if the changes desired by designers at any point during the process are not contained among the variables of the model, the model needs to be re-written. This issue is discussed at length by Daniel Davies in his PhD dissertation (Davis 2013). Davis's work addresses techniques for generating more flexible parametric models. The use of such techniques or the creation of other is important subject matter for the improvement of search methods in architectural design.

The use of parametric models in search processes is also subject to problems in their coding strategies as shown in section 11.2.1. There are many other issues related to the models and their coding that are not discussed in this PhD thesis, such as "epistasis". The study of efficient coding strategies could possibly arrive at general and practical information that can be applied by designers in many different models, in order to help them avoid search problems. Coding problems are certainly specific to the type of search algorithm being used.

This PhD research employed only one kind of search algorithm, the genetic algorithm. It also employed the same kind of GA and always used the same operators for all case studies. A comprehensive comparison of different search algorithms, operators and search inputs can also be of great help. The relative efficiency of the algorithms can be established in relation to each other when applied to architectural search problems. Algorithm efficiency is thus related to speed and convergence, how fast does the algorithm find the real Pareto front for example. Robustness of the algorithms is also an important issue to study, how the algorithms perform under very complex and different problems. All of this issued relate to exploration and exploitation, the balance between these two is of outmost importance, and it is determined not only by the algorithms themselves, but also by the search inputs we give them (e.g. in genetic algorithms, the number of individuals, generations, mutation probability and the genetic operators chosen).

In the first part of the thesis, interactivity was signaled as an important characteristic of the search process due to the nature of architectural problems. The case studies presented interactive features only before and after the design process. An interactive parametrization method made for the purposes of interaction *during* the search process was partially developed for this thesis, but it did not produce sufficient results for it to be included in this dissertation. Further research on interaction during the design process is certainly an important step. This issue relates closely with the time consumption in the creation of parametric models and their flexibility. In

order for interaction to be present during the search process, designers must be able to modify parametric models in a very quick way, almost in real time.

Case studies in this dissertation involve three very different disciplines, making use of search processes to involve them directly in early in design. Many discipline-specific research can also be done in the future.

The building envelope is a subject of study that is rich with possibilities for a search process, and especially so for multi-disciplinary work. Acoustic insulation was not a topic of study in this work. Sound transmission is very much related to structural integrity, in that rigid structures tend to have smaller vibrations and in higher frequencies. Sound insulation is addressed many ways, one such approach os that of having very massive elements, rigidity is achieved by having massive envelopes. Another possibility is to generate envelopes that are more rigid because of their shape, thus further improving both structural capacity and reducing vibration transmission. Structural analysis can be incorporated with acoustical models that can help shed light on the interaction between sound waves and the vibration of building envelopes. The opportunity of embedding structural, acoustical and energy performances in architectural shapes is very appealing. Normally, envelopes are made up of a big number of separate components that achieve performances on their own, thus decoupling efficiency with shape, and this is arguably not interesting architecture.

Complex shapes were not employed in energy studies in this thesis. This is so because it was not clear, during the development of this research, whether existing energy performance calculation software are able to reproduce, in a sufficiently accurate way, the physical phenomena involved in the transmission of heat for complex shapes. Adequate modeling is certainly possible with computationally expensive techniques such as Computational Fluid Dynamics (CFD). Most of these methods model only one aspect of the problem, CDFs for example do not model radiative transmission of heat. Coupled analysis is a technique that puts together many models in order to create a complete simulation of the physical phenomena involved. This kind of analysis is surely very time consuming and not ideal for search processes, but it might be a good place to start. The opportunities of complex shapes in creating energy efficient buildings are an interesting enough subject to warrant such research.

List of Figures

Bibliography

Abdou, A. & Guy, R. W. (1996), 'Spatial information of sound fields for roomacoustics evaluation and diagnosis', *The Journal of the Acoustical Society of America* **100**(5), 3215–3226.

Akama, T., Suzuki, H. & Omoto, A. (2010), 'Distribution of selected monaural acoustical parameters in concert halls', *Applied Acoustics* **71**, 564–577.

Akin, O. (1986), *Psychology of Architectural Design*, Pion.

Alberti, L. B. (1966), *L'architettura De re aedificatoria, 1452, testo latino e traduzione di Giovanni Orlandi, con introduzione e note di Paolo Portoghesi*, Edizioni Il Polifilo, Milano.

Alexander, C. (1966), *Notes on the Synthesis of Form*, Harvard University Press, Cambridge.

Ando, Y. (1983), 'Calculation of subjective preference at each seat in a concert hall', *The Journal of the Acoustical Society of America* **74**(3), 873–887.

Ando, Y. (2007), Concert hall acoustics based on subjective preference theory, *in* T. D. Rossing, ed., 'Springer Handbook of Acoustics', Springer, New York, chapter 10, pp. 351–386.

Argan, G. C. (1963), 'On the typology of architecture', *Architectural Design* (33).

Atal, B., Schroeder, M. & Sessler, G. (1965), Subjective reverberation time and its relation to sound decay, *in* '5th INternational Congress of Acoustics', number G32.

Attia, S. (2012), A Tool for Design Decision Making, Zero Energy Residential Buildings in Hot Humid Climates, PhD thesis, Universitè catholique de Louvain, Louvain La Neuve.

Attia, S. & De Herde, A. (2011), 'Early design simulation tools for net zero energy buildings: a comparison of ten tools', *Proceedings of Building Simulation* pp. 14–16.

Attia, S., Hamdy, M., O'Brien, W. & Carlucci, S. (2013), 'Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design', *Energy and Buildings* **60**(0), 110 – 124.

Balmond, C. (2012), *Informal*, Prestel, Berlin.

Barron, M. (1971), 'The subjective effects of first reflections in concert halls—the need for lateral reflections', *Journal of Sound and Vibration* **15**(4), 475 – 494.

Barron, M. (1995), 'Interpretation of early decay times in concert auditoria', *Acustica* **81**, 320–331.

Barron, M. (2005), 'Using the standard on objective measures for concert auditoria, iso 3382, to give reliable results', *Acoust Sci Techno* **26**(2), 162–169.

Barron, M. (2009*a*), *Auditorium Acousitcs and Architectural Design*, Spon Press, London.

Barron, M. (2009*b*), When is a concert hall too quiet?, *in* 'Proceedings of the 19th International Congress on Acoustics', ICA, International Congress on Acoustics.

Barron, M. (2013), 'Objective assessment of concert hall acoustics using temporal energy analysis', *Applied Acoustics* **74**(7), 936 – 944.

Barron, M. & Marshall, A. H. (1981), 'Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure', *Journal of Sound and Vibration* **77**(2), 211 – 232.

Bassuet, A. (2011), 'New acoustical parameters and visualization techniques to analyse the spatial distribution of sound in music spaces', *Building Acoustics* **18**(3), 329–347.

Beranek, L. (2003), 'Subjective rank-orderings and acoustical measurements for fifty-eight concert halls', *Acta Acustica united with Acustica* **89**, 494–508.

Beranek, L. (2008), 'Concert hall acoustics - 2008*', *Journal of Audio Engineering* **56**(7/8), 532–544.

Beranek, L. (2011), 'The sound strength parameter g and its importance in evaluating and planning the acoustics of halls for musica)', *The Journal of the Acoustical Society of America* **129**(5), 3020–3026.

Beranek, L., Jaffe, J. C., Nakajima, T., Kahle, Eckhard Kirkegaard, R. L. & Clements, P. (2010), How acousticians listen, *in* 'Proceedings of the International Symposium on Room Acoustics, ISRA 2010', Melbourne, Australia.

Beranek, L. L. (2004), *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, Springer, New York.

Blickle, T. & Thiele, L. (1995), A comparison of selection schemes used in genetic algorithms, Technical report, Gloriastrasse 35, CH-8092 Zurich: Swiss Federal Institute of Technology (ETH) Zurich, Computer Engineering and Communications Networks Lab (TIK, Zurich.

Block, P. (2009), Thrust Network Analysis, PhD thesis, Massachusetts Institute of Technology, Cambridge.

Bogar, D., Rapone, G., Mahdavi, A. & Saro, O. (2013), Ga-optimization of a curtain wall façade for different orientations and climates, *in* 'Proceedings of IBPSA Italy BSA 2013', IBPSA.

Bollinger, K., Grohmann, M. & Tessmann, O. (2010), 'Truss structures - from typology to adaptation', *Journal of the International Association for Shells and Spatial Structures* **51**(165), 190–194.

Bolt, R. & Doak, P. (1950), 'A tentative criterion for the short-term transient response of auditoriums', *The Journal of the Acoustical Society of America* **22**(4), 507–509.

Borong, L., Qiong, Y., Ziwei, L. & Xiaoru, Z. (2013), 'Research on parametric design method for energy efficiency of green building in architectural scheme phase', *Frontiers of Architectural Research* **2**(1), 11 – 22.

Bradley, J. (2011), 'Review of objective room acoustics measures and future needs', *Applied Acoustics* **72**(10), 713 – 720.

Bradley, J. S. (1994), 'Comparison of concert hall measurements of spatial impression', *The Journal of the Acoustical Society of America* **96**(6), 3525–3535.

Bradley, J. S. (2004), 'Using iso 3382 measures, and their extensions, to evaluate acoustical conditions in concert halls', *Acoust Sci Techno* **26**(2), 170–178.

Brooks, F. P. J. (2010), *The design of design: Essays from a computer scientist*, Addison-Wesley, Boston.

Burry, M. (2011), *Scripting Cultures*, John Wiley & Sons, Chichester.

Cache, B. (2009), 'De architectura. on the table of content of the ten books on architecture', *Candide. Journal for Architectural Knowledge* **1**, 9–48.

Carpo, M. (2003), 'Drawing with numbers: Geometry and numeracy in early modern architectural design', *Journal of the Society of Architectural Historians* **62**(4), 448–469.

Carpo, M. (2011), *The Alphabet and the Algorithm*, The MIT Press, Cambridge.

Carpo, M., ed. (2013), *The Digital Turn in Architecture 1992-2012*, John Wiley & Sons, Chichester.

Cerdá, S., Giménez, A., Romero, J., Cibrián, R. & Miralles, J. (2009), 'Room acoustical parameters: A factor analysis approach', *Applied Acoustics* **70**(1), 97–1–9.

Ceruzzi, P. (2003), *A History of Modern Computing*, 2nd edn, The MIT Press, Cambridge.

Chase, S. (2002), 'A model for user interaction in grammar-based design systems', *Automation in Construction* **11**(2), 161–172.

Chase, S. (2005), 'Generative design tools for novice designers: Issues for selection', *Automation in Construction* **14**(6), 689–698.

Cingolani, S. & Spagnolo, R., eds (2005), *Acustica Musicale e Architettonica*, Utet, Torino.

Coello Coello, C., van Veldhuizen, D. & Lamont, G. (2002), *Evolutionary algorithms for solving multiobjective problems*, Kluwer academic publishers, New York.

Colquhoun, A. (1969), 'Typology and design method', *Perpecta* **12**.

Corrado, V. & Paduos, S. (2010), *La Nuova legistlazione sull'efficienza energetica degli edifici, requisiti e metodi di calcolo*, Celid, Torino.

Cross, N. (2007), 'Forty years of design research, design studies', *Design Studies* **28**(1).

Cui, C., Ohmori, H. & Sasaki, M. (2003), 'Computational morphogenesis of 3d structures by extended eso method,', *Journal of the International Association for Shells and Spatial Structures* **44**(141), 51–61.

Dalenback, B. I. L. (1996), 'Room acoustic prediction based on a unified treatment of diffuse and specular reflection', *The Journal of the Acoustical Society of America* **100**(2), 899–909.

Davelaar, E. J. & Raaijmakers, J. G. W. (2012), Human memory search, *in* P. M. Todd, T. T. Hills & T. W. Robbins, eds, 'Cognitive Search, Evolution Algorithms and the Brain', The MIT Press, Cambridge.

Davis, D. (2013), Modelled on Software Engineering; Flexible Parametric Models in the Practice of Architecture, PhD thesis, School of Architecture and Design; RMIT University, Melbourne.

Day, A. (1965), 'An introduction to dynamic relaxation', *The Engineer* **219**.

de Vries, D., Hulsebos, E. M. & Baan, J. (2001), 'Spatial fluctuations in measures for spaciousness', *The Journal of the Acoustical Society of America* **110**(2), 947–954.

Deb, K. (2001), *Multi-Objective Optimization using Evolutionary Algorithms*, John Wiley & Sons, Chichester.

Derix, C. (2010), Mediating spatial phenomena through computational heuristics, *in* 'Proceedings of the 30th Annual Conference of the Association for Computer Aided Design in Architecture', Vol. Proceedings of the 30th Annual Conference, Association for Computer Aided Design in Architecture (ACADIA), New York.

Edelkamp, S. & Schrödl, S. (2012), *Heuristic Search*, Morgan Kaufmann Publishers, Waltham.

Fanger, P. (1970), *Thermal comfort: Analysis and applications in environmental engineering*, Danish Technical Press.

Floreano, D. & Mattiussi, C. (2002(1996)), *Manuale sulle Reti Neurali*, expanded and fully revised edn, Il Mulino, Bologna.

Fonseca, C. & Fleming, P. (1998), 'Multi objective optimization and multiple constraint handling with evolutionary algorithms i: A unified formulation', *IEEE Transactions on Systems, Man, and Cybernetics* **28**(1), 26–37.

Frazer, J. H. (1995), *An Evolutionary Architecture*, Architectural Publications Association, London.

Fry, B. (2008), *Visualizing Data*, O'Reilly, Sebastopol CA.

Fujii, K., Hotehama, T., Kato, K., Shimokura, R., Okamoto, Y., Suzumura, Y. & Ando, Y. (2004), 'Spatial distribution of acoustical parameters in concert halls: Comparison of different scattered reflections', *Journal of Temporal Design in Architecture and the Environment* **4**(1), 59–68.

Gade, A. C. (1989), Acoustical survey of eleven european concert halls, Technical Report 44, The Acoustics Laboratory, Tecnical University of Denmark, Copenhagen.

Gagne, J. M. L. & Andersen, M. (2012), 'A Generative Façade Design Method Based on Daylighting Performance Goals', *Journal of Building Performance Simulation* **5**(3), 141–154.

Glymph, J., Shelden, D., Ceccato, C., Mussel, J. & Schober, H. (2004), 'A parametric strategy for free-form glass structures using quadrilateral planar facets,', *Automation in Construction* **13**(2), 187–202.

Goia, F., Haase, M. & Perino, M. (2012), Optimal transparent percentage in façade modules for office buildings in different european climates, *in* 'EuroSun 2012 - ISES-Europe Solar Conference', ISES, Rijeka (Croatia), pp. 1–8.

Goldberg, D. (1989), *Genetic algorithms in Search, Optimization & Machine Learning*, 1st edn, Addison-Wesley Professional, Boston.

Goldsmith, M. (1953), The tall building: The effects of scale, Master's thesis, Illinois Institute of Technology, Chicago, Illinois.

Goodman, N. (1976), *Languages of Art: An Approach to a Theory of Symbols*, Hackett Publishing.

Haas, H. (1951), 'Uber den einfluss eines einfachechos auf die horsamkeit von sprache', *Acustica* **1**(49).

Hamdy, M., Hasan, A. & Siren, K. (2011), 'Applying a multi-objective optimization approach for design of low-emission cost-effective dwellings', *Building and Environment* **46**(1), 109 – 123.

Heyman, J. (1995), *The Stone Skeleton: Structural Engineering of Masonry Architecture*, Cambridge University Press.

Hidaka, T., Beranek, L. & Nishihara, N. (2008), A comparison between shoebox and non-shoebox halls based on objective measurements in actual halls, *in* 'Proceedings of Acoustics 08 Paris', Paris, pp. 327– 332.

Hills, T. T. & Dukas, R. (2012), The evolution of congnitive search, *in* P. M. Todd, T. T. Hills & T. W. Robbins, eds, 'Cognitive Search, Evolution Algorithms and the Brain', The MIT Press, Cambridge.

Hofstadter, D. (1982), 'Variations on a theme as the crux of creativity', *Scientific American* .

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems*, 1st edn, The University of Michigan, Ann Arbor.

Hutchinson, J. M. C., Stephens, D. W., Bateson, M., Couzin, I., Dukas, R., Giraldeau, L.-A., Hills, T. T., Mery, F. & Winterhalder, B. (2012), Searching for fundamentals and commonalities of search, *in* P. M. Todd, T. T. Hills & T. W. Robbins, eds, 'Cognitive Search, Evolution Algorithms and the Brain', The MIT Press, Cambridge.

ISO 3382-1:2009 (International Standards Organization, 2009), *Acoustics-Measurement of Room Acoustic Parameters I: Performance Spaces*.

Jordan, V. (1970), 'Acoustical criteria for auditoriums and their relation to model techniques', *The Journal of the Acoustical Society of America* **47**, 408–412.

Jurkiewicz, Y. & Kahle, E. (2008), Early reflection surfaces in concert halls - a new quantitative criterion, *in* 'Proceedings of Acoustics 08 Paris', Paris, pp. 6489–6494.

Jurkiewicz, Y., Wulfrank, T. & Kahle, E. (2012), 'Architectural shape and early acoustic efficiency in concert halls (l)', *The Journal of the Acoustical Society of America* **132**(3), 1253–1256.

Kahle-Acoustics & Altia-Acoustique (2006), Acoustic brief, Technical report, Philarmonie de Paris.

Kämpf, J. H. & Robinson, D. (2010), 'Optimisation of building form for solar energy utilisation using constrained evolutionary algorithms', *Energy and Buildings* **42**(6), 807 – 814.

Kilian, A. (2000), Design Exploration trough Bidirectional Modeling of Constraints, PhD thesis, Massachusetts Institute of Technology, Cambridge.

Kilian, A. & Ochsendorf, J. (2005), 'Particle-spring systems for strunctural form finding', *Journal of the International Association for Shell and Spatial Structures* **46**(147).

Klockner, A. (n.d.), 'Meshpy'.
URL: *http://mathema.tician.de/software/meshpy*

Klosak, A. K. & Gade, A. C. (2008), Relationship between room shape and acoustics of rectangular concert halls, *in* 'Proceedings of Acoustics 08 Paris', Paris, pp. 2163–2168.

Koza, J. (1992), *Genetic Programming: On the programming of computers by means of natural selection*, 1st edn, The MIT Press, Cambridge.

Krokstad, A., Strom, S. & Sørsdal, S. (1968), 'Calculating the acoustical room response by the use of a ray tracing technique', *Journal of Sound and Vibration* **8**(1), 118–125.

Kuttruff, H. (1993), 'Some remarks on the simulation of sound reflection from curved walls', *Acustica* **77**.

Kuttruff, H. (2000), *Room Acoustics*, 4th edn, Spon, New York.

Lacatis, R., Giménez, A., Barba Sevillano, A., Cerdá, S., Romero, J. & Cibrián, R. (2008), Historical and chronological evolution of the concert hall acoustics parameters, *in* 'Proceedings of Acoustics 08 Paris', Paris.

Lam, Y. W. (1996), 'The dependence of diffusion parameters in a room acoustics prediction model on auditorium sizes and shapes', *The Journal of the Acoustical Society of America* **100**(4), 2193–2203.

Lawson, B. (2006), *How Designers Think: The Design Process Demystified*, Architectural Press.

Llinares, J., Llopis, A. & Sancho, J. (1996), *Acustica Arquitectonica y Urbanistica*, Servicio de Publicaciones - Universidad Politecnica de Valencia, Valencia.

Lokki, T., Patynen, J., Tervo, S., Siltanen, S. & Savioja, L. (2012), 'Engaging concert hall acoustics is made up of temporal envelope preserving reflections', *The Journal of the Acoustical Society of America* **129 - Express Letters**(6), 1–10.

Marshall, A. (2001), 'How to make a frank gehry building', *New York Times Magazine* .

Marshall, L. G. (1994), 'An acoustics measurement program for evaluating auditoriums based on the early/late sound energy ratio', *The Journal of the Acoustical Society of America* **96**(4), 2251–2261.

Mechri, H. E., Capozzoli, A. & Corrado, V. (2010), '{USE} of the {ANOVA} approach for sensitive building energy design', *Applied Energy* **87**(10), 3073 – 3083.

Méndez Echenagucia, T., Astolfi, A., Jansen, M. & Sassone, M. (2008*a*), 'Architectural acoustic and structural form', *Journal of the International Association for Shell and Spatial Structures* **49**(3), 181–186.

Méndez Echenagucia, T., Astolfi, A., Jansen, M. & Sassone, M. (2008*b*), Architectural acoustic and structural form, *in* IASS, ed., 'Shell and Spatial Structures: New Materials and Technologies, New Designs and Innovations- A Sustainable Approach to Architectural and Structural Design', IASS, Acapulco-Mexico.

Méndez Echenagucia, T., Astolfi, A., Shtrepi, L., van der Harten, A. & Sassone, M. (2012), Esplorazione multi obiettivo nella progettazione acustica architettonica, *in* '39 ° Convegno Nazionale dell'Associazione Italiana di Acustica', AIA, Rome-Italy.

Méndez Echenagucia, T., Astolfi, A., Shtrepi, L., van der Harten, A. & Sassone, M. (2013*a*), EDT, C_{80} and G driven auditorium design, *in* 'Proceedings ISRA 2013 Toronto', Toronto-Canada.

Méndez Echenagucia, T., Astolfi, A., Shtrepi, L., van der Harten, A. & Sassone, M. (2013*b*), Interactive design methods for complex curved reflectors in concert halls, *in* 'Proceedings ISRA 2013 Toronto', Toronto-Canada.

Méndez Echenagucia, T., Astolfi, A., Shtrepi, L., van der Harten, A. & Sassone, M. (2013*c*), NURBS and mesh geometry in room acoustic ray-tracing simulation, *in* 'Proceedings of the AIA-DAGA 2013 conference Merano', Merano-Italy.

Méndez Echenagucia, T., Pugnale, A. & Sassone, M. (2013), Multi-objective optimization of concrete shells, *in* P. J. Cruz, ed., 'Structures and Architecture. Concepts, Application and Challenges', CRC Press/Balkema, Leiden, pp. 217–218.

Méndez Echenagucia, T. & Sassone, M. (2012), Multi-objective oriented design of shell structures, *in* S. D. Kim, ed., 'IASS-APCS 2012 From Spatial Structures to Space Structures', IASS-APCS, Seoul-South Korea.

Méndez Echenagucia, T., Sassone, M., Astolfi, A. & Croset, P. A. (2013), Multi-objective search in the early phase of architectural design, *in* P. J. Cruz, ed., 'Structures and Architecture. Concepts, Application and Challenges', CRC Press/Balkema, Leiden, pp. 341–342.

Méndez Echenagucia, T., Sassone, M. & Pugnale, A. (2014- IN PRINT), Computational morphogenesis, *in* S. Adriaenssens, P. Block, D. Veenendaal & C. Williams, eds, 'Shells for Architecture - Form finding and structural optimization', Routledge Architecture, chapter 17.

Menges, A. & Ahlquist, S., eds (2011), *Computational Design Thinking*, John Wiley & Sons, Chichester.

Meyer, J. (2013), Concert hall design 1960 - 2010 – a historical review, *in* 'Proceedings of the AIA-DAGA 2013 conference Merano', pp. 992–993.

Miles, J., Sisk, G. & Moore, C. (2001), 'The conceptual design of commercial buildings using a genetic algorithm', *Computers & Structures* **79**(17), 1583 – 1592.

Mommertz, M. (1995), Simulation der schallubertragung in räumen mit gekrummten wandflächen., *in* 'Fortschritte der akustik - DAGA 1995'.

Monks, M., Oh, B. & Dorsey, J. (2000), 'Audioptimization: Goal-based acoustic design', *IEEE ComputerGraphicsandApplications* **20**(3), 76–90.

Ochoa, C. E., Aries, M. B., van Loenen, E. J. & Hensen, J. L. (2012), 'Considerations on design optimization criteria for windows providing low energy consumption and high visual comfort', *Applied Energy* **95**(0), 238 – 245.

Oguchi, K., Toyota, Y. & Nagata, N. (1988), 'A study on the characteristics of early reflections in concert halls', *The Journal of the Acoustical Society of America* **184**(S130).

Otto, F. & Rasch, B. (1996), *Finding Form: Towards an Architecture of the Minimal*, Edition Axel Menges.

Oxman, R. (2010), 'Morphogenesis in the theory and methodology of digital tectonics', *Journal of the International Association for Shells and Spatial Structures* **51**(165), 195–205.

Patynen, J., Tervo, S. & Lokki, T. (2013), 'Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses', *The Journal of the Acoustical Society of America* **133**(2), 842–857.

Paulson, B. C. (1976), 'Designing to reduce construction costs', *Journal of the Construction Division* **102**(4), 587–592.

Peel, M. C., Finlayson, B. L. & McMahon, T. A. (2007), 'Updated world map of the köppen-geiger climate classification', *Hydrology and Earth System Sciences* **11**(5), 1633–1644.

Pelorson, X., Vian, J.-P. & Polack, J.-D. (1992), 'On the variability of room acoustical parameters: Reproducibility and statistical validity', *Applied Acoustics* **37**(3), 175 – 198.

Petersen, S. & Svendsen, S. (2010), 'Method and simulation program informed decisions in the early stages of building design', *Energy and Buildings* **42**(7), 1113–1119.

Picon, A. (2010), *Digital culture in Architecture*, Birkhauser Verlag, Basel.

Pugnale, A. & Sassone, M. (2007), Morphogenesis and structural optimization of shell structures with the aid of a genetic algorithm, *in* 'IASS Symposium 2007. Shell and Spatial Structures: Structural Architecture - Towards the future looking to the past', IASS, Venezia, pp. 289–290.

Rapone, G. & Saro, O. (2012), 'Optimisation of curtain wall façades for office buildings by means of {PSO} algorithm', *Energy and Buildings* **45**(0), 189 – 196.

Rapone, G., Saro, O. & Zemella, G. (2013), Multi-objective optimization of external shading devices for energy efficiency and visual confort, *in* 'Proceedings of IBPSA Italy BSA 2013', IBPSA.

Redish, D. A. (2012), Search processes and hippocampus, *in* P. M. Todd, T. T. Hills & T. W. Robbins, eds, 'Cognitive Search, Evolution Algorithms and the Brain', The MIT Press, Cambridge.

Rittel, H. (1972), 'On the planning crisis :systems analysis of the 'first and second generations", *Bedriftsokonomen* **8**, 390–396.

Rittel, H. & Webber, M. (1973), 'Dilemmas in general theory of planning', *Policy Sciences* **4**, 155–169.

Rocker, I. (2008), *Versioning: Architecture as series?*, Graduate School of Design, Harvard University, Cambridge.

Rocker, I., Picon, A., Carpo, M. & Meredith, M. (2011), 'The eclipse of beauty: Parametric beauty', Lecture video.
URL: *http://www.gsd.harvard.edu/#/media/the–eclipse–of–beauty–parametric–beauty –mario–carpo–michael.html*

Rogers, D. (2001), *An introduction to NURBS: With Historical Perspective*, Morgan Kaufmann Publishers, San Francisco.

Rossi, A. (1975), Tipologia, manualistica e architettura, *in* 'Aldo Rossi, Scritti scelti sull'architettura e la città', CLUP, Milano.

Rossing, T. D., ed. (2007), *Springer Handbook of Acoustics*, Springer, New York.

Sasaki, M. (2005), *Flux Structure*, TOTO, Tokyo.

Sassone, M., Méndez Echenagucia, T. & Pugnale, A. (2008), On the interaction between architecture and engineering: the acoustic optimization of a reinforced concrete shell, *in* I. IACM/USACM, ed., 'Proceedings of the 6th International Conference on Computation of Shell and Spatial Structures: IASS-IACM 2008 - "Spanning Nano to Mega"', IASS IACM, p. 231.

Sato, S. i., Hayashi, T., Takizawa, A., Tani, A. & Kawamura, H. (2004), 'Acoustic Design of Theatres Applying Genetic Algorithms', *Journal of temporal design in Architecture and the environment* .

Schec, H. (1974), 'Force density methods for form finding and computation of genral networks', *Computer Methods in Applied Mechanics and Engineering* .

Schooler, L. J., Burbess, C., Goldstone, R. L., Fu, W.-T., Gavrilets, S., Lazer, D., Marshall, J. A., Neumann, F. & Wiener, J. M. (2012), Search environments, representation and endoding, *in* P. M. Todd, T. T. Hills & T. W. Robbins, eds, 'Cognitive Search, Evolution Algorithms and the Brain', The MIT Press, Cambridge.

Scolari, M. (2012), *Oblique Drawing: A History of anti-perspective*, The MIT Press, Cambridge.

Shewchuk, J. R. (1996), Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator, *in* M. C. Lin & D. Manocha, eds, 'Applied Computational Geometry: Towards Geometric Engineering', Vol. 1148 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 203–222. From the First ACM Workshop on Applied Computational Geometry.

Stratil, S. (2010), 'Digital master builders evolutionary formfinding in the information age', *Journal of the International Association for Shells and Spatial Structures* **51**(165), 232–240.

Suga, K., Kato, S. & Hiyama, K. (2010), 'Structural analysis of pareto-optimal solution sets for multi-objective optimization: An application to outer window design problems using multiple objective genetic algorithms', *Building and Environment* **45**(5), 1144 – 1152.

Tang, P.-N., Steinbach, M. & Kumar, V. (2005), *Introduction to Data mining*, Addison-Wesley.

Tibbits, S., van der Harten, A. & Baer, S. (2011), 'Python for rhinoceros 5'. **URL:** *http://python.rhino3d.com/content/130-RhinoPython-primer*

Todd, P. M., Hills, T. T. & Robbins, T. W. (2012*a*), Building a foundation for cognitive search, *in* P. M. Todd, T. T. Hills & T. W. Robbins, eds, 'Cognitive Search, Evolution Algorithms and the Brain', The MIT Press, Cambridge.

Todd, P. M., Hills, T. T. & Robbins, T. W., eds (2012*b*), *Cognitive Search, Evolution Algorithms and the Brain*, The MIT Press, Cambridge.

Tomasoni, E. (2008), Le Volte in muratura negli edifici storici: tecniche costruttive e comportamento strutturale, PhD thesis, Universitá degli Studi di Trento.

Turrin, M., Kilian, A., Stouffs, R. & Sariyildiz, S. (2009), Digital design exploration of structural morphologies integrating adaptable modules, *in* 'Cultures and Visions: CAADFutures 2009', CAAD, pp. 800–814.

Turrin, M., von Buelow, P. & Stouffs, R. (2011), 'Design explorations of performance driven geometry in architectural design using parametric modeling and genetic algorithms', *Advanced Engineering Informatics* **25**(4), 656 – 675. ¡ce:title¿Special Section: Advances and Challenges in Computing in Civil and Building Engineering¡/ce:title¿.

US Department of Energy (2013), 'Energyplus documentation'.
URL: *http://apps1.eere.energy.gov/buildings/energyplus/pdfs/gettingstarted.pdf*

van der Harten, A. (2011), 'Customized room acoustics simulations using scripting interfaces.', *The Journal of the Acoustical Society of America* **129**(4), 2366–2366.

Vercammen, M. (2008), The reflected sound field by curved surfaces, *in* 'Proceedings of Acoustics 08 Paris', Paris, pp. 3473– 3478.

Vercammen, M. (2010), 'Sound reflections from concave spherical surfaces. part ii: Geometrical acoustics and engineering approach', *Acta Acustica united with Acustica* **96**, 92–101.

Vercammen, M. (2012), Sound Concentration Caused by curved Surfaces, PhD thesis, Eindhoven University of Technology.

Vidler, A. (1996), The third typology, *in* K. Nesbitt, ed., 'Theorizing a new agenda for architecture', Princeton Architectural Press, New York.

Vorlander, M. (2008), *Auralization:Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer-Verlag, Berlin Heidelberg.

Vorlander, M. (2011), Models and algorithms for computer simulations in room acoustics, *in* 'Proceedings of the International Seminar on Virtual Acoustics', Valencia, pp. 72–82.

Walz, A., Kilian, A. & Schindler, S. (2009), 'Programming knowledge', *Candide. Journal for Architectural Knowledge* **1**, 49–68.

Wang, W., Rivard, H. & Zmeureanu, R. (2006), 'Floor shape optimization for green building design', *Advanced Engineering Informatics* **20**(4), 363 – 378.

Wang, W., Zmeureanu, R. & Rivard, H. (2005), 'Applying multi-objective genetic algorithms in green building design optimization', *Building and Environment* **40**(11), 1512 – 1525.

Wei Whiting, E. J. (2012), Design of Structurally-Sound Masonry Buildings Using 3d Static Analysis, PhD thesis, Massachusetts Institute of Technology, Cambridge.

Wetter, M. & Wright, J. (2004), 'A comparison of deterministic and probabilistic optimization algorithms for nonsmooth simulation-based optimization', *Building and Environment* **39**(8), 989 – 999. ¡ce:title¿Building Simulation for Better Building Design¡/ce:title¿.

Witt, A. J. (2010), 'A machine epistemology in architecture. encapsulated knowledge and the instrumentation of design', *Candide. Journal for Architectural Knowledge* (02), 37 – 88.

Wright, J. & Mourshed, M. (2009), Geometric optimization of fenestration, *in* 'Proceedings of the Eleventh International IBPSA Conference', IBPSA, Glasgow, Scotland, pp. pp.920–927.

Xiangyang, Z., Ke'an, C. & Jincai, S. (2003), 'On the accuracy of the ray-tracing algorithms based on various sound receiver models', *Applied Acoustics* **64**(4), 433–441.

Zitzler, E., Thiele, L. & Deb, K. (2000), 'Comparison of multiobjective evolutionary algorithms: Empirical results', *Evolutionary Computation* **8**(8), 125–148.