Thèse n° 7345

Impact of facade details on the reliability of performance-based decisions for early-stage neighborhood designs

Présentée le 2 juillet 2020

à la Faculté de l'environnement naturel, architectural et construit
Laboratoire de Performance Intégrée au Design
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

Minu AGARWAL

Acceptée sur proposition du jury

Prof. A. Meibom, président du jury
Prof. M. Andersen, Dr L. Pastore, directrices de thèse
Prof. C. Hopfe, rapporteuse
Prof. C. B. De Souza, rapporteuse
Prof. A. Schlueter, rapporteur

École
polytechnique
fédérale
de Lausanne

2020

Acknowledgments

In the last five years I have relied heavily on the faith and generosity of several colleagues, family and friends in order to bring this work together. This thesis would thus not be complete without a sincere acknowledgment of their contributions.

Firstly my advisor, Marilyne Andersen, one of the kindest and most compassionate persons I have worked with. I sincerely thank you for your immense patience during the development phases of this project, generosity, criticism and attention that have given much shape and definition to this work. I thank my co-advisor Luisa Pastore for relentlessly making sure that this work stays relevant for its intended audience and for her sharp editorial eye. People often say that research is a lonely affair. Marilyne and Luisa, your dedication and spirit have propelled me throughout. Working with you been a blessing. I also thank you both and Jan Weinold for creating a true atmosphere of collaboration at LIPID. It is a place with quiet exuberance and cheerful enthusiasm for knowledge creation. I am grateful for every hour spent there, working peacefully and joyfully.

Sincere thanks to the jury members Prof. Christina Hopfe, Prof. Clarice Bleil De Souza and Prof. Arno Schluter for examining this thesis with sympathy for its unique structure and for sharing their experience and knowledge with me. I also thank Prof. Michel Bierlaire for his incisive comments during the candidacy exam that helped me build perspective as a researcher.

Special thanks to Emilie Nault for sharing insights from her doctoral work, which ultimately gave rise to the research questions addressed in this thesis. Also, thank you Parag Rastogi for your active interest in this work, sharing time, ideas, scripts and tips that have greatly helped me along. Thank you Maria Garcia Riera for your enthusiasm and the many resources that you generously shared with me in the early phases of this project. Also thank you to Gerald Danseux for his hard work with the grasshopper scripts and a highly productive summer in 2018.

Sincere thanks to Prof. Marc Olivier Boldi for introducing me to decision theory. Without his invaluable suggestions, this thesis was at a high risk of becoming a lost opportunity!

To all office mates, Shevy, Maria A, Giorgia, Kynthia, Giuseppe, Jean-Denis, Thomas, Peter, Forrest, Victoria, Clotilde, Caroline, Sneha, Stephan and Geraldine, I thank you for siting through the many research meetings that perhaps should have been called research ramblings. You constant willingness to engage have helped me arrive at a moment of clarity that I can finally be proud of. Sergi, Martine, Joelle, Judith, Loic, for sharing lunches, French translations, thoughts and suggestions. Thanks to Sneha, and Judith for proof reading parts of

this manuscript. Special thanks for Forrest, Giueseppe and Giorgia for the coffee breaks on the sofa. You all are great therapists!

To my dearest friends in Switzerland - Helen and Chuck, thank you for a gentle introduction to Lausanne. Your sincere friendship and wisdom have helped me and my family through tough times and made the good times, memorable. Also thank you Chuck for your excellent book recommendations for a gentle introduction to decision theory! To Anne Berthe Ayik for being a friend, mother and grandma to my kids during my many absences. Also many thanks to my Indian friends in Switzerland (Ambika, Sushim, Sudipto, Noireeta) for yoga lessons, unending supply of chai and heated political debates. They provided mental breaks that every PhD student desperately needs.

To my father Padam, thank you for showing me the immense joy of learning. You went away too soon but you gave me enough. To my mother Anita, mother-in-law Asha Lata and sister Vithi for showing me what it means to be strong. Words cannot express my gratitude to my husband Prashant, my PR consultant, my chief promoter, my permanent co-author in life. Without you this thesis would have never come to be. Thank you to my two wonderful children Prasoon and Minerva, my greatest cheerleaders.

Finally I wish to thank the researchers who have laid the foundations of knowledge to which I have endeavored to add a small grain. While many of the works that I engaged with in an intellectual manner are acknowledged in the list of references, what remains perhaps unacknowledged in the citations is the effort that goes into making knowledge accessible to others (like me). After having written this thesis, I appreciate this more than ever.

Lausanne, June 13, 2020
Minu Agarwal

Abstract

Neighborhood-scale projects often commence with the conceptualization of several massing-schemes as potential design solutions. There is growing interest in using building performance simulation (BPS) to evaluate and rank such conceptual stage schemes in order to choose ones that best support performance goals related to energy and indoor comfort. However, such evaluations are typically made at a time of deficiency in information on building level attributes that influence performance, raising questions regarding their usefulness for reliable decision-making.

In this thesis, a new method, that builds upon existing risk assessment methods, is introduced to calculate the risk of performance loss faced by a conceptual stage BPS user/decision maker (DM). The proposed method considers three sources of risk of performance loss (1) reversal in ranks of design proposals (2) latency effect or a delayed discovery of performance gain (3) insufficient performance gain or loss of expected performance gain. These losses result from a design choice made between competing design proposals based on conceptual design stage BPS results that would be rendered invalid under future design development possibilities. To observe these losses and estimate the risk, a virtual progression of the design process is done through incremental facade levels of detail (fLOD) resulting in several future design scenarios. The resulting risk value combines the overall chance and magnitude of loss in the future design scenarios. It is further categorized as high' or 'low' risk based on the number of design paths that lead to future design scenarios with unacceptable loss.

This risk assessment method was tested by running a number of relative performance comparisons between pairs of competing neighborhood design proposals (N=780), based on three commonly used indoor environment related performance metrics: spatial daylight autonomy (sDA), annual heating and annual cooling demand. The results led to several important findings. First, while many performance evaluations lead to risk-free decisions, the number of high-risk cases was large enough (e.g. 1 in 5 comparisons on sDA) to suggest reconsideration of conceptual stage decision-making practices for projects where several design alternatives need to be ranked. Second, the likelihood of high-risk cases regarding performance loss depends on the metric, and becomes irrelevant only when design alternatives differ significantly in their evaluation outcome already at a low level of detail (LOD). Third, rank assignments based on either daylight (sDA) or annual cooling demand were found to be afflicted by all three sources of risk - rank reversal, latency effect, and insufficient performance gain, in annual heating demand evaluations, almost all cases of high-risk were due to latency effect.

Abstract

The aim of this risk assessment method is to link the reliability in decision making to fLOD present in the BPS models used for the performance evaluation. An original visualization aid is proposed to facilitate the DM's understanding of risk, and identify the appropriate BPS model fLOD for making reliable conceptual stage decisions. A design competition case study was used to test usefulness of this approach in assessing risk, when going through an actual design process. This thesis also presents a novel approach for evaluating decision-making practices in an experimental manner in the BPS domain that can inform future policy making.

Key Words: *Conceptual stage Design, Level of Detail, Massing Models, BPS User, Design Process, Risk Assessment, Decision Making, Reliability*

Résumé

Les projets à l'échelle du quartier débutent en règle générale par la proposition de plusieurs solutions de design sous forme de master-plan conceptuels. L'évaluation et le classement de ces master-plans conceptuels à partir de simulations de la performance des bâtiments (BPS) suscitent un intérêt croissant afin de choisir ceux qui présentent les meilleures performances énergétiques et de confort intérieur. Toutefois, de telles évaluations sont généralement effectuées en dépit du manque d'information sur les caractéristiques des bâtiments influençant leur performance. Cela soulève des questions quant à leur utilité pour la prise de décisions.

Dans cette thèse, une nouvelle méthode, construite à partir de méthodes existantes d'évaluation des risques, est introduite pour calculer le risque de perte de performance auquel fait face un utilisateur/décideur (DM) des BPS en phase de conception. La méthode d'évaluation des risques proposée tient compte de trois sources possibles de risque de perte de performance : (1) l'inversion du classement des propositions de design (2) l'effet de latence ou la découverte tardive d'un gain de performance (3) le gain insuffisant ou la perte de gain par rapport à la performance attendue. Ces pertes résultent d'un choix de design fait entre des propositions concurrentes conçues en fonction des résultats du BPS en phase de conception, alors ceux-ci sont invalidés dans les futures alternatives de design. Pour observer ces pertes et estimer le risque, une progression virtuelle du processus de conception est effectuée par augmentation incrémentales des niveaux de détails de façade (fLOD) afin d'obtenir plusieurs scénarios de conception futurs. La valeur du risque qui en résulte combine le risque global et l'ampleur de la perte pour les différents scénarios futurs. Le risque est alors classé comme étant « élevé » ou « faible » selon le nombre d'alternatives de design qui mènent à des scénarios de conception futurs avec des pertes inacceptables.

Cette méthode d'évaluation des risques a été mise à l'essai en comparant deux à deux les performances relatives de nombreux projets de quartier (N=780), en fonction de trois mesures courantes d'analyse de l'environnement intérieur : l'autonomie spatiale en éclairage naturel (sDA), la demande annuelle de chauffage et la demande annuelle de refroidissement. Les résultats ont donné lieu à plusieurs constatations importantes. Premièrement, bien que de nombreuses évaluations de la performance mènent à des décisions sans risque, le nombre de cas à risque élevé était relativement important (p. ex., une comparaison sur cinq pour la sDA) pour suggérer l'importance d'un réexamen des pratiques de prise de décisions à l'étape conceptuelle pour les projets où plusieurs alternatives doivent être classées. Deuxièmement, la probabilité de cas à risque élevé dépend du paramètre évalué et elle devient sans importance si les résultats des alternatives de design diffèrent considérablement dès un faible niveau de

détail (LD). Troisièmement, on a constaté que les classements en fonction du critère de la lumière du jour (sDA) ou de la demande de refroidissement sont soumis aux trois sources de risque - inversion de rang, effet de latence et gain de performance insuffisant. Dans les évaluations annuelles de la demande de chauffage, presque tous les cas de risque élevé étaient dus à l'effet de latence.

L'objectif de cette méthode d'évaluation des risques est de relier la fiabilité de la prise de décision au fLOD présent dans les modèles BPS utilisés pour l'évaluation de la performance des bâtiments. Une aide à la visualisation originale est proposée afin de faciliter la compréhension du risque par le DM et de déterminer le modèle BPS approprié pour prendre des décisions conceptuelles fiables. Une étude de cas d'un concours de projets a été utilisée pour évaluer l'utilité de cette approche dans l'évaluation des risques, dans le cadre d'un processus de conception réel. Cette thèse présente également une nouvelle approche pour évaluer les pratiques de prise de décision de manière expérimentale dans le domaine du BPS qui peut éclairer la future prise de décision.

Mots Clés : *Conception de l'étape conceptuelle, Niveau de détail, Modèle de masse, Utilisateur du BPS, Processus de conception, Évaluation des risques, Prise de décisions, Fiabilité*

Contents

1 Introduction

Almost 70% of the world population is projected to live in urban areas by 2050 [UN world urbanization, 2018]. The United Nations estimates that worldwide population of city-dwellers will continue to grow until 2030 at a net rate of about two million per week [UN world urbanization, 2018]. The need for construction of new buildings continues even in highly urbanized countries like Switzerland where urban population growth rates in some regions[1] are comparable to world average[2]. In the period, 2007-2017, 390,000 new residential units were added to the Swiss housing stock[Swiss FSO, 2019].

The size of individual dwellings has also been on the rise in several parts of the world, increasing the living area per person [Weber and Perrels, 2000]. Recently, a large number of housing cooperatives in Zurich, Switzerland preferred new construction over renovation of their 1950s estates [ValueS, 2011]. The existing buildings in this case had layouts that could no longer be renovated to meet today's lifestyle requirements. Such issues further heighten the challenge of meeting energy use targets such as the 2000 watt society. 2000 watt society is an energy use reduction target adopted by the Swiss federal government that that limits a person's use to 2000 watts per year by the year 2050[3]. The current energy need per person in Switzerland exceeds 7000 watts. Reducing it to 2000 watts in the next 30 years is challenging but a need of the hour[4].

New building standards not only address energy use but a wide range of issues like access to views, direct sunlight and daylight for the health and well-being of building residents [for Standardization CEN, 2019]. An inward development agenda has been put forth by cities around the world to promote densification. However, ensuring sufficient solar radiation access in buildings gets increasingly difficult at higher density [Darren, 2006]. The task of designing energy efficient buildings, while supporting a wider sustainability agenda and the lifestyle needs of occupants presents a formidable challenge to architects today.

[1] Zurich population growth rate 1.3%/year 2007-2017 [Swiss FSO, 2019]
[2] World average urban population growth rate 1.1 -1.3%/year 2000-2017 [World Bank, 2018]
[3] http://www.2000watt.ch/fr/societe-a-2000-watts/
[4] [IPCC, 2018] call for 45% reduction in CO2 emissions from 2010 levels, by the year 2030.

1.1 Building design as a decision making problem

While building designers today have to respond to a greater number of challenges, they have always needed to evaluate and substantiate their design decisions. Robinson [1990], while exploring the role and meaning of research in architectural design stated that *"...architect's expertise has to be validated, which is not possible if the proposal is seen as arbitrary"*. She further states that *"The problem is not that we do things that are not right, but rather that we cannot provide explanation to others of how we know that they are right"*. Empirically derived or judgment based 'rules-of-thumb'[5], protractor-like tools[6], scale models[7] and reference projects[8] have been providing architects with a basis for making design decisions for many decades.

These design aids can help the designer anticipate certain aspects of the quality of space once the building is built. Each of these methods, however, comes with its own limitations. For example, rules-of-thumb cannot take into account the specific design, site and climatic context of a project. They are conceived in an idealized setting and thus can only be used as guidelines. Scale models lack flexibility and are difficult to use to improve a design in an iterative manner.

In the last 15 years, building performance simulation (BPS) tools have increasingly filled such gaps to provide performance evaluations to designers while they are working on developing their design ideas [Miller et al., 2019]. BPS replicates specific aspects of the building design in the form of computer based mathematical models, that can be used to simulate the various physical phenomena in and around the buildings, based on fundamentals of physics. Growth in use of BPS tools in the design process was preceded by decades of research that linked building design features to comfort/health related effects on occupants[9] and energy use[10]. They invoke higher degree of confidence in designers regarding their design choices [Lam et al., 1999] and can provide feedback to a designer on complex design strategies at any point in the design process, specific to the site and climatic context. They are typically used with the objectives of lowering the building's energy demand while improving the indoor environment for building occupants and are well suited to the precise nature of today's energy[11] and indoor

[5] Tregenza and Mardaljevic [2018] chronologically list several rules-of-thumb and their basis of development to achieve daylight sufficiency in indoor spaces (for example 45° vertical sky angle at the center of the window is considered necessary for daylight sufficiency)

[6] For example 'Shading masks' introduced in Olgyay et al. [1957] allow a designer to evaluate seasonal and diurnal access to direct sunlight in the building interior

[7] Traditionally used for qualitative understanding of daylight interaction with building [Reinhart, 2014]. Bodart et al. [2007] provides an extensive overview of possible design applications of this technique

[8] Interviews with architects show that reference projects play a key role in early design decision making [De Wilde et al., 1999]

[9] Ghaffarianhoseini et al. [2018] provide an overview of studies on negative and positive health impact of building design and operation on building occupants

[10] Pacheco et al. [2012] provide an overview of key architectural design features that impact building energy use.

[11] For example, the Swiss vision for "2000 watt society" (www.stadt-zuerich.ch) seeks to limit energy demand to 2000 watts per person by 2050 in Switzerland

environment related design standards[12].

While BPS based performance evaluations provide valuable information to the designer, it is often not sufficient by itself for making good design choices. Concerns that *"easy-to-generate reams of simulation output are just as likely to overwhelm as empower the user"* [Tregenza and Mardaljevic, 2018] are starting to emerge.

Investigations into decision making using BPS had begun even while the use of BPS was in its infancy (De Wit and Augenbroe [2002]; De Wit [2001]). These early efforts recognized that decision making using BPS tools is essentially a problem of making decisions under uncertainty. De Wit and Augenbroe [2002] suggests that performance based design decisions need to be evaluated after subjecting BPS models to relevant sources of uncertainty and viewing design performance in a probabilistic manner.

Consider a common early design stage decision making problem as illustrated in figure 1.1. Figure 1.1 (i) shows the given design problem - a site where a specific set of needs to be met by a built structure. Figure 1.1 (ii) shows multiple design proposals developed by the architect that could be potentially suitable design solutions. The early stage design process can be dived into three sub-stages (MacMillan et al. [2002],[Liu et al., 2003],[Boyko et al., 2006]):

1. Divergence, where an array of possible design solutions are created. This activity involving generation of multiple design solutions is a means of exploring the design solution space.
2. Transformation, where all proposed design solutions are adjusted to make them more suited to the design brief.
3. Convergence, where design solutions are vetted against the objectives to select the best solution(s).

Figure 1.1 reflects a decision maker (DM)[13] in the convergence phase where he/she now needs to choose from among the available design alternatives. When this choice between the various competing design alternatives is made based on energy and indoor environmental quality related performance, then performance estimates are needed. However, BPS based performance estimates are subject to severe uncertainty at the early design stage when it is not known how the project is going to evolve through the various design stages. And yet, decisions must be made for the design process to move forward. Early design stage decisions such as massing, placement of buildings, window opening sizes have a trickle down effect on a large

[12] The new European daylight standard, EN 17037 [for Standardization CEN, 2019] requires a minimum of 300 lux for 50% of the indoor area of 50% of the daylit hours.

[13] The term DM is used in this thesis to signify member(s) of the design team who are final decision making agent(s). The DM could be the architect, BPS specialist advising the architect or the client, the engineer, the client or a combination of these professionals. The presumption is that the DM is interested in making design decisions based on performance evaluations from BPS.

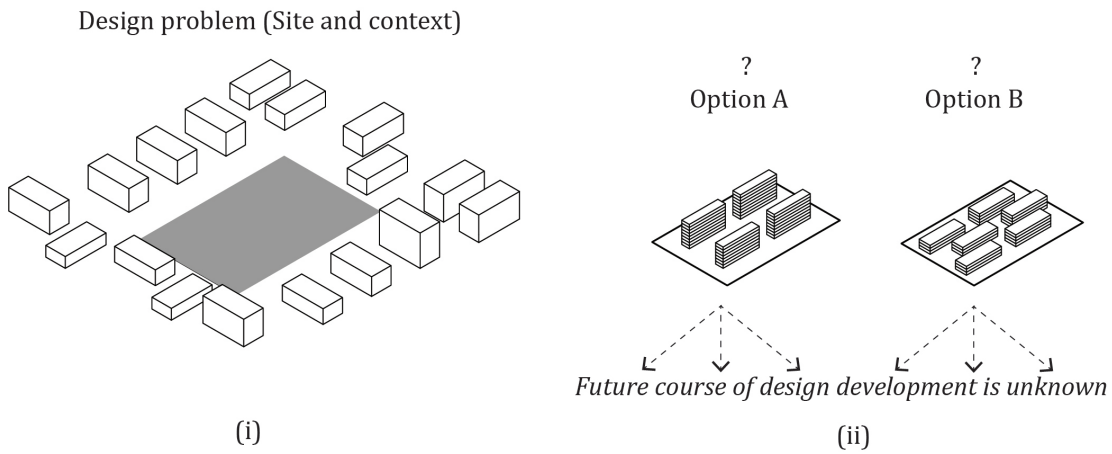


Figure 1.1 – Common design decision problem where a design alternative needs to be chosen from multiple design proposals.

number of subsequent design decisions [Ratti et al., 2005] and yet these decisions need to be made under uncertainty.

Methodological approaches to decision-making have been suggested (for example, Hester et al. [2017], Basbagill et al. [2013][14] that allow the decision maker (DM) to address uncertainty in early design stage performance estimates from unknown detailed design features that will materialize at later design stages. If this uncertainty is ignored, the DM stands the risk of making sub-optimal design decisions or introducing a 'design gap' which suggests notable performance difference between *"the selected design, and a truly optimum desig"* [Wright et al., 2016]. While the use of BPS tools and BPS based evaluations in the early design stages, is encouraged universally[15] as a means to reduce the design-gap, simplistic analyses and decision making methods[16] could potentially impede such an objective. This thesis allows BPS based decision makers to understand the potential loss from not adopting more sophisticated decision making methods at the early design stage.

1.2 Impact of design decisions at the neighborhood scale

To understand the impact of design choices at the neighborhood scale two broad categories of methods, namely, 'top-down' models and 'bottom-up' models [Reinhart and Davila, 2016] are used. Top-down models examine characteristics of certain sections of a city, neighborhood or buildings to understand their implication on energy use. When a particular design proposal for a neighborhood or building is to be analyzed, 'bottom-up' models are used which model various physical phenomenon inside and around buildings to estimate indoor conditions and

[14] These and others methods relevant to the early design stage are further discussed in Chapter 2.
[15] Architects [2012], RIBA and Sinclair [2012], ASHRAE [2018], CIBSE [2015]
[16] Approaches such as 'one design feature at-a-time' approach (also known as 'what-if' approach)

subsequently need for energy use for heating, lighting, cooling and ventilating buildings can be calculated. 'Bottom-up'models are largely building performance simulation (BPS) models.

Using BPS, several researchers have demonstrated the potential impact of the built environment on building occupants and the consequent energy use. At the urban scale Ratti et al. [2005] gathered urban texture samples (using digital elevation models) from three different European cities and analyzed them for weather data of London. Overall variation caused solely by urban texture was found to be 10% when 50% glazing ratio was used. Using these models they were further able to compare the effect of two different facade design strategies for an existing urban morphology and found that assigning higher glazing ratio to over-shadowed facades improved performance by 6-7% compared to uniform application of glazing ratio.

Rode et al. [2014] tested if the heating demand could be predicted based on the urban morphology (morphology represented using geometrical properties like surface/volume ratio and site coverage). They showed that each geometrical property has a different degree of correlation to heating demand depending on the morphology but some indicators have a more consistent correlation than others. They also concluded that urban morphology's solar cross section has a key role to play in determining the 'ideal' glazing ratio, when glazing ratio is based on the balance between beneficial direct solar gain and conduction loss through windows. Their findings indicate that design features are dependent on each other for optimality.

Sattrup and Strømann-Andersen [2013] created detailed thermal and daylight simulation models based on abstracted versions of eight different types of large residential developments commonly found in northern Europe. They found that the overall energy impact of choice of form was 16% (Copenhagen, Denmark). However much greater variation was found in daylight autonomy (DA) [Reinhart et al., 2006] achieved with the change in urban forms. While variations in heating demands could be compensated to some extent with well insulated envelopes, the variation in daylight (32%-52% average DA) appears to be a critical contribution of urban form. A fixed window to floor area ratio (W/FA) ratio of 30% and residential occupancy type was input.

Ratti et al. [2005] argue that while an effect of 10% may appear small, the urban scale is at the top of the design development hierarchy and that effect of urban scale decisions in not limited to the overall morphology. Rode et al. [2014] show that that urban form has a role to play in the thermal balance of the buildings that urban form is comprised of. Also the impact of urban form is not consistent on all performance metric or levels of urban density [Sattrup and Strømann-Andersen, 2013].

These and other studies at the urban and neighborhood scale show that planning for low-energy, daylit environments needs to begin at the urban and neighborhood scale. A number of design tools have thus come into the foray that can help designers to begin working on their design strategies for daylight and passive solar heating at the neighborhood scale. For example, Vartholomaios [2015]; Okeil [2010]; Knowles [2003] have suggested methods for shaping neighborhood form to improve solar access during a period when the buildings

are expected to be in heating demand mode. DeKay [2010] suggests a similar approach to improve daylight access at the neighborhood/block scale of design. The 'Solar Envelope' method [Okeil, 2010], that creates a bounding volume for new developments such that they do not over shadow surrounding existing buildings has been incorporated into Ladybug[17], a conceptual/early-stage design tool. Such tools are providing designers new methods for producing conceptual design alternatives. Next, we shall discuss how the use of BPS tools is changing the design decision making processes.

1.3 BPS as a design-decision making tool

BPS has become an indispensable tool for building design and in the last 15 years its use has grown by several folds Miller et al. [2019]. A recent survey of BPS based consultants/specialists and architects working with such consultants revealed a shift in the dynamics of design decision making, from the hands of the architect to the BPS consultants [Alsaadani and De Souza, 2016]. The role of BPS tools thus appears to be transitioning from design choice validation to playing an active role in decision making. Architectural design industry [Architects, 2012; RIBA and Sinclair, 2012] and municipal bodies [Housing, 2018; SIA , 2017] all over the world are advocating the use of BPS through out the design process. While, BPS tools have been largely viewed as a means to support the architectural design processes [Attia et al., 2012], new guidelines suggest changes in the design process itself are needed in order to fully exploit the ability of BPS tools to reduce the 'design gap'. For example, Leadership in Energy and Environmental Design (LEED v4.0), a popular building rating system introduced one reward credit for an integrative design process(IDP). IDP calls for proactive sensitivity analysis using BPS to inform massing, orientation and facade related design decisions (among others) before the commencement of the conceptual design stage. Similar guidelines are included in ASHRAE 209-2018 (Energy simulation aided design for buildings) standard for use of building performance simulation tools in the design process [ASHRAE, 2018]. IDP is based on a multidisciplinary approach to decision making, where multiple design features are considered concurrently as early as possible in the design process and is seen as a critical step in achieving high performance design [AIA, 2007].

Contrary to IDP, a sequential design approach is one in which design features are decided upon in a largely one-at-a-time manner. This approach has developed over time where designers tend to 'zoom in' gradually from site level issues to building and then subsequently work on building interior related issues. Reinhart and LoVerso [2010] in the context of designing for daylight stated that *"design practitioners tend to initially follow an 'outside in' approach... the building form is developed... once a basic building form has been conceived, different facade variants can be explored"*. While such an approach is conducive for spatial organization for various design elements as per their geometrical hierarchy, it introduces large degree of uncertainty in early design stage BPS based performance evaluations.

[17] https://www.ladybug.tools/

Typically relative performance evaluations (e.g. see figure 1.1) are preferred at the early design stage[Attia et al., 2012b] based on the premise that if all unknown building parameters are kept the same across BPS models for each design alternative, we should be able to estimate the performance difference between the alternatives and subsequently rank them in order to choose one or at least reject some of the poor performing alternatives. However, this assumption has remained largely untested.

1.4 BPS evaluations: The chance and cost of being wrong

BPS based approach would typically be considered as evidence based approach to building design. However, without employing proper procedures of modeling and robust decision making, the purpose of using BPS tools could be compromised. In this thesis we examine whether performance loss could be incurred in the absence of any robustness measures in BPS modeling and decision making practices. Such a loss is a virtual loss, as it is the difference between 'what is' and 'what could have been', resulting from sub-optimal decision making. The loss studied in this thesis is the unrealized performance potential, lost due to simplistic decision making methods. When such an event occurs, effort made in generating multiple design alternatives is lost as the DM is unable to make the correct choice. In the absence of robustness measures, the design team also exposes itself to the risk of violating the design performance goals that maybe voluntary or obligatory in nature. To summarize, while the performance loss maybe virtual, following losses are real. They are not studied explicitly in this thesis but are implied in the performance loss when it occurs:

Loss of design effort and time: The largest cost incurred during the design phase, is the time and expertise of various design team members. Once the design requirements have been established, design team members use them to establish design solution characteristics, search for appropriate design solutions and participles. Finally the requirements and solutions are combined to produce design alternatives. The design alternatives may go through multiple round of refinement. A recent survey of architects and engineers found that the conceptual design phase lasts 12 weeks on average and 3 design alternatives are produced [Jusselme et al., 2020][18]. Computational design techniques can reduce/ease the architect's burden, can produce large number of design alternatives (>1000) (e.g. Wilson et al. [2019]) and potentially expand the design space exploration. However the challenge of robust decision making still remains.

Loss of BPS modeling effort and time: Implicit in the lost effort due to improper selection between design alternatives, is the loss of the BPS based modeling effort that led to improper ranking of alternatives. As mentioned earlier, an important application of BPS tools is to make justifiable design choices. However that purpose is defeated if due to simplistic modeling and

[18]As per findings from online survey with 414 participants reported in Jusselme et al. [2020], the average length of the conceptual design stage is estimated to be 12 weeks. The median length is 8 weeks. Also the survey participants reported an average of 3 design variants generated per project during the conceptual design stage.

subsequent decision making procedures, the design decisions made are unreliable. BPS is still viewed as an added task in the design process and any additional efforts made for improving the design using BPS (beyond mandatory compliance checks) are seen as additional financial burden [Alsaadani and De Souza, 2016]. Meanwhile, the design team may remain unaware of the performance loss or design-gap resulting from decision making process as the design-gap is unexplored design potential. A design activity (conceptual design stage performance evaluations), that may require additional efforts to be included in the project scope, has the potential to be ineffective.

Loss of design flexibility: Recent studies suggest that the amount of design flexibility (freedom to choose design features) can be highly sensitive to the performance goals that a project needs to meet [Basbagill et al., 2014; Nault et al., 2020]. For example Nault et al. [2020] showed that a small relaxation in the performance goal from 20 to 21.4 kg CO2/m2-year, can make 2800 additional combinations of design feature specifications compliant (15 design features were considered including building shape, window openings, materials and mechanical system type). These 2800 newly compliant combinations represent a 21.3% increase in design flexibility, achieved by a modest 7% shift in the performance goal. If we consider a loss of a similar order (say 7%) due to absence of any robustness measures in conceptual/early decision making, it could imply greater strain on the designer at later design stages in order to stay compliant with the performance goals of the project.

Industry consortium ASHRAE has put forth building simulation standard [ASHRAE, 2018] which can help the design team/DM to reduce the 'design-gap' at the early design stages and avoid the losses mentioned above. The standard stipulates goals in terms of the performance analysis that must be carried out at each design stage (8 stages have been identified in the standard). At the conceptual design stage (referred to as modeling cycle 2 in the standard), it says that the performance for all design alternatives must be compared. It also requires that performance improvement strategies be proposed for all design alternatives before rejecting any of them. In other words, standard requires that performance potential of each alternative must be explored in depth before accepting or rejecting any alternative.

However it does not specify what kind of methods could be followed for adequately exploring performance potential for each design alternative. Additionally, even if proper due diligence and best practices (e.g. ASHRAE 209 standard) are followed at the conceptual design stage but in the final design stages of the project, some design elements decided upon at the conceptual/early design stage are changed, does it mean that all early stage analysis is nullified? Can performance analysis methods ensure robustness to such vagaries of the design process? These are crucial issues that need to be addressed while laying emphasis on the use of BPS tools in the conceptual/early design stage.

1.5 Thesis contribution and organization

In this thesis, we evaluate reliability of ranks assigned to conceptual stage design alternatives in a sequential design process that induces high uncertainty at the early stages of design, but is simpler (compared to IDP). This endeavor supports a simple, but not simplistic decision making process to find balance between design-time, design-practices and robustness at the conceptual design stage. To do so, a new robustness metric is proposed to estimate the risk of performance loss when making under a simple sequential design process. The proposed the metric addresses the following question:

Would the design choice made between competing design proposals, based on conceptual design stage BPS results, remain justifiable irrespective of design choices taken at later design stages?.

The proposed risk based robustness metric is named Expected Relative Performance Loss (ERPL) and in this thesis the risk in conceptual stage decisions is estimated from uncertainty in important facade related design features. Uncertainty in performance evaluations from other sources such as aleatory sources (e.g. future evolution of the built context, weather and occupant behavior related variability) though important, are not considered in this work. At the very early/conceptual stages of design if uncertainty in important design details is coupled with aleatory forms of uncertainty, it may result in performance ranges that are too large to be of use for decision making [Tregenza, 2017][19].

The risk values are presented in relation to the availability of precise information regarding design features in the BPS models. The precision in design features in BPS models is signified using a model level-of-detail (LOD) framework where low LOD implies low precision in (or absence of) inputs regarding several important design features. The DM can relate the risk of performance loss to the BPS model LOD that was used for decision making. The risk metric ERPL is thus also proposed as a criterion for selecting model LOD where traditionally, the choice of BPS-model LOD is based simply on the amount of design information available at hand.

The specific focus of this thesis is on neighborhood scale design problems where the designer has opportunity to work on a large amount of built area. Reconciliation of design constraints while meeting needs of multiple buildings can lead to several distinct design solutions. Choosing between these conceptual/early stage neighborhood design proposals becomes an important juncture in the design process as continuing to develop all design solutions can quickly become prohibitively expensive. Following is the chapter wise summary of the thesis:

Chapters 2: This chapter presents the need for robustness evaluation when making design

[19] Tregenza [2017] found the variation in indoor illuminance values to be 0.5 to 1.5 times the mean due to future development scenarios of the built context. Variations in heating and cooling energy demand of a similar order have been observed due to uncertainty in occupant behavior [Hoes et al., 2009]

decisions using BPS and state of the art for robust decision making at the early design stage. Techniques that take an integrated design approach are discussed as a means to improve performance outcomes of the design process with emphasis on ranking of design alternatives at the early design stage under uncertainty.

Chapter 3: This chapter presents the experimental approach that was developed to evaluate the risk (of performance loss) in early design decisions. The method for generating example decision making situations is explained in detail, along with the methodology for calculating the risk of performance loss in each of the decision making situation. The risk calculation method (for ERPL) is explained both in mathematical terms and graphically using some example data. Finally, the analytical framework applied on the data generated from the experiment is presented.

Chapter 4: The results from the experiment described in Chapter 3 are presented here. The summary of results includes the frequency, nature and magnitude of risk in early design decisions (as per the experimental findings). The potential need of a DM to qualify a risk value in a particular decision as being high enough to take risk mitigating actions or being low enough to ignore the risk is explored. The incidence rate of encountering high risk in conceptual stage decision making is also shown. Using statistical analysis methods, the overall risk values observed in various instances of conceptual stage decision making are examined to identify the conditions that induce risk.

Chapter 5: Using the data from the experiment described in Chapter 3, the ability of higher LODs to reduce the risk in relative performance loss is examined. In the calculation of risk done in Chapter 4, several assumptions are made regarding the decision maker (DM). Here the impact of various assumptions, namely, performance difference threshold needed for assigning ranks, threshold for risk (or risk tolerance) based on ability to take remedial actions is examined. These assumptions reflect a DM's sensitivity to performance gain and process related constraints respectively. These factors can play a role in the risk assessment, which then effects the prudent, further course of action.

Chapter 6: In this chapter, the potential of ERPL as a decision making tool is explored. The main value that the ERPL based risk assessment brings to a DM is the information whether a decision made using the BPS models at a given LOD is reliable or not. However, high risk conditions may spark a DM's curiosity as to why the risk is high. To satisfy such curiosity, a graphical visualization of relative performance outcomes from future design scenarios at higher LOD is presented as a possible augmentation to the ERPL value to further aid decision making. Using a case study (neighborhood scale design competition in Switzerland), the ability of ERPL to rank design alternatives, compared to other robustness metrics, is examined and discussed.

Chapter 7: This chapter summarizes the main findings and present their implications for BPS based DMs. Possible significance of the thesis beyond the presented scope is discussed along with the limitations of the work and future avenues of development.

2 Early-stage design-decisions using BPS

Performance based design decision making (while the design process is still ongoing) inherently implies decision making under uncertainty (DMUU) (De Wit and Augenbroe [2002]; Hopfe and Hensen [2011]). Performance related uncertainties are especially severe at the early design stage. This stage is characterized by a large number of design decision that are yet to be made and BPS models which contain little design information that is reliable. At the same time, the early design stage is considered important for the following reasons:

- **Flexibility in design exploration:** Given the large degree of design flexibility, the early design stage has been and continues to be an area of high interest for researchers in the BPS domain [Shi et al., 2017]. BPS is viewed as an important tool to aid early design explorations exploration using optimization and statistical analysis techniques [Østergård et al., 2017].
- **Influence on the performance outcome:** Early design stage decisions such as massing/form, placement of building(s) relative to other buildings and window openings are known to have large effects on multiple aspects of building performance and ability to deliver comfort to occupants in a passive manner [Ratti et al., 2005].
- **Irrevocability in design decisions made early on:** Early design decisions form the basis for many design decision made later on in the design process. Any changes desired in the decision made early on, would imply significant re-design costs. In response to such challenges, several municipalities offer a phased design approval process to avoid time-consuming detailed design development before a preliminary proposal has been checked and approved [Pedro et al., 2011]. Such phased approval processes protect designers from re-design costs but further heighten the importance of early design stage decisions.

Figure 2.1 shows the importance of integrating BPS tools early in the design process to benefit from the unique opportunities present at this stage. However, studies also suggest that high

degree of uncertainty present at this stage could be detrimental to reliable decision making. Figure 2.2 shows an abstracted graphic based on findings in [Struck et al., 2009] where the uncertainty in performance evaluation at the commencement of the design process may be too high to provide reliable information to the DM for making decisions. Several methodological interventions have been suggested by researchers to counteract high uncertainty at the early design stage. They can be divided into two main categories:

1. Methods to **reduce uncertainty** by taking an integrated design approach. In this approach the BPS user is directed towards design features that are most effective in reducing uncertainty. This is commonly achieved using sensitivity analysis (Section 2.1.3). This approach is also called 'data-driven design space exploration' [Jusselme et al., 2019] but in the context of this thesis, this approach has been referred to as uncertainty reduction.

2. Methods for **robust decision making** under uncertainty (DMUU). Under this approach the robustness of design decisions is tested under uncertainty (Section 2.2). The interest of such methods is not to reduce uncertainty from unknown design decisions or unknown conditions. Rather, robustness to uncertainties is rewarded and in some cases used to rank design alternatives.

Both these approaches are useful to the early design stage DM in different ways. Uncertainty reduction methods guide a DM's priorities in design decisions. Under this technique, the DM is encouraged to 'fix' certain design features and as a result he/she can increase chances of achieving desired performance outcomes. Recent studies, emphasize uncertainty reduction while maximizing design flexibility [Basbagill et al., 2013b], [Jusselme et al., 2017]. This technique implies prioritizing design features that not only reduce uncertainty in performance but do so with minimal need to 'fix' other design decisions. This reflects the potential interest of an early design stage DM in robust decisions but not at the cost of loosing design flexibility. Robustness metrics for DMUU can be used to evaluate the reliability of decisions given prevailing/modeled uncertainty. However, the DM then needs to determine if the reliability in decision making is acceptable to him/her or not.

Uncertainty reduction techniques come with the need to specify additional design features that the DM has not done so far. This chapter surveys existing literature to understand the extent to which such methods can improve performance outcomes of the design process. It looks into robustness metrics for DMUU to identify ones that are most suited to early design stage decision making. To operationalize these techniques for more informed decision making, several uncertainty propagation methods are available. These techniques shall also be examined in section 2.1.3.

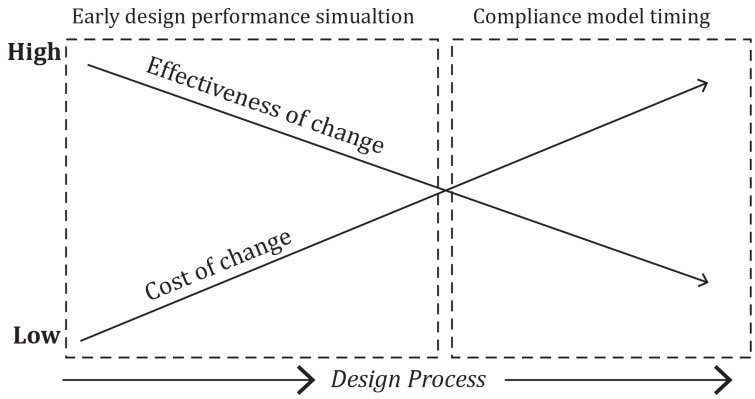


Figure 2.1 – Commonly known argumentation to integrate BPS at the early design stage (large effect on performance at low cost). Figure adapted from AIA's guide for integrating BPS in design process [Architects, 2012].

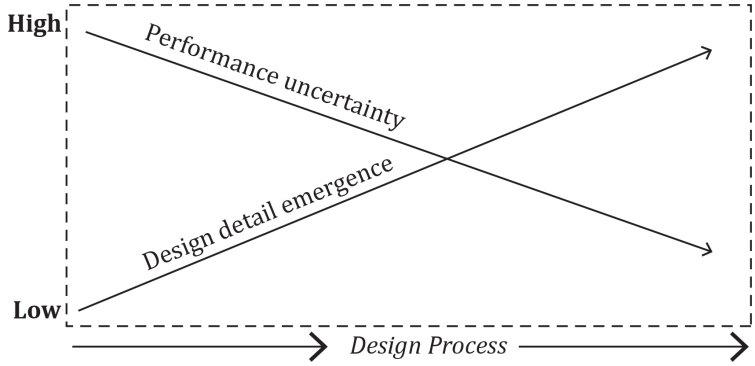


Figure 2.2 – Diagram based on findings in Struck et al. [2009] indicating that uncertainty at the commencement of the design process is high, and drops as more design features get specified.

2.1 Making reliable early design decisions

2.1.1 Use of BPS for early design decisions

The most common decision making application of BPS tools at the early design stage is for comparing and ranking design alternatives [Attia et al., 2012b]. Attia et al. [2012b] conducted an online survey of 471 architects and engineers in late 2008-early 2009 to understand the potential of, and challenges in use of BPS tools to support design decisions. The questionnaire had five main types of questions relating to information management, accuracy, guidance, interoperability and integration in design process. With regards to information management within BPS tools, architects voted *"Creation of comparative and multiple alternatives"* [1] as the most important feature of BPS interfaces.

There are a few other important aspects of the early design stage decision if we examine the

[1] "Creating of comparative and multiple alternates" here refers to representation and simulation of multiple design alternatives in BPS tools for comparative performance evaluation.

literature related to use BPS tools at the conceptual stage for design development:

- **The multi-criteria nature of the design performance evaluation** is strongly reflected in user feedback to existing energy performance based early design support tools (Attia et al. [2012]; Hitchcock et al. [2008]). For example, the energy simulation tool, ZEBO Attia et al. [2012] can only handle energy issues, while users expected other environmental and economical indices as well. Including multiple metrics can make building performance assessment a powerful tool for conflict resolution among different stakeholders with disparate expectation as it distills design into performance metrics.
- **Relative performance evaluation** is often the main focus of early design stage performance evaluation. Compagnon et al. [2015] suggest that it is more useful for early-design-phase support tools to be able to present the magnitude of change in performance from one design option to another. Calibrating the performance evaluation (e.g. significant or perceptible difference between alternatives) could be a desired goal of early design metrics rather than define absolute target values that are often climate dependent, take a lot of computation time and specificity to be calculated.
- **Simplified metrics**(e.g. geometry and building external-surface solar-irradiation based) Nault et al. [2015] as opposed to detailed thermal and daylight modeling based metrics (e.g. Spatial Daylight Autonomy(sDA), annual heating demand) are often used at the urban and neighborhood scale to lower the computational load. Reducing computational load for faster evaluation of design alternatives, remains an important issue for use of BPS tools at the early design stage [Østergård et al., 2017].

Estimating the indoor environmental quality is the end goal, irrespective of the type of performance metrics used (Nault et al. [2015]; Compagnon [2004]). In this thesis, we used detailed thermal and daylight simulation methods that are more susceptible to uncertainty in design details (compared to simplified metrics), but are of the greater value to the designer. Next we shall examine the state of design development of early stage designs that can be expected to understand the extent to which early design stage is compatible with BPS evaluations.

2.1.2 Design information-deficiency in early design stage BPS models

High degree of abstraction in early design architectural models (CAD based or otherwise) is common and these simplified models are known to facilitate the design process. Quick and simple models are often developed to examine and discuss design ideas (Akin and Moustapha [2004]; Kvan and Thilakaratne [2003]). Kvan and Thilakaratne [2003] highlight that models in the architectural design process, especially early on, tend to be diagrammatic or representative in nature and they develop along the design process. Mahdavi [2003] states that *"Models are entities that represent other entities. While represented entities may be arbitrarily complex, models can be highly "reduced", i.e. they may focus only on a limited sub-set of the features of*

the represented entity while abstracting from other features." Thus models can contain varying amount of design information depending on the extent of design development.

Acknowledgment of design-information deficiency

Struck et al. [2009] observed 10 teams of students working on a building design in a design studio over a period of 10 weeks. The development of various aspects of the design was tracked in terms of design attributes (e.g. volume, area, window-wall-ratio (WWR)), components (e.g. structural system type, window framing systems) and relationships (e.g. space and occupancy relationship signified by air flow patterns and comfort). In the first week, 1-6 design aspects were identified by various teams which grew to 10-28 design aspects at the end of 10 weeks. These numbers indicate the aspects that were considered, but may or may not have been fully resolved by the student. The study also noted widening gap among design teams in terms of the number of design aspects that they chose to work with at week 10. While some groups focused on fewer aspects (10 in number) and thus deliberated on the conceptual design for longer, other groups finalized the concept and delved into a larger number of design aspects much sooner.

This gradual arrival of design information poses a challenge to BPS evaluations as a number of design inputs need to be specified in BPS models in order to execute them. For example, the simplified Light and Thermal (LT) method [Baker and Steemers, 1996] later used by Ratti et al. [2005] to study the effect of urban form on annual heating demand requires 30 additional inputs (apart from the urban geometry) regarding the building level design characteristics such as window-to-floor area ratio and shading device type among o thers. Performance assessment of the conceptual stage designs thus requires constant values to be assumed for unknown parameters (e.g. window area and placement) and may also ignore some building details that may be specified later (such as fixed shading, balconies et c.). Uncertainty due to undecided building parameters makes these assessments difficult to carry out, requires a number of assumptions to be made and raises questions regarding their reliability.

Treatment of design-information deficiency in BPS tools

Early design stage BPS tools such as Zero Energy Building design tool (ZEBO) [Attia et al., 2012] and Urban Modeling Interface (UMI) [Reinhart et al., 2013], among others, have made concerted efforts to fill gaps in the design information at the early design stages. For example (UMI) [Reinhart et al., 2013] an urban scale performance simulation tool, deals with geometrical details and material property specification related deficiencies in conceptual stage 3D models. It divides the 3D volumes floor-by-floor and can perform internal zone divisions as well. Material properties can be assigned using customizable templates. ZEBO does not add geometrical details but provides templates that populate the BPS model with material properties. The templates provide data corresponding to code compliant design features. The user has to only select applicable code to populate the BPS model with neccassary data.

Such automated workflows allow the users to overcome gaps in their knowledge of the design details and BPS related terminology [Attia et al., 2012].

Like UMI, SUNtool [Robinson et al., 2007], Young Cities [Huber and Nytsch-Geusen, 2011] can automate the process of transforming simple neighborhood massing geometries into BPS models. UrbanSolve [Nault et al., 2018] also builds BPS models (thermal and daylight) from simple 3D geometry. It uses a combination of BPS modeling of some of the physical phenomena and meta-modeling on a reference set which significantly reducing the processing time. The development of these tools is well supported by observational studies examining the current nature of early design stage BPS practices. Surveys of BPS users point to a general consensus in the simulation community that simple early design simulation models are more conducive to the design process (de Souza [2009]; Attia et al. [2012]).

This consensus is also reflected in assessment criteria of early design BPS tools [Attia and De Herde, 2011] and simulation guides (Saxena et al. [2011]; ASHRAE [2018]). Saxena et al. [2011] daylight simulation guide acknowledges design information deficiencies that can be expected at the early design stages. The report suggests three levels based on 'level of detail' (LOD) of geometry and precision in inputs. Level one is considered suitable for early design stages, while the highest level (level 3) is considered suitable for modeling existing buildings. At level one, geometry can be kept simple (e.g. simple 2D windows), reflectance properties can be set to predefined default values. The guideline suggests that 'optimistic' inputs maybe used to understand the highest performance potential of the design.

Building information model (BIM) based CAD tools are increasingly integrated to the design process and can be used to integrate three main types of design information: 1) geometrical data 2) semantic data (eg.g material properties) 3) topological (relationships and dependencies between design components) [Schlueter and Thesseling, 2009]. BIM based CAD tools help to address the multi-dimensional nature of design information. With greater interoperability between BIM based CAD tools and BPS tools using common file exchange schemes (IFC, gbXML) greater precision can be expected in BPS model inputs. While one may have tools that can support better data transfer to BPS models and various computational techniques to tackle data deficiencies, the design still remains in the early design state and the BPS model relies on a number of assumptions.

BPS user feedback collected for the development of the Commercial Fenestration design tool (COMFEN) for net zero energy building design [Hitchcock et al., 2008] indicated that early design tools need to address ideas and issues that are site and end-user specific. In other words early design tool need to *"model building performance in all it real-world messiness"*. This recommendation addresses the performance gap that is often found in predicted performance and findings of post-occupancy evaluation (POE) s tudies. Donn et al. [2009] as a solution, recommend early design tools to move away from 'input data' and start thinking in terms of 'input information'. To further elaborate, instead of relying on single, static inputs, several variations of input data such as user behavior need to be considered in order to gain confidence

in performance estimates made. Effective communication of these assumptions was also found to be a critical component in acceptability of the simulation tools as reported by user evaluations of the ZEBO tool [Attia et al., 2012a].

2.1.3 Methods for quantifying and reducing uncertainty

In the previous section, various efforts to fill in the information-gaps in the early design proposals were examined that can enable detailed annual performance simulations. However in the absence of design information, if evaluations rely on a large number of assumptions, uncertainty in design features is ignored. Studies show that there can be enough uncertainty in early design performance evaluations due to unknown design information to impact design decisions [Tregenza, 2017] and that different assumptions in early design BPS models can lead to different results [Xia et al., 2008; Brembilla et al., 2018]. For example, Brembilla et al. [2018] tested if different assumptions related to interior and exterior surface reflectance properties could effect performance evaluations of some designs more than others. They examined sensitivity of four daylight metrics, measured for four different classrooms designs under uncertainty in material reflectance properties. The design with large windows (69% WWR on one face) was found to be more sensitive to exterior reflectance properties (varied between 0.05 to 0.6), while a room with small windows (25% WWR on one face) was found most sensitive to reflectance properties of interior walls. Thus sensitivity to assumptions was found to vary case-by-case. Among the performance metrics, Useful Daylight Illuminance (UDI)-c combined (100–3000 lx) was found to be most sensitive in terms of relative ranks among the classrooms designs.

However lack of information is not necessarily an impairment to decision making. Uncertainty analysis (UA) can help a decision maker, make decisions in the presence of uncertainty. Under UA, the important sources of uncertainty are identified and then introduced as variable inputs into the BPS models to produce probabilistic performance evaluations. UA can have multiple decision making applications. The simplest application can be to judge the variation or the range in that can be expected in the performance evaluation. In daylight simulations Tregenza and Mardaljevic [2018] found a possible ±20% effect on horizontal illuminance under a uniform sky at the center of a room with unknown design properties such as glazing light transmittance, internal and external reflectance properties. Tregenza [2017] further state that "conclusions that could be drawn from a statement that the illuminance at a workplace is 300 lux are different from those implied by the statement that the illuminance is 300 ±60 lux".

Rather than ignoring uncertainty, by modeling it, the DM can account for it in their decision making. In the BPS domain, several methods have been used to quantify uncertainty.

Uncertainty quantification

Two distinct methods of modeling uncertainty exist: 1) probabilistic; and 2) non-probabilistic [Colyvan, 2008]. The probabilistic approach is well suited for modeling design features whose exact values are not known but a probability distribution of their values is available. Under the non-probabilistic approach, all future possibilities are assumed to be equally probable. A set of scenarios are built to represent unknown states by sampling values of unknown entities.

In the BPS domain, for uncertainty quantification from unknown design features, Monte Carlo (MC) simulation has been used extensively [Lomas and Eppel, 1992; Macdonald and Strachan, 2001]. MC simulations have been used to model uncertainty in unknown building operation/occupancy, internal loads, material properties, future weather and such other inputs into simulation models. The unknown design feature, say the glazing type or window-wall-area ratio (WWR), are represented as a range of possible values. Sometimes, instead of a uniform range, specific distributions are used to represent expected probabilities associated to certain values of the design factor. Then, for each Monte Carlo simulation, random values of all unknown entities are drawn from their respective distributions or uniform ranges. Each new Monte Carlo simulation thus produces a new possible design configuration and/or operational conditions (e.g. occupancy, weather). Figure 2.3 (a) shows a diagrammatic representation of this technique where values are samples from five sources of uncertainty (marked as df1, df2...df5). This diagram shows two samples generated by sampling unknown entities and applying them to the design alternative at hand, in this case, a massing scheme. For comparison, 2.3 (b) shows the design of experiment (DOE) approach called a *full factorial* approach which is used for sensitivity analysis but not for UA. Full factorial approach, often implies that only extreme values of design features are used. This is useful to gauge sensitivity but does not necessarily produce a representative prognosis of performance.

However, the application of Monte Carlo simulation technique has seen limited use in representation of architectural design features. Any design feature, in order to be included in a Monte Carlo simulation, needs to be parameterized either as continuous or categorical variable. Hester et al. [2017] used Monte Carlo simulation to represent possible window designs in single family homes in Chicago. They used continuous uniform distributions for the total WWR for the whole house, and the window area share of one prominent wall. Basbagill et al. [2013] used Latin Hypercube sampling, a more efficient form of sampling method for Monte Carlo simulations. They included additional design features such as fixed shading devices and building shape. Discrete values were used to represent some parameters (for example, inclusion of fins and horizontal shade was a binary input (yes/no). The depth of shading device was then parameterized as a uniform distribution. Once, one adds design feature such as placement of windows on the facade and placement of shading devices with respect to the windows, random sampling techniques could produce possibilities, a number of which maybe considered unfit from an architectural design standpoint. Controls such as correlations between design feature and design 'rules' would be required to result in design configurations which could be considered valid and acceptable.

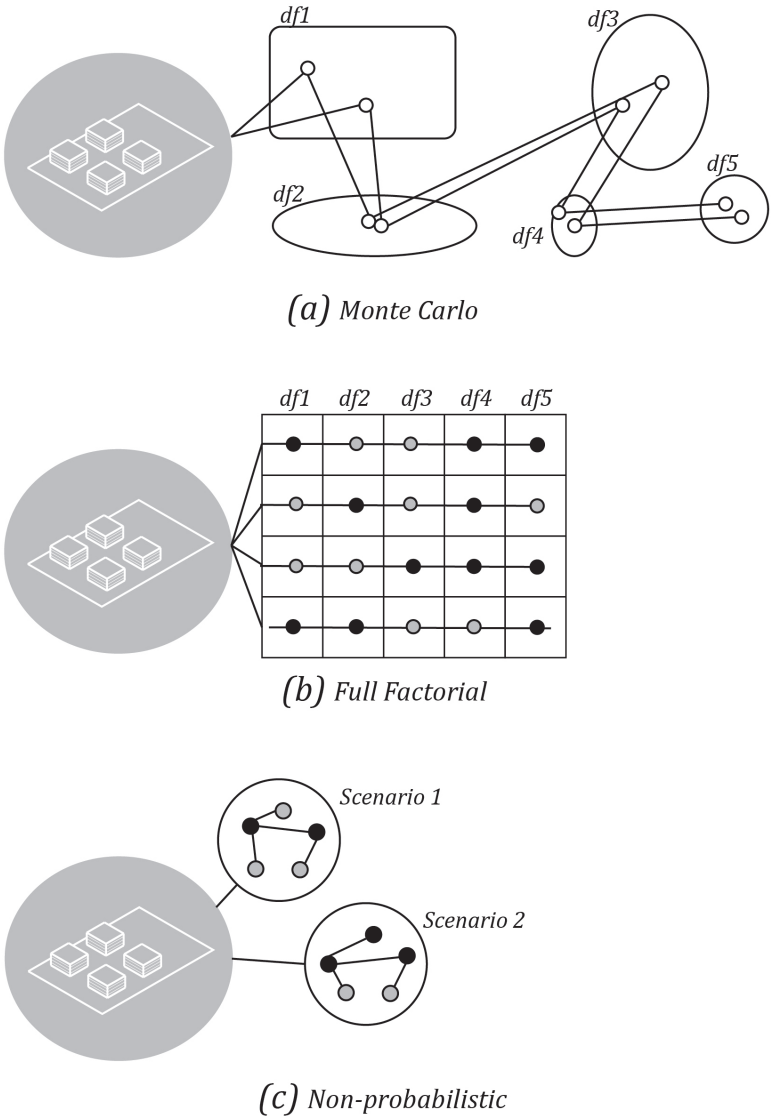


Figure 2.3 – Sampling methods commonly used in BPS. Part (a) adapted from Sun et al. [2014].

Recently, machine learning (ML) techniques have been used to generate design scenarios (for example, [Chaillou, 2019]). A common workflow for using ML in this area is to first have a random generator (much like the MC method) that produces random design solutions. The randomly generated design is, then, passed through a 'discriminator' to assess the validity of the randomly generated design. The discriminator is trained using ML, from training set of pre-existing design solutions. This technique overcomes the limitations of both, the MC approach (possibility of too many unacceptable design configurations) and the scenario based approach (data collection, design rules and scalability). However, The ML tools for validating design configurations are still evolving and are more useful in generalizing the algorithm at a universal scale. ML algorithms may not always be accurate in validating the feasibility of designs, depending on the data provided in the training set. For example, they may exclude all

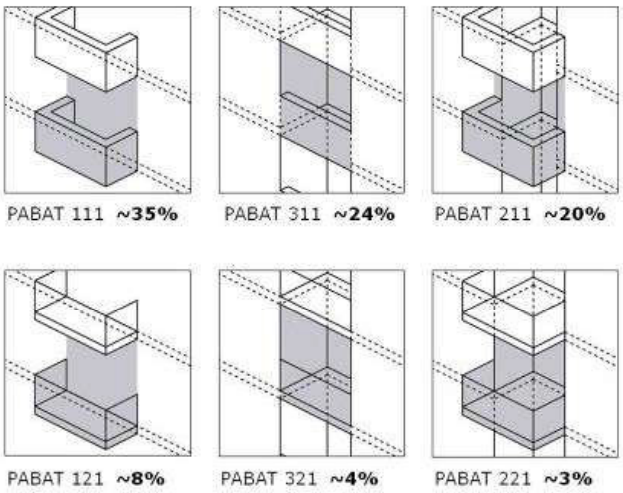


Figure 2.4 – Types of balconies found in Swiss residential buildings built between years 1919-1990. Term PABAT refers to the coding scheme used in the report to signify balcony types in a particular building typology, the percentage values indicate the occurrence of each type of detail within the typology [Schwehr and Fischer, 2010].

feasible solutions which were, inadvertently, excluded from the training data. As a result, for a specific context, manual intervention may still be required to develop a set of comprehensive and feasible design solution possibilities.

The scenario based approach has also been used in order to only generate valid design configurations ([Porritt et al., 2012; Hachem et al., 2011]). Under this technique, features of interest are parameterized as categorical variables and a rule based approach is used to fill in other design details that are important but are either difficult to parameterize or are not within the scope of the study. This method only produces valid/acceptable design configurations. For example simple rules such as *"size of south-facing ground floor windows is 12% of the ground floor area; while first floor south facade windows are 8% of first floor area"* [Hachem et al., 2011] are established. Such rules are combined with other feature such as building shape and different design scenarios are produced. However these rules are difficult to generate. In the study by Hachem et al. [2011] rules were derived from another study that provided recommendations regarding distribution of glazing in single family homes in high latitude locations. Large scale observational studies such as [Schwehr and Fischer, 2010] can be also be used to derive rules and/or design scenarios (Figure 2.4 shows the nature of the data that can be gathered through such studies). However, such data collection exercises can only be used to list prevalent design features that would be dictated by feature such as prevailing construction codes, design trends, cost etc. feature such as orientation of glazing are typically a designer's response to the specific site, built context and design needs. Rules regarding such features are best defined in terms of design scenarios and not drawn from a sample of existing buildings. 2.3 (c) shows a diagrammatic representation of the scenario based approach where relationship rules between factors are respected and specific combinations of unknown features maybe assembled.

Methods for uncertainty reduction

Various studies (Jusselme et al. [2019]; Hester et al. [2017];Basbagill et al. [2014]; Attia et al. [2012a]) use a combination of uncertainty propagation and sensitivity analysis (SA) to propose robust design paths that ensure synergy between early design and detailed design stage decisions. In order to reduce uncertainty for improving the performance outcomes of the design process, researchers have suggested the following approaches:

1. Identifying **influential future design features** using uncertainty analysis (UA) followed by sensitivity analysis (SA). Studies such as these rank most influential uncertain design features. (e.g. Hopfe et al. [2011])
2. Sensitizing the BPS user towards impact of future design decisions. These studies focus on presenting and **comparing various benefits of increasing the design development** through future design features. (e.g. Basbagill et al. [2014])
3. Providing **interactive exploration of future design features for uncertainty reduction**. The approach mentioned in item 1 is conducive to ranking sensitivity levels to individual features. Interpretation of interaction between multiple features in terms of metrics of sensitivity is difficult. Interactive visualizations can help BPS user understand relationships between multiple sources of uncertainty (e.g. future design decisions). Users can also understand the implications of having flexibility versus constraints in some design features. (e.g. Jusselme et al. [2017]; Nault et al. [2020])

All these approaches for uncertainty reduction direct the BPS user towards a more integrated design process [2]. An integrated design approach is regarded as one of the paths towards high performance through design. One of the key aspects of an integrated design approach is simultaneous consideration and resolution of multiple design features. Thus each design alternative (e.g. conceptual stage massing-scheme) would be considered along with a multitude of other design details (e.g. construction type, facade details) before choosing among the design alternatives(ASHRAE 209-2018, section 6.1, 6.2) [ASHRAE, 2018]. Such a process implies greater rigor in determining the design solution. Hester et al. [2017] found that energy demand stabilized only after 40% (8 out of 22) of the design features have been determined while exploring the early design solution space for a single family home in Chicago, USA. Some of these decisions were conceptual design decisions (total built area) while others such at attic U-value may typically be decided only in later design stages. Jusselme et al. [2019] adopted a data-driven design approach for a lab/office space project in Fribourg, Switzerland. While using global warming potential (GWP) as the performance criterion, they found that 21% of the design features (3 out of 14) represented 81% of the variance based sensitivity to performance. This design solution space exploration relied on 20,992 design alternatives generated by sampling 14 unknown/undecided design features.

[2] "Integrated design process is one where project team members look for synergies among systems and components, the mutual advantages that can help achieve high levels of building performance, human comfort, and environmental benefits" LEED v4 (www.usgbc.org/leed/v4)

These methods thus help a designer create a more detailed design (or design specification) for a conceptual/early design proposal which forms the initial 'seed' for the analysis. However, it is common in the early design phase, to compare the relative performance of two or more competing design proposals (e.g. massing-schemes) for further development. In case of relative comparisons, the decision rests not on the performance of the individual proposal (or 'seed') but on the relative performance difference between the proposals. Relative performance evaluations are typically used to rank design alternatives. In the following section, techniques adopted by researchers to rank various design alternatives under performance uncertainty shall be examined. These methods do not call for reduction in uncertainty, but use the prevailing uncertainty to inform decision making.

2.2 Decision making under uncertainty (DMUU)

De Wit and Augenbroe [2002] was one of the early works that highlighted the need to view the use of BPS in design, as a problem of decision making under uncertainty (DMUU). The study also showed that is was not sufficient to quantify uncertainty for decision making. Simply quantifying the uncertainty will only leave the DM confused. As an example, De Wit and Augenbroe [2002] presented the case of a DM trying to make the binary decision of whether a cooling system should be installed in a naturally ventilated office building (located in Netherlands) to prevent overheating hours from exceeding the stipulated limit. The BPS model based on fixed assumptions showed no need for a cooling system (overheating hours found to be 100, well under the maximum permissible limit of 150 hours). However the UA, which addressed uncertainty in knowledge of building characteristics and various empirically derived model inputs, showed that there was one in three chance of exceeding the limit of overheating hours. While the UA provided the DM with new information that could make him/her consider a cooling system but it does not help him/her make a clear decision. Further incorporation of concepts (such as utility derived from higher probability of respecting comfort requirements) is needed to aid decision making. De Wit and Augenbroe [2002] showed that the decision making problem is an action taking problem. The DM needs to be provided with methods to understand the implications of the uncertainty on design problem.

2.2.1 Classification of DMUU methods

A number of methods exist to support DMUU. Fisher et al. [1950], Neyman and Pearson [1928], Wald [1950] and Savage [1951], are considered to have provided four classical approaches [Raiffa, 1968].

The first two methods (by Fisher and Neyman Pearson) can be used test the probability of incorrect conclusions being drawn when entities are compared. They can be used to calculate the False Signal Rate (FSR) from a known case (sample) with respect to unknown/uncertain conditions (population). These findings of these methods are not conditional on any prior knowledge/inclinations of the DM.

Wald [1950] introduced the concepts of loss and value as tools of decision making. He also brought about a shift to a scenario based approach where probabilities are not assigned to uncertain states (scenarios). He assessed the expected value of a choice/decision based on the potential loss or value it could deliver. Similar to Fisher et al. [1950] and Neyman and Pearson [1928], [Wald, 1950]'s methods evaluate each choice independent of other alternative/competing choices to calculate the 'Expected Value' of each decision alternative.

Savage [1951] built upon the work of [Wald, 1950], by transforming loss from a decision, into regret. He defined regret as "the difference between the actual *payoff* [3] one receives for the decision that is made and the optimal payoff that would be obtained if the best decision was made" [Savage, 1951]. For a given scenario, the loss from a particular decision is calculated in relation to the optimal payoff that could have been realized if the optimal decision had been taken. The evaluation of a choice/decision in terms of regret is thus subject to other competing alternatives that are present.

McPhail et al. [2018] provides a different classification of DMUU methods as metrics for robust decision-making in the fields of design and engineering. Robust decision-making can acquire different meaning in different design contexts. For example, robustness could mean insensitivity to uncertainty, avoiding regretful decisions or avoiding negative outcomes such as failure to meet design requirements under uncertainty. The following types of robustness metrics can meet the needs of a diverse set of DMs:

1. **Expected value metrics (EVM) [Wald, 1950]** Expected value metrics include *maximin* and *maximax* metrics that are suited to two extreme types of DM. Maximax is suited to a highly optimistic DM who is interested in the best-case scenario. A DM following the maximax criterion would choose the alternative that delivers the maximum (highest) payoff, even if its average payoff or worst possible outcome maybe worse than other alternatives. Maximin is for the pessimistic or risk-averse DM who is interested in the worst case outcomes under uncertainty. A DM using the maximin criteria chooses the alternative whose worst payoff is better than the worst of other alternatives. That is, under the most adverse conditions, the alternative that delivers the best performance is chosen using the maximin criterion. These metrics are calculated under a non-probabilistic treatment of uncertainty. All uncertain scenarios are considered equally likely which is a useful assumption when the probabilities of scenarios cannot be known. Arrow and Hurwicz [1972] proposed the Hurwicz's optimism-pessimism rule that allows a DM to assign weights to the the worst and best payoff and calculate a weighted mean of these pay-offs. A decision can then be taken based on this weighted mean value.

2. **Higher-moment metrics**: These metrics are derived form various statistical properties of the performance values under uncertainty. While EVM metrics [Wald, 1950] look at payoffs under specific conditions, higher-moment metrics like standard deviation and variance are calculated based on the entire set of possible outcomes under uncertainty.

[3] In the context of BPS, payoff can be understood as the performance delivered by a design alternative.

3. **Regret-based metrics [Savage, 1951]**: *Regret* accounts for the full set of possible outcomes (compared to EVM) under uncertainty and regards any outcome that is worse than the most optimal, as regret or *opportunity loss*. The alternative with the least-maximum regret is considered the fittest under the *mini-max* regret decision rule. Like the EVMs, regret is also calculated under a non-probabilistic treatment of uncertainty.

4. **Satisficing-metrics(Simon, 1956)**: Satisficing metrics measure the probability of meeting the performance threshold that each alternative is held to. These are commonly used in the context of building design where performance targets are established before the design process begins. For example, these targets could be related to comfort levels or the energy/resource use to meet indoor comfort levels.

Figure 2.5 is adapted from McPhail et al. [2018] to show the metrics discussed above. The figure is diagrammatic in nature showing relative position of this metrics when ordered by risk aversion of the DM. A risk averse DM would be more interested in making decisions based on possible loss, while an optimistic DM would be interested in potential for gains.

Figure 2.6 shows the method for identifying the best alternative based on Hurwicz's optimism-pessimism, Wald's maximin and Savage's regret (shown partially) robustness metrics. In this figure the calculation is illustrated using six completing alternatives (A,B...F), each modeled under the same set of scenarios. The best (in blue) and worst performance (in orange) of each design alternative is explicitly marked. The weighted average between these best and worst observed performance of each alternative is marked by a gray marker(*). The alternative with the highest weighted average (alternative - C) would be deemed best under Hurwicz optimism-pessimism criterion. Alternative E would be deemed best under Wald's maximin criterion as its worst performance is better than the worst of others. The calculation of regret using the minimax principle tracks the performance of the alternative across all scenarios.

In each scenario, the most optimal choice is identified and the regret resulting from choosing others is calculated with respect to the optimal choice. As seen in this figure, the optimal choice can be different under different scenarios. Alternative C could be most optimal under the best case scenario (blue markers) and alternative E would be best under the worst case scenario (orange markers).

2.2.2 Use of DMUU methods in the context of BPS

The above mentioned metrics have been used in several BPS based studies that focus on task of design decision making. Some of these studies are listed below in table 2.1. More specifically, the table shows studies (non-exhaustive list)[4] that used one or more of the metrics

[4] In some studies problem specific robustness measures have been devised. These are not included in table 2.1. For example, Gang et al. [2016] combined uncertainty in performance and system reliability (low failure rate) together by monetizing both of them. Under prevailing sources of uncertainty, they calculated the operational cost and system failure cost for each alternative. Higher cost from either of the two (operation cost and system failure cost) are optimized (minimized) simultaneously to find robust optimal design.

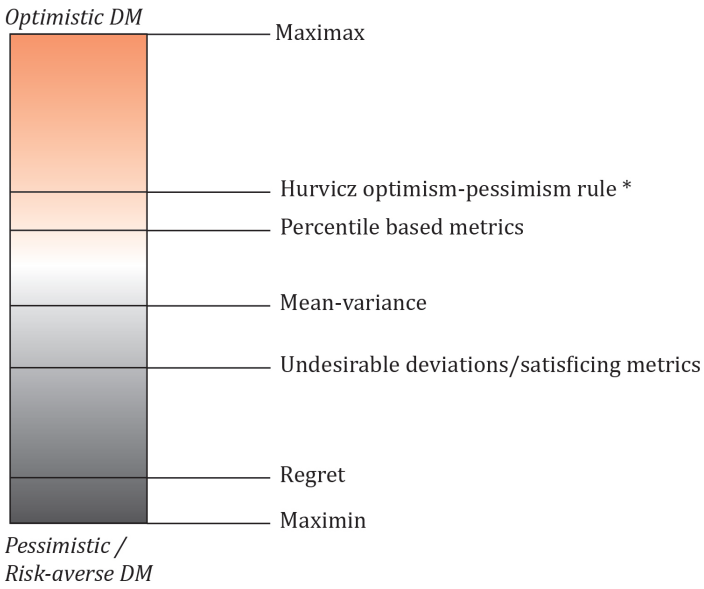


*Figure 2.5 – Robustness metrics arranged in order of risk-aversion in DM. *Hurwicz optimism-pessimism rule can be placed anywhere on this scale as is the weighted ratio of Maximax and Maximin.*

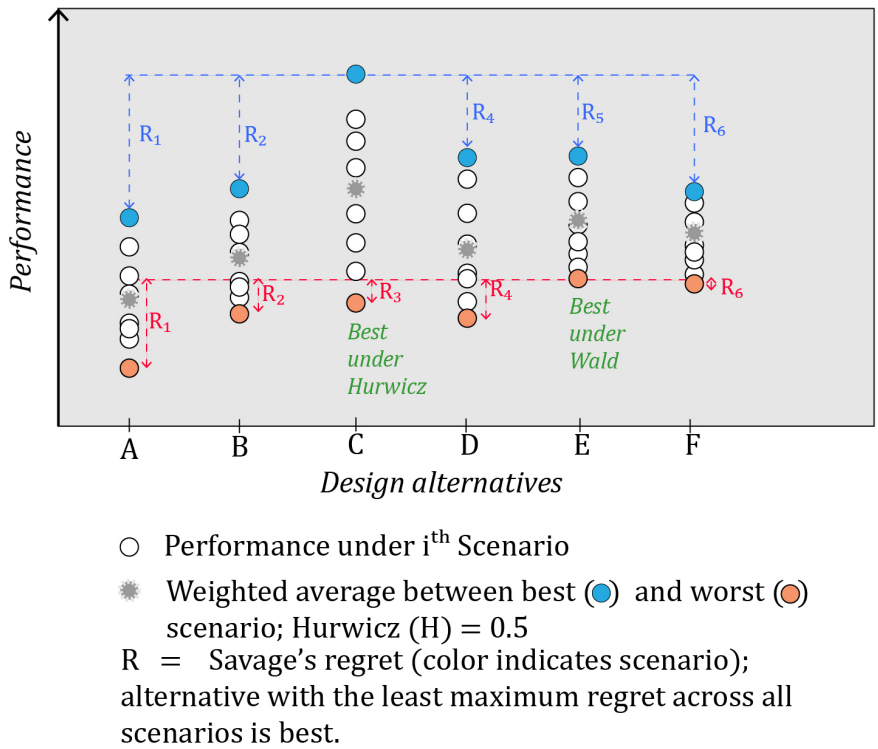


Figure 2.6 – Calculation for Hurwicz's optimism-pessimism, Wald's maximin and Savage's regret (shown partially) robustness metrics and subsequent identification of best alternative (Higher value on Y axis is considered better). The data shown is for the purpose of illustration only. The figure is partially adapted from [Rysanek and Choudhary, 2013]

mentioned above in order to rank design alternatives under uncertainty. Some studies that used satisficing metrics could also be classified under the category of 'higher-order measures'. In such cases the study is still categorized as relying on satisficing metrics as that reflects a special condition that needs to be met. The studies mentioned here are used in a few specific design contexts (arranged in no particular order):

- **Context-1: robust-optimal** In this context, a DM's task is to choose one design alternative out of several optimal design alternatives. To assist this, one or more robustness metrics can be used to compare the performance of the alternatives under prevailing uncertainty. Uncertainty could be related to design features beyond the scope of the optimization procedure or environmental conditions such as weather or operational conditions such as occupancy.
- **Context-2: most-compliant** Instead of pursuing a design that delivers the highest performance compared to other alternatives, the DM maybe interested in the design alternatives that meets a specific design requirement under a wide variety of conditions and assumptions made in the BPS models. It is common to have such requirements as a results of occupant-comfort needs, building codes or financial constraints.
- **Context-3: better-than-the-rest** Under this design context, the DM is trying to address the design-gap. However, unlike context 1, the DM does not use an optimization approach and has a limited set of feasible design solutions. The DM needs to choose the alternative that is most likely to prevail over the others.
- **Context-4: most-reliable** In this design context, the DM is interested in the design alternative that is most insensitive to prevailing uncertainties. This would provide the DM, flexibility in future planning (or design decisions) as the chosen alternative can be expected to be robust to a wide variety of conditions.

As described earlier, table 2.1 lists several BPS based studies that use one or more robustness metrics mentioned earlier (expected value metrics, higher-moment metrics, reget based metrics and statisficing metrics). While robustness is the most common key word, other words such as reliability and confidence were also used depending on the context of use.

Among the studies mentioned in table 2.1, Rezaee et al. [2015], is distinct as it used satisficing metric to meet a requirement of the decision making process (compared to the typical application of meeting a design requirement). They anticipate that DMs could have a certain threshold of performance difference (ϕ) that they would need to observe between two design alternatives . BPS model inputs $P=p_1,p_2...p_m$ were divided into two subsets, one representing design parameters that have been decided upon (P_{dec}) and others that are undecided (P_{undec}) at any given point of the design process. Features in the set (P_{undec}) become sources of uncertainty. Hypothesis testing methods were used to evaluate if the performance difference between two design alternatives would exceed (ϕ). However, such a method only attests to the

Table 2.1 – BPS based studies using robustness metrics for ranking design alternatives. The metric type used in the study is indicated by **(E=EVM; H=Higher-moment metrics; R=Regret; S=Satisficing)**.

BPS based study	**Context of use/ Objective of DM**	**Basis for ranking design alternatives under uncertainty; type of robustness metric used**
De Wit and Augenbroe [2002]	most-compliant	Expected utility from higher probability of respecting indoor comfort threshold; **S**
Domínguez-Muñoz et al. [2010]	most-compliant	Lower probability of peak load threshold being exceeded; **S,H**
Sun et al. [2014]	most-compliant	Probability of meeting the load 99.6% of all hours; **S,H**
Huang et al. [2015]	most-compliant	Probability of meeting design requirement; **S,H**
Hoes et al. [2009]	most-reliable	Low relative standard deviation (standard deviation/mean) in various indoor comfort and energy performance metrics; **H**
Nik et al. [2015]	most-reliable	Low relative standard deviation; **H**
Hoes et al. [2011]	most-reliable	Low relative standard deviation for multiple criteria; **H**
Parys et al. [2012]	most-reliable	Low difference between output under uncertainty and the idealistic deterministic case; **S**
Hopfe et al. [2013]	most-reliable	Lower range in performance under uncertainty; **H**
Chinazzo et al. [2015]	most-reliable	Robustness index based on variance with separate index for energy saving;**H**
Rezaee et al. [2015]	better-than-rest	Probability of exceeding the competing design alternative by a fixed scaler amount. This scaler amount is referred to as the design decision threshold; **S**
Gang et al. [2015]	robust-optimal	Mini-max regret, that is, the minimum of the maximum regret of design alternative; **E**
Rysanek and Choudhary [2013]	robust-optimal	Expected value and regret based metrics, ranks differ using different metrics; **E,R**
Nikolaidou et al. [2017]	robust-optimal	Several robustness metrics used to rank Pareto-optimal design alternatives. Different metrics resulted in different ranks. **E,H,R**
Kotireddy et al. [2018]	robust-optimal	mini-max regret criterion; **R**
Kotireddy et al. [2019]	robust-optimal	mini-max regret criterion;**R**

means of two distributions being differentiated by ϕ with a given confidence (say 90%). The final design may on may not correspond with the mean value.

Regret and EVM metrics were found comparable as robustness metrics and more successful in differentiating between alternatives as compared to statistical measures [Rysanek and Choudhary, 2013; Kotireddy et al., 2018]. These metrics focus on potential for loss and are well suited to a DM who wants to avoid making incorrect decision at the early design stage. Regret takes into account performance of various alternatives in a more comprehensive manner (compared to EVM) across all scenarios. It is also more conducive for relative performance comparisons as it compares each design alternative to the most optimal one (out of the given set of alternatives) under a given scenario. However, it has certain limitations. The regret metric, accounts for loss from not choosing the most optimal alternative. In case an alternative is found that has zero regret then a DM could be confident in having found an alternative that prevails over all other alternatives under all observed conditions. However, in case none of the alternatives have zero regret, that is, all alternatives are sometimes worse than other alternatives, then the DM would need further input for making a decision.

Su and Tung [2012] extended regret (opportunity loss) based decision making criterion to Expected Opportunity Loss (EOL) which incorporates probability of loss along with the potential amount of loss. EOL is a risk based (rather than loss) metric for decision making. While regret evaluates the closeness (in terms of performance) of a design alternative to the most optimal decision in each scenario, EOL evaluates the potential for loss resulting from choosing an alternative, simply given the other alternatives that are available. EOL does not rely on the identification of the optimal choice (from non-optimal ones). Rather it differentiates between the chosen and all other alternatives. EOL integrates the probability density function resulting from pair-wise comparison of a chosen design alternative to its competing design alternatives under conditions of loss.

2.3 Summary-What is needed?

While a number of studies have illustrated the magnitude of uncertainty in early design stage performance evaluations, it is not known to what extent the decision making is afflicted. Decision making rarely relies simply on the raw performance values received from the BPS models. The DM examines the performance evaluations through his/her decision making mechanisms before arriving at a conclusion. Uncertainty regarding detailed design decisions, has an understandably significant impact on performance evaluation of conceptual/early design proposals and must be considered when the DM is trying to identify alternatives that are 'most-compliant'. In such cases the the DM relies on the performance evaluation to judge the ability of the design to meet a specific performance threshold. The extent to which uncertainty at the early design stage effects relative performance comparisons and ranking of design alternatives remains understudied in three respects:

Evaluating need for uncertainty analysis in early-design decision-making: While a number

of studies have illustrated the benefits of incorporating uncertainty analysis in various design problems, the findings from these works remain tied to the case studies that are used. Also, limited attention has been given to understanding and presenting the interaction between uncertainty in performance estimates and the decision making mechanisms that DMs adopt. There is need for additional work from which findings can be generalized and can help us understand the role of uncertainty analysis in the early design decision making.

Robustness metric for relative performance evaluations: Choosing an appropriate robustness metric that is well suited to the decision making problem is essential. Relative performance evaluation is a comparison between rival or competing design alternatives. They are inherently pair-wise comparisons. The robustness metrics adopted in this context, for assisting decision making, must acknowledge this nature of relative comparisons. The expected opportunity loss (EOL) risk metric proposed by Su and Tung [2012] appears well suited to relative performance comparisons commonly done in the BPS domain but has attracted limited scientific attention. EOL is based on regret and accounts for loss under various scenarios. It is suitable to DMs who are mainly interested in avoiding loss and less concerned about potential for unexpected gains due to uncertainty. This approach could be considered appropriate for the early design stage then the DM mainly wants to avoid being wrong.

Mechanism for reporting reliability of design-decision making: The need for greater transparency in presentation of BPS results is expressed by both users of BPS tools [Attia et al., 2012a] and researchers [Clarke, 2015]. The quality of decision making, achieved using BPS tools must also be reported with clarity to all those involved in decision making. The quality of decision making depends on the specificity of the design information present in the BPS model (i.e. known versus unknown conditions), the design alternatives being compared and the decision analysis methods used. Robustness metrics can address design-information-deficiency present at the early design stage. However if the degree of known information is very low, robustness may also remain low. Thus robustness evaluation is related to degree of prevailing uncertainty in design features at the time of decision making. There is need for fast and effective communication of both the uncertainty that is present and the resulting reliability in decision making to the DM.

In addition, when a robustness metric is used for guiding decisions, an acceptable level of robustness needs to be identified. Quantifying the loss or even the risk of loss can be insufficient for decision making without mechanisms to evaluate the risk as all available alternative my carry some degree of risk. A further evaluation of the risk as being 'high' or 'low' is needed to help a DM determine further course of action such as reducing uncertainty to lower the risk.

3 A Method to Evaluate Risk in Conceptual Stage Decisions

A design project evolves and develops not just through the generation of design ideas and proposals, but also through the essential task of making decisions. The specific focus of this thesis is on conceptual stage design decisions that require a DM to choose one from multiple design alternatives. Given the low state of design development at this stage, this thesis suggests that the decision maker (DM) should consider such a choice risky (potentially sub-optimal). A decision made without giving due consideration to the large number of outstanding design decisions yet to be made (about various design features), could be unreliable.

This thesis puts forward a methodology for evaluating the risk of performance loss from early design decisions based on BPS evaluations that suffer from information deficiency. The methodology is presented using a simple, conceptual-stage decision making exercise that involves choosing one, out of two massing-schemes. The lowest scale of decision making (choosing one out of two), under a single performance criterion is used to reveal the inner workings of the method. It maybe noted that choosing one design out of two, and, one out of many, can both be treated as binary choices, where the DM's task is to divide the pool of choices into two categories - 'chosen' and 'discarded'.

Figure 3.1 shows the condition of such a DM who compares two conceptual design stage alternatives (labeled A,B) on a specific performance metric (e.g. Spatial Daylight Autonomy (sDA)) and needs to choose between them. This thesis suggests that an important choice that confronts the DM (besides the choice between the two design alternatives (A,B)) is whether to choose right away or delay the choice until more design details are determined. The black square in figure 3.1 indicates this decision fork. If the DM does not wait and chooses one (say A, indicated by A*) based on BPS results relying on assumed values for uncertain design features, then the DM exposes him/herself to some risk. If the DM is able to delay the decision making, some of the assumptions in the conceptual design stage BPS models can be replaced by actual design information.

Conceptual design stage BPS models reflect the low state of design development and can be deficient in important details. At the same time, the process to arrive at the final design

involves multiple disciplines, stakeholders, cycles of development and evaluation. Therefore, uncertainties regarding the nature of the final design cannot be promptly reduced. Therefor, this thesis attempts at answering the following research questions:

- Can conceptual design stage BPS based performance estimates be relied upon to rank design alternatives?
- How to estimate the risk of performance loss from conceptual stage design decisions, due to unknown future design choices?
- How can the risk estimate be used to improve the reliability of decision making?

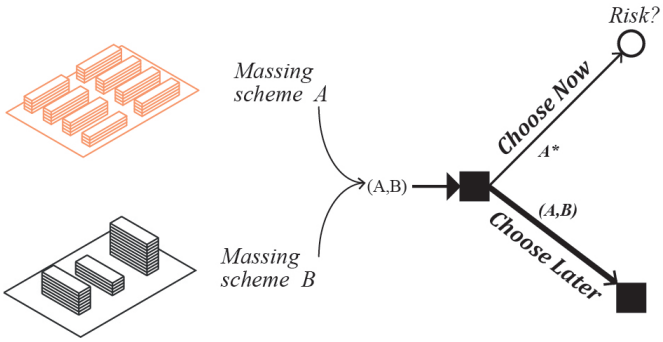


Figure 3.1 – Core conceptual stage decision making problem addressed in this thesis (graphical notations as per Raiffa [1968]).

To address the first question above, an experimental approach was taken where a number of comparisons between conceptual stage design proposals are conducted to gauge how often conceptual stage decisions could be considered reliable. The experimental setup is described in detail in section 3.2, after presenting the overall approach in section 3.1. This chapter also introduces and describes a new robustness metric for estimating risk in a given comparison between two given design proposals (section 3.2.3) [1]. Other additional issues relating to using of risk assessment in improving reliability of decision making are discussed at the end of the chapter summary.

3.1 Experimental evaluation of risk - Overview of Methodology

This thesis evaluates the risk of performance loss at the conceptual design stage due to uncertainty in design choices that are yet to be made. The risk metric used to do so is based on opportunity loss [Savage, 1951]. The risk metric proposed and used in this thesis is better understood through the following interpretation of opportunity loss-

[1] Parts of this section are replicated from a publication entitled '*Performance evaluation at the conceptual stage for neighborhood-scale design: How useful are BPS models for massing schemes?*' currently undergoing revisions to be submitted to Journal Building Performance and Simulation.

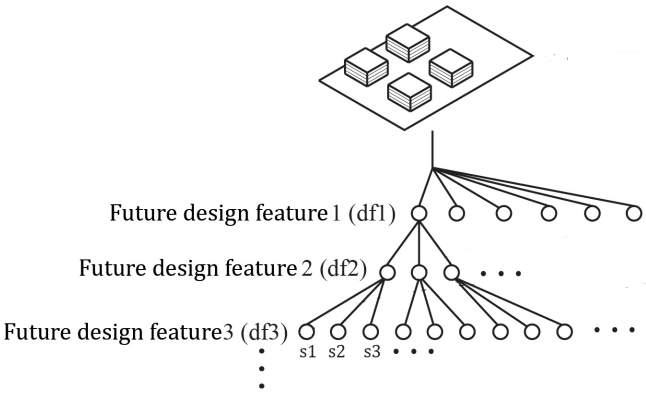


Figure 3.2 – Tree diagram showing that size of experiment grows as more design features are considered. Each circular marker indicates a specific scenario (df = design feature).

"It is the difference between the wrong choice you took and the best alternative available, i.e. the one you would have chosen if you had the perfect information" [Hubbard, 2014].

The source of risk examined in this thesis is the design information gap in BPS models between the conceptual and final design stage. In order to gauge how often such a risk impedes performance based decision making at the conceptual design stage, multiple instances of decision making were created.

To enumerate the potential information gap, each conceptual design stage alternative was modeled in its final stage of design development. However, there are many design development paths are possible. Figure 3.2 shows the 'tree' of the design paths or scenarios that grow from a conceptual design stage proposal. The tree indicates that many design paths (each branch is a design path) could be taken during the course of design development, resulting in many future design scenarios. As more design features (indicated as df in the figure) are accounted for in the scenarios, the tree grows vertically and laterally in order to represent the design possibilities within a design feature. For example, if the design feature added is 'balcony', various types of balcony designs need to be considered in the tree of design scenarios.

Numerous design features need to specified for a conceptual design to arrive at the final design stage. Also virtually unlimited design paths are possible. Thus it was important to set a 'horizon' or scope to the risk analysis. Key detailed design features that would be relevant to the risk evaluation at the conceptual design stage were selected based on the literature and the performance metrics that a DM may want to use. Owing to proximity, Switzerland was chosen as the geographical context for the experimental evaluation of risk. Residential buildings were used as the building function type for this study as they are the largest energy consumers by building type in Switzerland and several other European countries[2].

[2] https://ec.europa.eu/energy/en/eu-buildings-database

3.1.1 Metrics for early design performance evaluation

An important aspect of the experimental design was the selection of performance metrics of interest. Building performance metrics are used to assess ability of a space in providing comfort to the occupants. They typically evaluate space quality on a single aspect at a time such as daylight access, view quality, visual comfort and thermal comfort. For this study the focus was on metrics commonly evaluated using BPS tools that could be considered meaningful even at the most rudimentary state of design development. Three broad aspects of performance addressed in this thesis are daylight access, passive solar heating and overheating avoidance. Multiple metrics for assessing these design objectives and several aspects were considered in choosing the metrics for the experiment.

Indoor daylight access is often evaluated in terms of horizontal illuminance, i.e. illuminance received across the room usually at the working plane height. This is done using a grid of virtual sensors laid on a horizontal plane in the building interior in a daylight simulation model. The BPS model includes the building structure and surrounding obstructions. Usually climate based daylight modeling (CBDM) accounts for an entire year's worth of weather conditions on an hourly basis. As a result of the simulation, a time-series of illuminance values is obtained at each virtual sensor. The illuminance time-series data from a design solution can be aggregated under different metrics such as Useful Daylight Illuminance (UDI) [Mardaljevic, 2015], Spatial Daylight Autonomy (sDA) [IESNA, 2012], Residential Daylight Autonomy (RDA) [Dogan and Park, 2019] and Residential Daylight Score (RDS) [Dogan and Park, 2019]. Similarly, for estimating heating and cooling demand, various methods can be used and the scope of the analysis also has to be considered. Table 3.1 shows the various metrics that were considered and table 3.2 for notes that further explain each metric in the context of this study.

As the objective of this thesis is to assess the risk due to absence of design details in conceptual stage BPS models, the metrics chosen for this study need to be relevant irrespective of the level of design development. This thesis seeks to assess adequacy (or lack there of) of design information in conceptual stage models rather than the appropriateness of metrics used. This was the main driver for selecting the metric for the study. Another consideration was sensitivity to design features well beyond early design stage or those that could change post-occupancy.

Several metrics were found better suited for evaluation only after internal layouts have been developed. In conceptual stage models, internal layouts are typically absent. Dogan et al. [2016] proposed 'auto-zoning' algorithms that can overcome this limitation of conceptual stage models. However assumptions in the algorithm can lead to vastly different internal layouts, resulting in significant performance implications [Dogan et al., 2016]. Internal layout related assumptions enforced on the conceptual stage model could affect the risk assessment and the findings of the experiment.

All of the daylight metrics as summarized in table 3.1 (sDA, UDI -c, UDI-a, RDA, RDS) are affected (to varying degrees) by the distribution of light in the building interior, which in turn is effected by the internal layout and surface properties of internal walls, ceilings and floors.

Table 3.1 – Commonly used building performance metrics considered for the study.

BPS based performance metrics	Units and brief description	Scope of metric	Key reference; also see notes in table 3.2 for inclusion or exclusion from study
Included metrics			
Spatial Daylight Autonomy (sDA)	%area of space that meets DA (30 lux, 50%)	Annual CBDM*, active shading included	[IESNA, 2012]; (a)
Annual Heating Demand	kWh/m2-year	Annual detailed thermal modeling based, Ideal HVAC system	NA; (b)
Annual Cooling Demand	kWh/m2-year	Annual detailed thermal modeling based, Ideal HVAC system	NA; (b)
Other metrics that were considered			
Useful Daylight Illuminance - combined (UDI-c)	% time between 100-3000 lux	Annual CBDM* , no active shading neccassary	[Mardaljevic, 2015]; (c)
Useful Daylight Illuminance - autonomous (UDI-a)	% time between 300-3000 lux	Annual CBDM* , no active shading neccassary	[Mardaljevic, 2015]; (d)
Residential Daylight Autonomy (RDA)	Average DA(300 lux) over 12 time-bins	Annual CBDM , no active shading neccassary, space-by-space** evaluation	[Dogan and Park, 2017]; (e)
Residential Daylight Score RDS	Score of 0 to 24 with 1 point each awarded for direct daylight access or average DA (300 lux, 50%) compliance over 12 time-bins	Annual CBDM , no active shading neccassary, space-by-space** evaluation	[Dogan and Park, 2019] ; (e)
Overheating hours	Hours indoor temperature exceeds comfort criteria	Annual detailed thermal modeling based , space-by-space** evaluation is advisable	NA; (f)
Energy use	kWh/m2-year	Annual detailed thermal modeling based, HVAC system modeled explicitly to include system efficiency and fuel type, internal zone divisions** are advisable	NA; (g)

(*) Climate Based Daylight Metric

(**) space-by-space/zone-by-zone evaluation implies that internal layout is required or recommended for evaluation

Table 3.2 – Notes for table 3.1.

(a)	**sDA** is the percentage of floor area that receives at least 300 lux for at least 50% of the annual occupied hours. sDA metric has a dual stringency measure (compared to UDI) where it only rewards an illuminance over 300 lux and when it is achieved for more thate 50% of the time. Brembilla et al. [2018] found sDA to be least sensitive (among four daylight metrics) to reflectance properties of interior surfaces which are typically decided in the later stages of the design and can be modified post-occupancy as well.
(b)	**Annual Heating Demand and Annual Cooling Demand** can be calculated based on the thermal comfort set point recommended for heated buildings by the local energy code (for example in Switzerland, [SIA 180, 2014]). A constant thermostat set point during occupied hours (representative of a programmable thermostat). No heating equipment efficiency is input to keep parity between cooling loads and to include effects of the building design features only. An ideal system with unlimited capacity and COP 1 is input. This type of performance assessment allows the DM to understand the implications of design decision and is recommended for the early design stage ASHRAE [2018]
(c)	**UDI-c** is another annual daylight metric like sDA where horizontal illuminance values in the range of 100-3000 lux are considered compliant. Brembilla et al. [2018] found UDI-c to be highly sensitive to reflectance properties of interior surfaces. Also, UDI-c has not found to be a strong criteria for discriminating between early design alternatives. For example, Agarwal et al. [2017] found six design alternatives to be within ±10% of each other. Thus ranking conceptual design alternatives could be difficult
(d)	**UDI-a** It is likely that UDI-a is less sensitive to indoor reflectance properties. However, applying an upper limit for preferred horizontal illuminance is a subject of ongoing research (Kleindienst and Andersen, [2012]; Wienold, [2009]). For this reason, both UDI-c and UDI-a could be regarded as especially restrictive for residential buildings.
(e)	**RDA, RDS** RDA and RDS both divide the annual time-series into three diurnal bins(morning, noon and evening) and four seasonal bins(summer, fall, winter and spring) and reward illuminance in excess of 300 lux in each time bin for more than 50% of the time. RDS in addition to daylight access, also rewards direct sunlight access. RDA and RDS are meaningful to use once the internal layout has been developed. These metrics acknowledge that rooms in a residential building will have specific orientations and all rooms cannot individually meet the daylight access requirements throughout the day in every season. These metric are thus not meaningful to analyze on an open floor plan.
(f)	**Overheating Hours** Number of hours that the indoor temperature exceeds thermal comfort standards (for example in Switzerland, [SIA 180, 2014]) can be used to evaluate the design performance during the summer period. To evaluate overheating hours, detailed thermal simulations are set up as 'free running' simulation models without any active cooling systems. The indoor temperatures achieved in the absence of any active cooling are then evaluated based on prevailing comfort codes. While the metric is useful to estimate incidence of overheating, models with with lumped zoning (single zone per floor) type models do not provide usable estimates [OBrien et al., 2011].
(g)	**Energy Use** is largely calculated in the same manner as (b), but with the inclusion of the HVAC system in order to reflect the energy expense associated with space conditioning. However, energy saving strategies can be devised and tested by modeling demand for space conditioning demand.

However, consideration for internal layouts were excluded is this study to avoid imposing debatable assumptions on the conceptual stage model. Among the daylight metrics sDA was found to be most appropriate for assessment without internal layouts being present (see table 3.2). For thermal simulation, absence of internal layout implies 'lumped zoning' meaning a single zone is modeled per floor. Lumped zoning results in underestimation of both heating and cooling demand, however; the effect is particularly severe on the estimation of overheating hours [OBrien et al., 2011]. In this case, assessing performance using annual heating demand and cooling demand based metrics were found better suited.

These metrics (sDA, annual heating demand and annual cooling demand) support the main objective of the DM while choosing the massing scheme, which is to maximize the daylight/-solar irradiation received at the building exterior surface [Compagnon, 2004]. They are also included in multiple sustainable design assessment methods such as LEED [3] and BREEAM [4] [Sharifi and Murayama, 2013] and were thus short-listed for this study.

3.1.2 Identification of risk inducing design features

Figure 3.2 indicates the process of enriching conceptual stage models with various design features to generate future design scenarios. As mentioned earlier, considering all design features would lead to a prohibitively large set of design scenarios. sDA, annual heating demand and annual cooling demand were the performance metrics short listed for the study. Performance on all these metrics is closely related to quantity of daylight/solar radiation received in the building interior. Self-shading, mutual shading between buildings and inter-reflections between buildings are the key physical phenomena that influence performance on all of these metrics. All these phenomena are closely related to the geometry of the massing scheme. Other design features that also affect intake of daylight/solar radiation into the building interior were identified based on existing literature for delimiting the scope of the study.

Several facade design features were found relevant and the following criteria were used for identifying design features that (if excluded from conceptual stage BPS models) could disrupt performance ranking of massing-schemes:

1. Does the design feature **affect intake of irradiation and daylight** into the building interior? This criterion was considered as the most important in defining the scope of the risk assessment.
2. Does the design feature have a **non-monotonic relationship** [5] with the performance evaluation metrics on its own or in conjunction with other design features. For example

[3] Leadership in Energy and Environmental Design (www.usgbc.org/leed/rating-systems/)

[4] Building Research Establishment Environmental Assessment Method (www.breeam.com)

[5] Monotonic relationship between two ordered entities implies such that the rate of change in one entity results in consistently positive (or negative) change in the other. In a non-monotonic relationship increase in one entity (typically the independent entity) does not imply increase in the other

higher WWR would have a monotonic relationship with a metric like daylight autonomy if no active blinds are considered in the design. However when active blinds are considered, then an increase in WWR does not always imply greater daylight access on an annual basis. A design feature that has a non-monotonic relationship with one or more performance metrics implies that there is at least one inflection point in its relationship with the metric. Such relationships are more likely to change the nature of relative performance evaluations between different designs.

3. Does the manifestation of the design feature **change with the geometry of the massing scheme** that they are applied to? For example, a WWR of 20% results in a different window area depending on the massing scheme, as the window area will depend on the total wall area of the massing scheme. Similarly, orientation of windows, depends on the orientation of the various vertical surfaces of the massing scheme. In such as case, the same design feature, at the same value (e.g WWR), can have different impact on different massing schemes. This generates interest from the point of view of risk assessment.

Based on the above criteria, window-to-wall ratio (WWR), window area distribution per orientation, fixed and active shading devices were identified as potential risk inducing design features. Facade design factors that were excluded are discussed in table 3.3 and related table 3.4. Table 3.3 indicates if a particular design feature has an effect on performance evaluation for the chosen metrics with the underlying . For example, window placement (vertically) can have a significant impact on sDA, but does not effect the intake of daylight/irradiation on its own. Glazing properties can have a large impact on the daylight access and space conditioning demands. However due to the monotonic nature of the relationship with the selected performance metrics, glazing selection related scenarios were not included. High influence of the prevailing building code can also be expected in the selection of glazing material. Table 3.4 (i) explains the criteria for choosing the window glazing properties that were kept constant through out the experiment.

3.2 Experimental work-flow

The act of evaluating performance and choosing from among competing design proposals at the conceptual design stage was carried out in an experimental set up. The risk involved in each comparison is assessed following four main steps:

1. Two competing design proposals are chosen from a manually prepared pool of massing-design proposals for a site. Figure 3.3 shows a simple schematic diagram of the experiment. Performance of the two proposals is evaluated on a given metric (sDA, heating demand or cooling demand). No specific facade related information is considered available and the design information is limited to the geometry of massing-schemes. At this time, the decision that would be made by a rational DM based on the performance

Table 3.3 – Various facade features that were considered for inclusion in future scenarios; **D = daylight; H= heating demand; C = cooling demand**.

Facade design feature	**Significant* effect on performance metric?****	**Significant* effects on solar radiation intake into the building interior?**	**Monotonic relationship with performance metric?**	**Effected by massing-scheme geometry?**	**Notes from existing literature; see table 3.4**
Included factors	(D, H, C)	(D, H, C)	(D, H, C)	(D, H, C)	
WWR	yes , yes, yes	yes , yes, yes	no, no, no	yes , yes, yes	(i)
Exterior active shading closure rate	yes , yes, yes	yes , yes, yes	yes , yes, yes	yes , yes, yes	(ii)
Window orientation	yes , yes, yes	yes , yes, yes	Non-ordinal factor	yes , yes, yes	(iii)
Exterior fixed	yes , yes, yes	yes , yes, yes	no, no, no	yes , yes, yes	(iv)
Excluded factors	(D, H, C)	(D, H, C)	(D, H, C)	(D, H, C)	Reason for exclusion; see table 3.4
Window Placement (vertical)	yes , no, no	no, no, no	unclear***	no, no, no	(v)
Window aspect ratio	no , no, no	no, no, no	unclear***	no, no, no	(vi)
Number of windows (#)	no, no, no	no, no, no	unclear***	no, no, no	(vi)
Window sill, frame depth and type	no, no, no	yes , yes, yes	unclear***	no, no, no	(vii)
Glazing type	yes , yes, yes	yes , yes, yes	yes, yes, yes	no, no, no	(i)

(*) 5% difference on performance from one design choice or one possible value to another. For example if the effect of window frame type can be more than 5%, it is considered significant enough for inclusion in experiment.

(**) Annual performance evaluation at the neighborhood scale.

(***) An existing study that isolates the effect of these factors could not be found.

(#) Number of windows here implies constant glazed area per facade but the effect of changing number of window openings.

Table 3.4 – Notes for table 3.3.

(i)	Rode et al. [2014] examined four different urban morphological and found that the urban morphology's solar cross section has a key role to play in determining the 'ideal' glazing ratio, when glazing ratio is based on the balance between beneficial direct solar gain and conduction loss through windows. Gilani et al. [2016] show that increase in WWR (from 40% to 60%) results in an increase in higher Annual Sunlight Exposure (ASE) [IESNA, 2012] implying greater blind closure rate. Vanhoutteghem et al. [2015] did comprehensive sensitivity tests among various window properties (size, glazing materials and shading) for daylight access and thermal comfort. Vanhoutteghem et al. [2015] showed that high U value (>0.9 W/m2K) could mean reduced design flexibility in terms of WWR (low WWR needed to balance conduction loss with direct solar gains) in order to meet performance goals on heating demand, overheating avoidance and daylight (daylight factor). At glazing efficiency of 1 (SHGC/visible light transmittance of glazing) effect of choices related to WWR on daylight and overheating in space maybe most pronounced. The findings of these studies were the basis for choosing the glazing property for the experiment (see table 4 in Appendix A.3) as it ensure that WWR remains an important factor for all metrics included in the study. At the same time a U value of 1.0 W/m2K was found to compliant with the local energy code
(ii)	BLINDSWITCH-A [Van Den Wymelenberg, 2012] exterior shading manual operation model was used. In this model primary triggers for blind operation are (a) irradiation falling on the exterior surface of window and (b) depth of penetration of direct radiation. BLINDSWITCH-A was considered more appropriate compared to blind operation models based only on incident irradiation thresholds or only based on interior area of floor receiving direct illumination [IESNA, 2012]
(iii)	At the urban scale Ratti et al. [2005] compared two glazing ratio assignment strategies for an existing urban morphology (office occupancy type assumed) and found that assigning higher glazing ratio to over-shadowed facades (facades with low sky-view-factor) was found to perform better than uniform application of glazing ratio by 6-7%. Sky view factor was used as one of two criteria for assigning glazing ratio in this study as well.
(iv)	Tzempelikos and Athienitis [2007] compared impact of active and passive shading devices on demand for artificial lighting in an office building. They found that the two approaches could bring about a difference of 20% in demand for artificial lighting. The minimum depth of fixed shading that brought about significant effect on performance was 1m [Tzempelikos and Athienitis, 2007]. The minimum depth of fixed shading/balcony in this was study was kept at 1.2m
(v)	Higher window placement can result in higher sDA [Reinhart, 2014], and this effect could be greater than horizontal placement in a room or the shape of the window Vartiainen et al. (2000). This factor could effect depth of direct irradiation penetration into the room as well , resulting in interaction effects with manual blind control model BLINDSWITCH. However penetration depth of direct radiation is a secondary trigger for blind closure in the BLINDSWITCH-A model. Window height does not directly effect quantity of irradiation and daylight intake. Also decisions regarding these factors are not related to the massing-scheme geometry in any manner.
(vi)	Gibson [2014] showed that 10% difference in daylight performance could be achieved by varying window geometry and placement with respect to interior walls, along with window orientation. However, this effect size was observed when window sizes and positions were allowed to vary randomly (keeping area constant). Windows were also allowed to placed in the roof. The effect size of window geometry by itself is thus expected to be < 10%. Also, all reasons for exclusion mentioned in (v), are applicable here.
(vii)	Effect of this factor is related to window geometry. Window geometry excluded from experiment because of (vi)

evaluation is recorded. If the DM finds the performance difference between the proposals to be low, he/she can choose to stay indifferent. This decision making outcome (not choosing any specific scheme) is also challenged in the risk estimation and evaluation. Assumptions have been made regarding the threshold for the performance difference (explained further, later in this chapter), that the DM needs to observe at this step to assign ranks.

2. Several enriched (i.e. more detailed) versions are prepared for each conceptual stage design alternative and their corresponding BPS models are set up. The future design scenarios contain requisite design details, relevant for making a robust choice between the massing-schemes. These future design scenarios are compared in a pairwise manner as shown in figure 3.4. Only the final scenarios (the leaf nodes or the end nodes) of the 'tree' of scenarios are compared. Intermediate states (for example when design features df1 and df2 are added) are only shown to indicate the process for arriving that final design scenarios.

3. The loss in performance due to any disagreement in choice of massing scheme that the DM would make in the future upon arriving at developed design state is estimated with respect to the decision that the DM had arrived at in step 1. The risk calculation largely relies on methodology followed in Su and Tung [2012] where the risk of performance loss emanates from rejecting an conceptual design option that appears to be unfavorable when evaluated in the absence of detailed design information. All deviations from Su and Tung [2012] are described in section 3.2.3.

A sufficiently large set of comparable massing schemes need to be generated for this experiment to be able to conduct a diverse set of performance comparisons between potentially competing design options on three metrics (sDA, heating demand and cooling demand). The approach chosen to produce this set of massing schemes is described in the next section.

3.2.1 Experimental subjects - neighborhood scale massing-schemes

Reliability of relative performance evaluation of conceptual design stage design alternatives is the subject of this experiment. For this purpose, existing building designs could have been collected that were generated by architects for a given site. However for any given design project, a limited number of design alternatives tend to be generated (on average three design variants generated per project during the conceptual design stage Jusselme et al. [2020]). To meet the analytical needs of this experiment, a number of comparable conceptual stage massing schemes were needed so that a number of relative comparisons could be conducted. The process of generating these schemes relied on a survey of recently built residential projects in Switzerland to identify building typologies and geometrical characteristics (e.g. floor to floor height, depth for floor plan, number of floors). Existing studies that have investigated the impact of building arrangement types on site and their impact on energy and daylight related performance, also offer useful guidelines for developing massing-schemes.

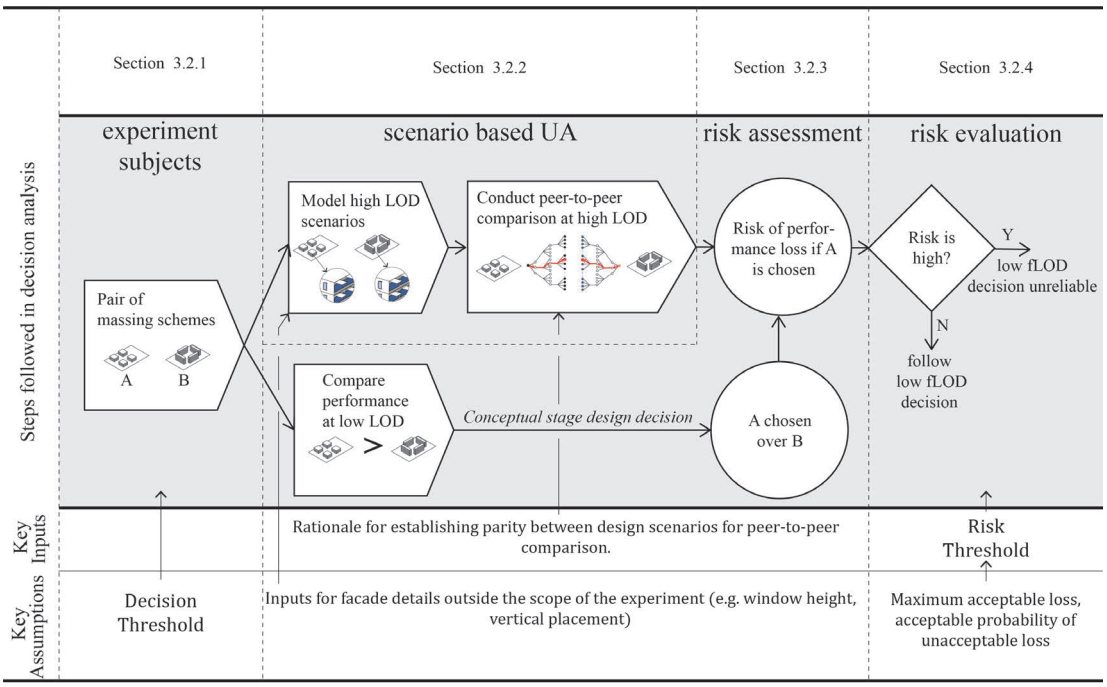


Figure 3.3 – Overview of methodology for experimental evaluation of risk.

Figure 3.4 – Pairwise comparison of future design scenarios of the two competing conceptual stage proposals.

The built density (total built area/site area) of multi-family neighborhoods in Switzerland tends to vary between 0.5 and 3.0 [Mohajeri et al., 2016]. Built density is one of the strongest known indicators of access to solar radiation at the urban scale ([Darren, 2006];[Chatzipoulka et al., 2016]). Vartholomaios [2015] estimated the maximum density for optimal passive solar gains to be 1.25. Capeluto and Shaviv [2001] contend that at density of 1.6 -1.8, it is possible to respects solar rights of all buildings in a neighborhood but is challenging at higher densities. In this thesis, a built density of 1.0 was determined to be an interesting area to investigate the effect of under-specification of design details, when the built form can potentially receive 'adequate' amount of solar radiation for implementing passive design strategies.

For setting up the experiment, drawings (plans) and photographs were gathered for recently built residential neighborhood projects in major Swiss cities (Zurich, Lausanne, Geneva, Basel and St. Gallen). Projects with less than three floors are excluded as many buildings with two floors could also be single family homes. Multi-family projects with three floors or more and constructed in years 2000-2015 were shortlisted from the following three sources in the sequence mentioned below:

- **Dichter of Zurich [2012]** - A magazine issue that collated 30 multi-family residential neighborhood projects in Zurich completed recently.
- **www.swiss-architects.ch** - A popular website documenting important work by Swiss architects.
- **City websites**- Websites of municipal bodies provide repository of all ongoing multi-family residential projects within the city limits on public land (data collection was done in November 2016).

50 projects were shortlisted as example projects based on the above criteria and were sorted into groups based on the internal layout of the buildings. Figure 3.5 shows four plan based typologies found to be prevalent in these examples. Two categories were found to clearly be the most prevalent and are described below:

1. **Linear bars** with each unit having two exterior exposures on opposite ends. 17 out of 50 neighborhoods were found be composed purely of buildings with such a plan (called *Linear bars* in figure 3.5 (a))
2. In the **block** type layout, each apartment unit occupies a corner, and each unit has two exterior faces in two different but adjoining directions. 12 out of 50 neighborhoods were found be purely composed of buildings with such a plan (called *Blocks* in Figure 3.5 (a))

These two typologies were selected for developing subject massing schemes for the experiments. Figure 3.5 shows two other (less) prevalent layout types that there observed but were not used.

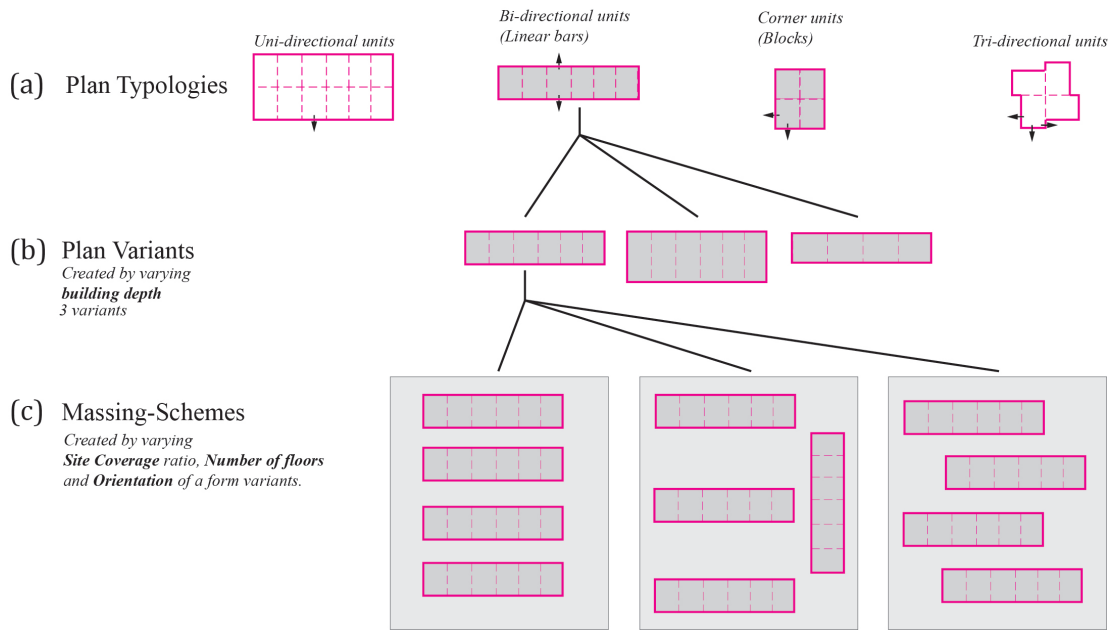


Figure 3.5 – (a) Typologies of building plans found in recently built residential neighborhoods in Switzerland (b) Plan variants created by varying depth (c) Buildings composed into neighborhood scale massing-schemes.

Building Plan Variations

Once the representative typologies were shortlisted based on the architectural plan type, their dimensional possibilities (in plan, at building level) were explored. Further variants for typologies such as linear bars could be created by changing the depth and/or changing the length of the building in plan. Change in depth of floor plan was selected as it is expected to exert greater influence on the other indoor environmental quality of each apartment unit. The range in the depth of the linear bar type building was found to be 20m to 10m in residential building examples gathered above. The length of the bar was based on the city of Lausanne bye-laws permitting a maximum building length of 35m for residential buildings [Perez and Rey, 2013]. In the example projects large variations were found in the length of the building depending on the shape of the site. However, for experimental purposes, since a fixed site was to be used, a fixed length (35m) was used for the liner bar type buildings.

No large variations were found in the aspect ratios of the block typology in existing buildings and thus only one variant type was created for the block typology. The average area per swiss apartment (125m2) was used to create the plan for a typical block type of building.Linear blocks were generated with three values of plan depth (20m, 15m, 10m) and are shown as plan variants in figure 3.5 (b).

The passive zone ratio (PZR) [Baker and Steemers, 2003] is measure of a building's self-sufficiency for passive heating and daylight. It is based purely on the geometry of the building in plan. PZR has been found to a strong indicator of spatial daylight autonomy when tested

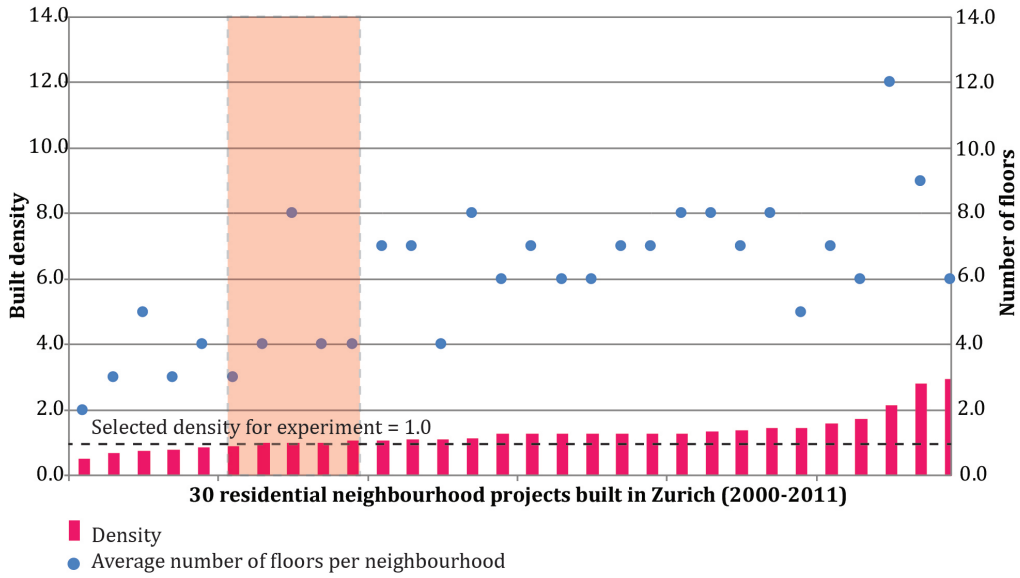


Figure 3.6 – Density, number of floors and people density data extracted for 30 residential neighborhood projects built in Zurich.

on built forms in Switzerland [Nault et al., 2015]. Three linear bar and one block type variant, respectively achieved a passive zone ratio (PZR) of 100%, 85% and 71%.

Neighborhood Composition Variants

Once the plan typologies were identified and geometrical variants of building plans within the typologies were created, these building level variants were composed into neighborhood scale massing schemes. Figure 3.6, shows the distribution of average project heights as observed in all projects (N=30) published in the DICHTER magazine of Zurich [2012], with corresponding built density. At the chosen density for the project (1.0), the variation in average building height for neighborhoods was found to be three floors to eight floors.

With built density, height and building plan variants as delimiters, 40 massing schemes were generated by manually arranging them on a rectangular site (basis for selection of size of site is described in Appendix A.1). The site is shown in figure 3.7 (a). The matrix in figure 3.7 (b) shows the compositions that were created starting from the simplest arrangement (left-hand side column) of regularly spaced buildings, all aligned in the same orientation and of the same height. Three compositions were generated on all linear-bar type buildings, namely, staggered placement in plan, regular placement but variation in building heights and courtyard shape. Impact of building arrangement related parameters (e.g. orientation, variation in building heights) have been shown by prior studies to impact energy consumption and daylight access [Cheng et al., 2006]. The courtyard shape could not be created using the block type buildings. In their case clustering of buildings was tried (instead of courtyard arrangement) which can be

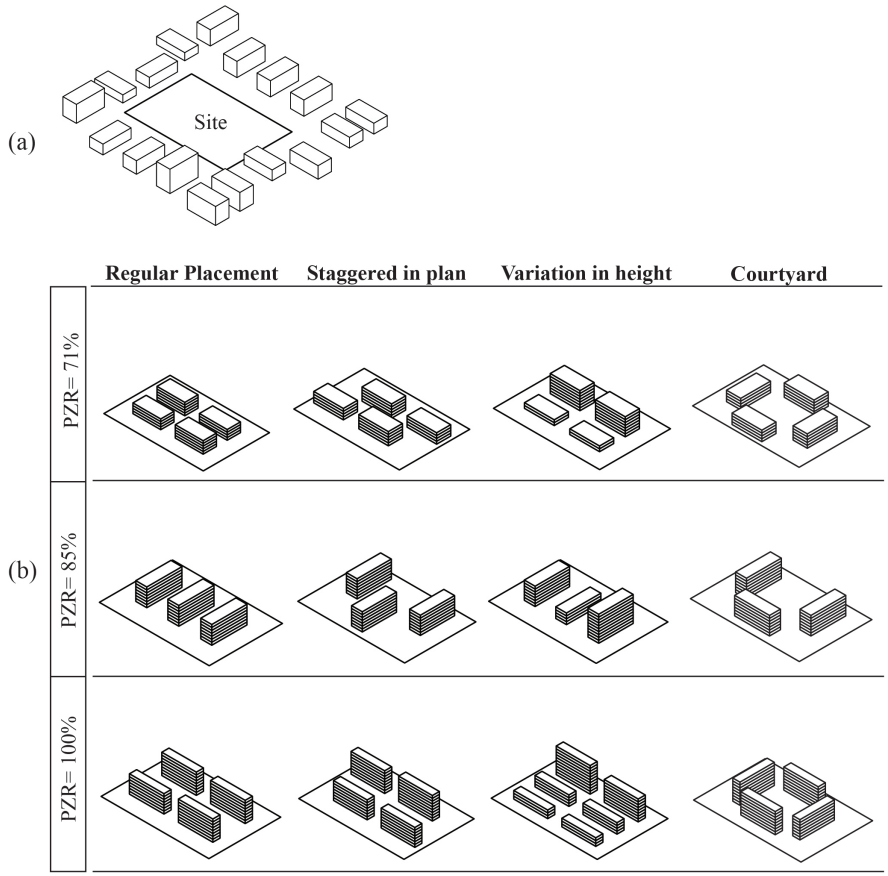


Figure 3.7 – Generation of neighborhood composition variants from simple regular arrangement of building.

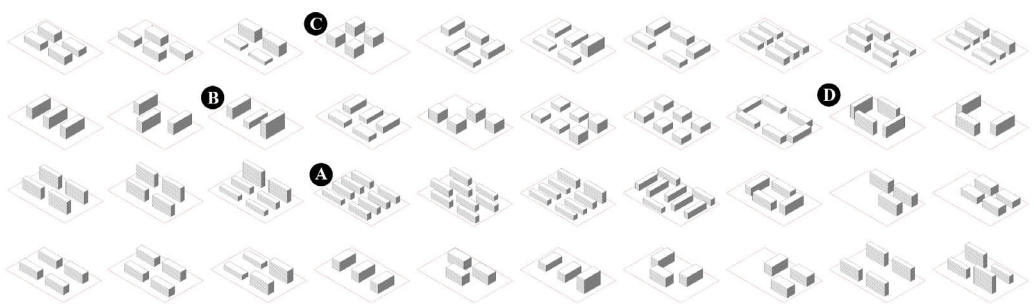


Figure 3.8 – Set of massing schemes generated as potentially competing design proposals for a given site. Scheme labeled A,B,C,D shall be used later in the chapter to graphically illustrate further working of the methodology.

Table 3.5 – Attributes of massing-schemes created manually for the experiment.

Massing scheme attribute	Maximum	Minimum
Number of buildings	8	3
Building depth (m)	20	10
Building height (in floors)	8	3
Passive zone ratio %	100	71
Site coverage ratio %	28	14
Floor to ceiling height (m)	3.5 (fixed)	
Density (total built area/site area)	1.0 (fixed)	

seen, for example in scheme C of figure 3.8. Figure 3.8 shows the final set of massing schemes that were generated from the composition process mentioned here. Table 3.5 shows a brief summary of properties of the neighborhood massing schemes. A more detailed summary of the various characteristics of the massing scheme data set used for the experiment is provided in table 5.

3.2.2 Experimental treatments - Incremental levels of detail (LOD)

In this next step of the experiment, the 3D model geometry of each massing scheme was enriched with selected facade design features to generate future design scenarios. The selected facade design features were sampled such that specific values regarding the geometrical attributes were selected in order to represent the design feature in the design scenarios. The overall process for sampling these design features is described in section 3.2.2 below.

Technique for sampling design features

Conceptual stage BPS models can contain two types of design information deficiencies. First, is the absence of design features that may not be known at the early design stage are considered too cumbersome to include in early stage models when several design alternatives are being compared. Second deficiency, is related to information that is likely to be changed at a later design and a placeholder/default value is input. The difference between a default input and the final value becomes a source of uncertainty in performance evaluation. Both of these sources of uncertainty are treated in the same manner and figure3.9 shows the method for producing future building design scenarios from these unknown design features.

Design features with geometrical or informational dependencies that often build on each other in an incremental manner[6]. The main characteristic of the method followed for sampling is nesting of design features. That is, the design features are arranged in an assumed order of

[6]At a conceptual level (nesting character of geometrical design information) was found to be similar to what is known as info-gap (IG) model [Ben Haim, 2006]. IG model *"is an unbounded family of nested sets all sharing a common structure...Ingenuity is sometimes needed to formulate an info-gap model capturing all prior information...without introducing unwarranted assumptions"*. However the overall IG decision theory supports robustness of the *satisficing* nature (refer table 2.1) and has thus not been pursued further

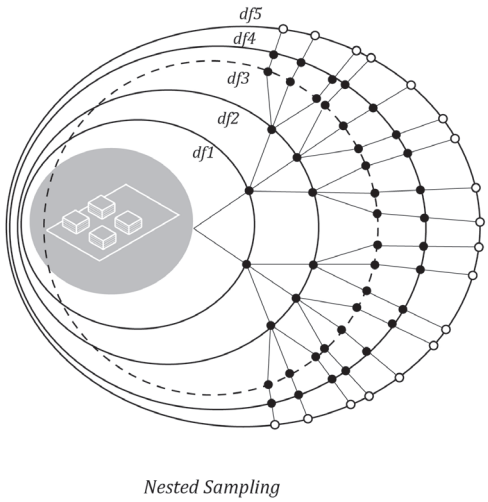


Figure 3.9 – Diagrammatic representation for the sampling method used in this study. Solid ellipses indicate design features that are to be included in an incremental manner in a specific sequence. Dashed circle indicates a design factor that is incremental in nature but free of order of specification. The white circles indicate the final scenarios. Black markers indicate intermediate states.

hierarchy. The design features that are higher in the hierarchy govern the permissible values for those that are lower in the hierarchy. The scenario resulting from the addition of each new design feature (denoted by 'df' in figure 3.9) is dependent on the features already present in a model. df1 and df2 in figure 3.9 indicate features that result in three distinct design scenarios. Multiple scenarios are considered in case they are expected to yield significant difference in performance (between the three design scenarios). df3 is binary and also sequence free. It indicates a feature that does not require other factors to be specified and has no geometrical dependencies on other factors. The binary nature reflects that multiple variations of this design feature are not found or not expected. Features df4 and df5 are also happen to be binary in this diagram. The white markers indicate the design scenarios that can emerge when design features df1-df5 are added to the conceptual stage model.

Implementation of LOD framework

In order to track and report the design features that are included in the experimental scenarios, a level of detail (LOD) framework has been used. In this framework, each new level of detail signifies inclusion of one additional design feature. There are multiple existing frameworks for LOD of 3D models. Level of Development specification (also abbreviated as LOD) framework proposed by the American Institute of Architects (AIA) [7] was found to be the closest to the approach adopted in this study[8]. However, the AIA LOD framework's hierarchy is driven by the typical design process and the sequence in which design elements tend to be specified. The

[7] https://bimforum.org/lod/

[8] The city GML standard is another LOD framework where no assumed or default inputs are permitted for unknown geometrical inputs: http://www.citygml.org/

informational priorities of architectural drawings, and BPS models are different. For example in detailed thermal simulation models, the shape and placement of windows has little impact on the performance estimate as long as the size of the window opening is accurate. This is not the case with architectural drawings. The nomenclature from the AIA LOD framework was not found align with the needs of BPS models.

A custom LOD framework was used in this thesis where the lowest facade level of detail (fLOD) is designated as facade fLOD0. At fLOD0 all facade design related information is assumed and subject to change. The specific inputs at fLOD0 and at higher fLODs are as follows:

1. **fLOD0:** At fLOD0 windows are input at an assumed 30% WWR as simple punched windows, uniformly distributed on all faces, all with the same height and sill height (2m and 0.75m respectively). The number of windows per face is estimated based on the number of apartments per floor. Once this number is determined at fLOD0 for a particular massing-scheme, it is kept the same at all fLODs.

2. **fLOD1 no-blinds:** In the second step, three variants are created, with three possible values (20%, 30%, 40%) of WWR. Each WWR value is used to revise the total resulting glazed area, which is then distributed uniformly on all vertical faces of the massing-scheme same as fLOD0. The window height and sill height is also kept unchanged.

3. **fLOD1:** In step 3, active shading devices are added in the form of shading schedules (on/off type) to be used in annual daylight, dynamic thermal simulation models. The fLOD number was not increased as the blinds are not modeled in terms of their geometry, but only as on/off schedules.

4. **fLOD2:** In the fourth step, the uniform distribution of glazing (on all faces) is modified to reflect a designer's intent to identify primary and secondary facades. For the secondary facade, the WWR is at least 10%. The remaining glazed area is assigned to the prominent facade. If the prominent facade's WWR begins to exceed 60%, the excess window area is re-assigned to the secondary facade. This trial-and-error process of assigning window area between various facades has been programmed into the grasshopper based workflow. Four variants are modeled at each fLOD1 WWR value, one where the primary facades are those with high sky view factor (SVF), second where primary façades have the lowest SVF, in the third primary façades face east or south, and fourth, primary façades face west or north. Active shading operation schedules are revised in accordance with this new distribution of glazing. At this fLOD 12 facade variants (3 X 4) are derived.

5. **fLOD3/high fLOD:** fLOD3 is the final step. Four possible balcony types are assigned to the prominent façade, as identified in step 4. The balconies are only assigned to the primary facades identified at fLOD2. Active shading operation schedules are revised again. This final fLOD, results in 48 façade variants per massing-scheme (3 X 4 X 4)

The conversion of 3-D models from fLOD0 to high fLOD was done using a custom work flow in grasshopper (a plug-in for Rhino 5.0 [McNeel, 2015]). Figure 3.10 shows an exploded view

of a massing-scheme at high flOD. The continuity of the design process and geometrical hierarchies between building elements is respected in generating high fLOD variants. The total window area is the first detail to be specified in terms of WWR, followed by the window placement/distribution related specifications. Once the window distribution is specified, the placement of balconies can be decided (glazed doors are also referred as windows).

Several combinations were contemplated for deciding the size of the experiment (48 high fLOD variants). For each design factor, any number of scenarios could be generated. The simulation time needed for the experiment was considered and sample size of high fLOD relative performance values that could be derived from 48 scenarios. Using the peer-to-peer comparison method mentioned in 3.2.3 48 scenarios would result in 192 possible relative performance comparisons. This was considered satisfactory for estimating the probability of loss. Overall this experiment thus called for 2,720 simulations per metric. The number of daylight simulations was 8 times this number given the simulation method followed (described in section 3.2.2) and total simulation and post processing time was estimated to be 35 days for daylight simulations using a 30 core machine. The estimated time for the daylight simulations was the chief factor for the scale of experiment from the point of computation time. The detailed thermal simulations were estimated to take 15 days utilizing a single core machine.

Generating future design scenarios

While the fLOD framework does provide a structure for the process of enumerating future design scenarios, there still remain numerous ways of specifying a particular design detail. For example, at high fLOD, when all opaque elements of a fixed balcony need to be specified, should the same type of balcony be specified on all the facades within a neighborhood? Should all window openings be associated with a balcony? In order for the risk assessment to be realistic, the facade design scenarios need to be coherent designs that an architect could actually produce. Architectural design process is complex, involving many different practices in order to make decisions. To generate design scenarios at high fLOD, the following questions needed to be answered:

What is the level of complexity in facade designs of residential buildings that is practical, or considered suitable by architects?

To address this issue, 30 projects were shortlisted from the set of 50 projects mentioned above and facade designs in these projects were studied. Observations were distilled into programmable rules. Characteristics relevant to the facade design features short-listed for the study (WWR, window area distribution, fixed shading/balconies) were documented and organized on a 3X3 matrix as shown in Figure 3.11. The x-axis indicates the number of balcony and glazing configurations found on the most complex facade of a project. On the y a-axis the number of distinct facade treatments found in a particular project are indicated. The scope of this study in terms of facade complexity was limited to two types of window configurations (windows with balcony attached and windows without balconies) and two types of facades

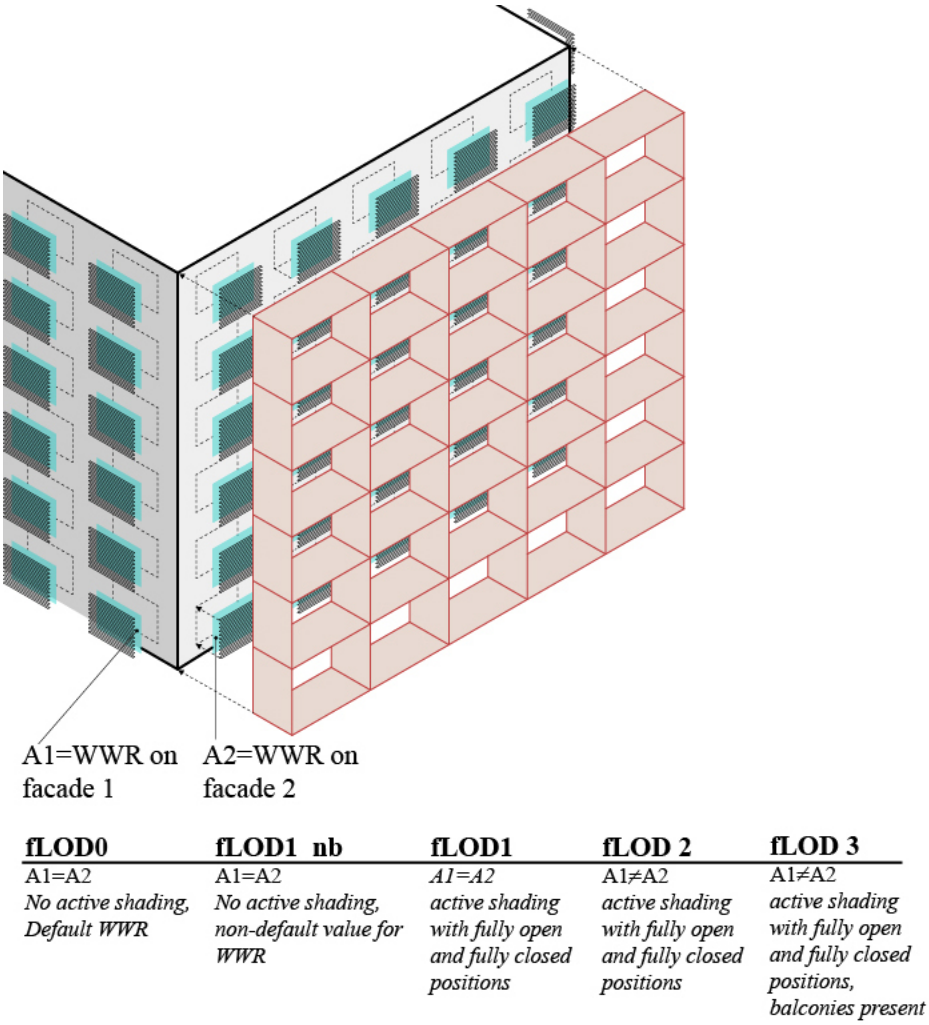


Figure 3.10 – Exploded view showing incremental levels of façade detail at which 3-D models are produced by the grasshopper workflow.

per building. More complex facades were found but were less prevalent.The average WWR value was found to be 30% and was used as the default WWR input at fLOD0.

At the same time, the facade design scenarios generated in this study are not meant to be representative of all architectural design possibilities. Rather, the intent behind the grasshopper-based workflow developed was to generate facade variants that could potentially diminish or amplify performance observed at fLOD0.

In order to ensure that rank changes that are observed, occur only due to the intended detail types included in the fLODs, it has been ensured that other design factors (number of window openings, vertical placement and height) remain unchanged. Window height (2m) and sill height is kept constant (0.75 m), irrespective of the WWR and all other factors. The only exception to this rule is allowed where balconies are present, the window is dropped to the

Figure 3.11 – Organization of Swiss residential building facades in levels of complexity. Images drawn from www.swiss-architects.ch and design firm websites.

floor level(+0.05m) to include reflection of radiation from the floor of the balcony into the building interior through a glass door/window panel. Number of windows is also kept fixed irrespective of fLOD. The effect of the factors included in high fLOD scenario need to be evaluated without introducing 'noise' from excluded facade factors. For example, the number of windows per facade per floor remains the same across all levels of detail and all scenarios.

Table 3.6 lists specific values (or range of inputs) corresponding to each fLOD. As mentioned earlier, the mean WWR was found to be 30% in the survey existing building facades. ±10% in value of WWR were chosen as experimental bounds. An overall WWR of less than 20% is not expected given that currently observed trends of high WWR in residential projects [9]. The maximum WWR on a single facade was assumed to be 60% (roughly equivalent to 2m high glazing all along the facade). The minimum WWR ratio was assumed to be 10% (roughly equivalent to 2 m x 0.5 m windows per apartment per facade). The minimum balcony depth (when present) was assumed to be 1.2m based on observations in [Tzempelikos and Athienitis, 2007] (discussed earlier in table 3.4).

[9] Increasing emphasis on daylight, view satisfaction and livability is seen though projects such as Model Homes 2020 [Foldbjerg et. al. 2015]which have so far reported high user satisfaction in post-occupancy evaluations. These homes are designed for 5% daylight factor in main rooms (instead of the typical 2%). This was achieved in some model homes using twice the amount of glazing area compared to what is typically seen for single family homes.

Table 3.6 – Façade inputs incrementally added/changed with each fLOD; values in table indicate changes per fLOD; **n.c.** indicates no change from lower fLOD.

Facade Attributes	**fLOD0**	**fLOD1-nb**	**fLOD1**	**fLOD2**	**fLOD3**
WWR	30%	20%,30%,40%	n.c.	n.c.	n.c.
WWR per facade	Uniform, 30%	Uniform, 20%,30%,40%	n.c.	Varies 60% to 10%	n.c.
Active shading	none	n.c.	Manual operation model	n.c.	n.c.
Fixed shading	none	n.c.	n.c.	n.c.	Present, depth varies from 1.2 m-2.4 m
Window placement					
Number of window units per facade	depends on size of floor plate	n.c.	n.c.	n.c.	n.c.
Window height	2.0 m	n.c.	n.c.	n.c.	n.c.
Window width	0.5 m or more	n.c.	depends on WWR per facade	n.c.	
Sill height	0.25 m	n.c.	n.c.	n.c.	0.25 m (0.05 m when balconies present)
Glazing Properties					
Visible light transmittance	0.6 (fixed)	n.c.	n.c.	n.c.	n.c.
SHGC	0.58 (fixed)	n.c.	n.c.	n.c.	n.c.
U-Value (W/m2-K)	1 (fixed)	n.c.	n.c.	n.c.	n.c.

Figure 3.10 shows the various details mentioned above that are applied to all massing-schemes in the experiment set shown in Figure 3.8.

Implementation and outcome of facade scenario generation methodology

The above mentioned steps for generating facade design scenarios were implemented using a grasshopper workflow that 'reads' the geometry of a subject massing-scheme and adds various facade details at low, high and intermediate fLODs. Each facade element is 'tagged' using layer names and object name attributes in the Rhino environment. This tagging is useful for calling out the facade details later, specific to a fLOD and scenario. The custom workflow was set up

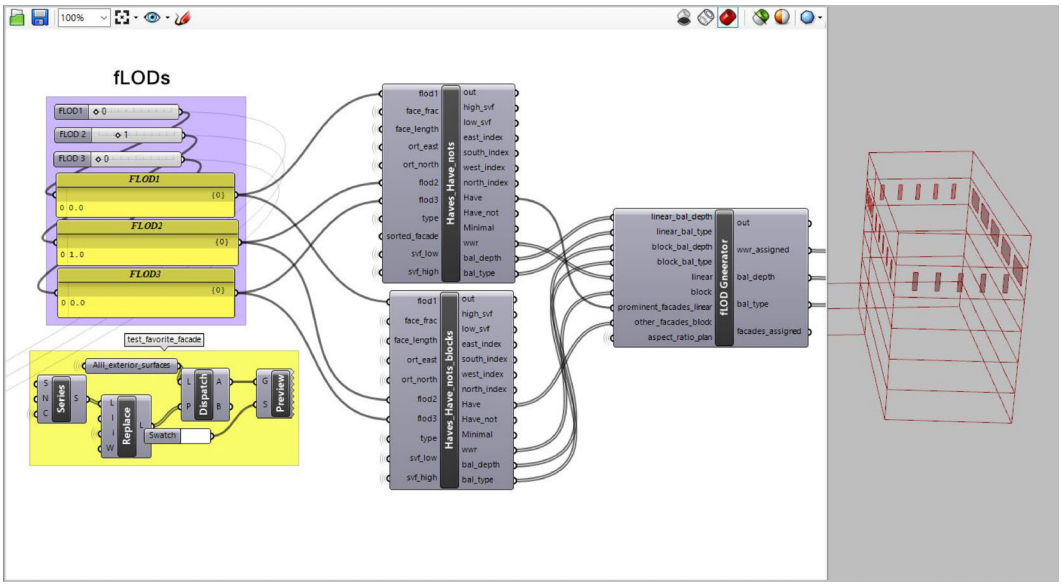


Figure 3.12 – Screen-shot of grasshopper workflow that 'reads' massing scheme geometry and assigns facade details as per the fLOD and scenario.

using a combination of available and additional components (created using Python, shown in figure 3.12). Figure 3.13 show partial set of scenarios for massing-schemes labeled C in figure 3.8.

Conversion of 3-D models to BPS models

As the final step of data generation for the experiment, all 3-D models developed at various fLODs were converted into BPS models and the corresponding performance values were calculated. In the BPS models, material properties of all envelope components, surface properties and simulation parameters were kept largely constant in all simulations conducted at all fLDOs.

All material property related inputs are specified in Appendix A.3. They were assigned while converting the geometry models into BPS models using Grasshopper, Honeybee and Ladybug plug-ins for Rhino 5.0. BLINDSWITCH-A [Van Den Wymelenberg, 2012]model was used for generation of blind schedules. The same blind schedules were used in both daylight and thermal simulations but the blind schedules are updated for each scenario at each fLOD to allow for interactions between active blind operation and other design details. BLINDSWITCH-A model takes external incident irradiation on the window and penetration depth of direct sunlight into the building interior as inputs. However partial blind occlusion positions in the BLINDSWITCH-A model were ignored given the scale of the simulations. The .epw weather file for Geneva, Switzerland was used.

Method for calculation of sDA: Eight DAYSIM simulations were carried out per scenario per

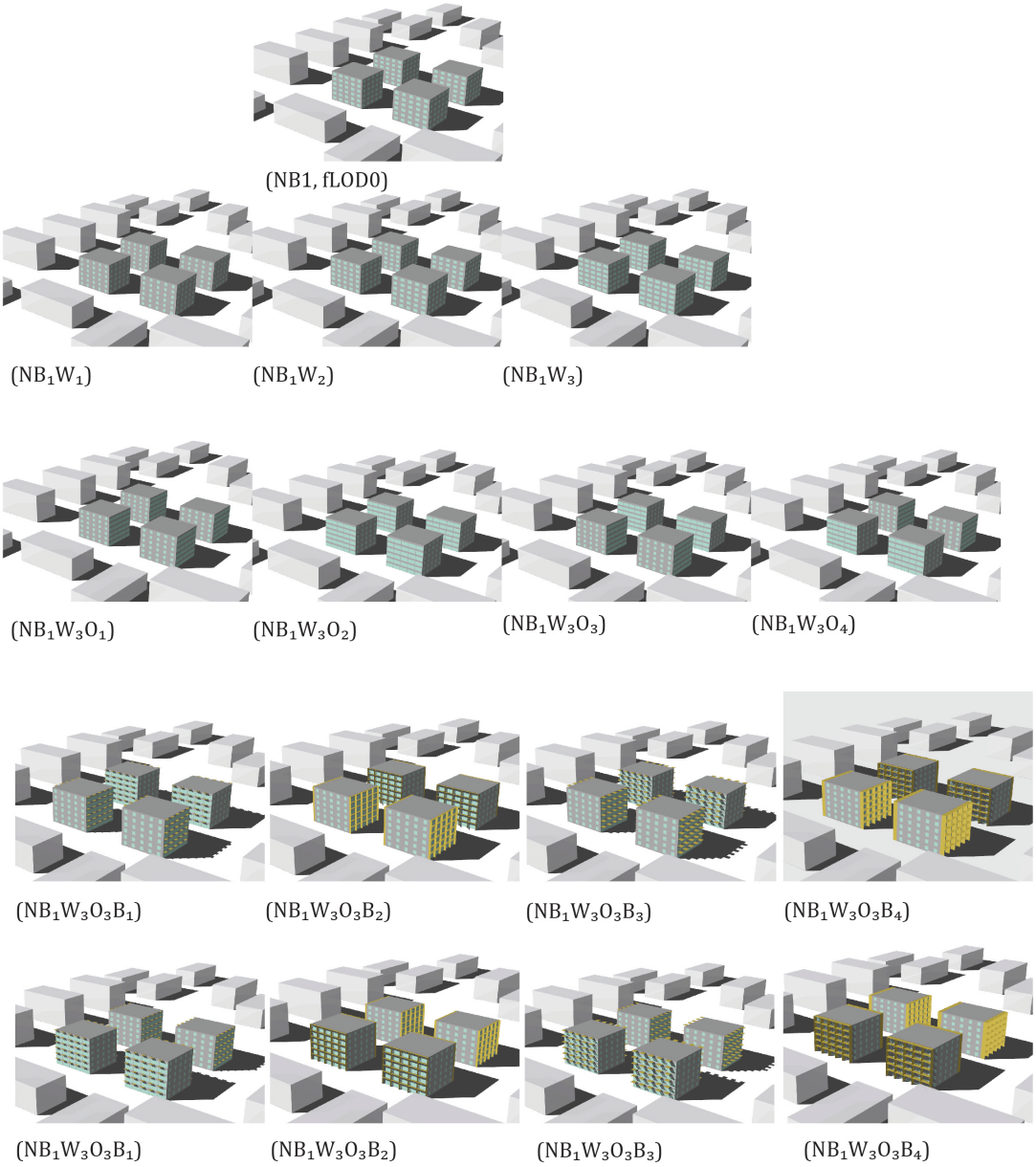


Figure 3.13 – Result of facade modeling workflow in grasshopper on massing-scheme 'C' in Figure 3.8. At high fLOD 8 out 48 variants are shown.'W' denotes WWR, 'O' - orientation related scenario, 'B' - balcony design related scenario, 'NB' - denotes neighborhood schemes

fLOD to model blinds in open and closed position per building facade. The eight files were then compiled into one composite illuminance file while applying the blind schedule specific to each building facade, using Matlab. The sDA value was also calculated using Matlab. Radiance simulation parameters input are based on Illuminating Engineering Society (IES) LM-83-12 including indoor and exterior reflectance properties (See Appendix A.3 for inputs used). 5 ambient bounces were used in all simulation where blinds are modeled in fully open position. 6 ambient bounces were used in all simulation where the blinds are modeled in a closed state (as trans material to represent combined transmittance of the closed blinds and the glazing). In order keep the size of the result files reasonable, the grid size for the placement of the interior photo sensors was kept 1m x 1m and was laid in the top and bottom floor of each building.

Method for thermal simulations: All detailed thermal simulations were carried out in EnergyPlus 8.3. All simulations were run using the solar radiation model "FullExteriorWithReflections" in EnergyPlus that accounts for shadow patterns on exterior surfaces due to detached piece of shading such as over hangs. Exterior reflections were also accounted. However interior distribution of radiation on interior surfaces was not calculated explicitly. Due to the lumped zoning type, more advanced models such "FulInteriorAndlExteriorWithReflections" were not considered necessary either. Monthly heating demand and cooling demand values were extracted. These were converted into annual energy use intensity values using Matlab.

Visual examination of simulation outputs: some examples

Two examples of raw performance values from simulations at fLOD0, fLOD3 and the intermediate fLODs are shown in Figure 3.15. The specific cases shown in Figure 3.15 are for massing-scheme pairs (A, B) and (C, D) shown in 3.8 and 3.14 here. The overall shaded regions show the range of performance possibilities at each fLOD. The sub-regions in darker shade show performance values for the pair of schemes at a particular WWR mentioned in the figure caption. While this figure does not clearly show conditions of loss, as those can only be observed when the scenarios are compared one-to-one, the figure was a used to visualize potential overlaps in performance among peer fLOD3 variants. On the sDA metric, in the comparison between schemes A and B, high chances of rank reversal are seen at fLOD1 and after.

Such plots were generated for several massing scheme pairs to understand the nature of the data. For example, one can tentatively observe a diverging effect (increase in difference between two schemes being compared and not within each scheme's performance) in the evolution of heating demand for schemes (C, D) where the performance difference between C and D is negligible at fLOD0 but for some façade designs, the performance difference between them could be significant at fLOD3. On cooling demand, in both pairs of evaluations, it maybe observed that the performance of both design options converge, and thus that it is likely that an appreciable difference in performance is not observed at fLOD3. On sDA based comparison of (A, B), the possibility for reversal of ranks is seen at fLOD1 and beyond. For (C, D) there appears to be no change in ranks at any fLOD on the sDA metric.

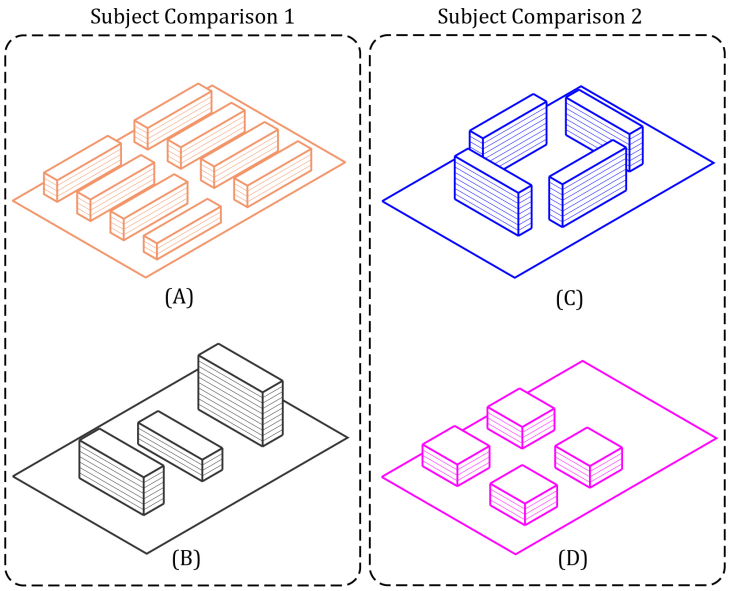


Figure 3.14 – Example pairs of massing schemes pairs (A,B) and (C,D).

3.2.3 Calculation of risk

Figure 3.16 (b)[10] shows the decision making problem modeled in this experiment in an abstract manner. The black squares indicate decision points where a DM must act. The circle is a chance node where the DM encounters risk due to unknown/ future design decisions based on the decision made upstream from the chance node. This decision tree is adapted for a BPS user who can get BPS based performance estimates of the design alternatives at any stage of the design process. The BPS based DM's dilemma is whether to abide by the findings of the current performance evaluation or wait till more design information is obtained. If the DM chooses to wait then he/she is led to another, similar, decision point later on in the design process to reconsider the decision based on a more detailed model. If he/she decides to abide by the current performance evaluation, then there is risk of performance loss due to pre-mature decision making.

In the experiment, for each instance of risk assessment, a pair of massing-schemes is drawn from the subject massing-scheme set. These are referred to as schemes Ai and Bi. A preferred massing-scheme is identified based on the performance values obtained at fLOD0. This event is notified as (Ai*, Bi) where Ai is the preferred massing-scheme or (Ai, Bi*) in case Bi is the

[10]For comparison, in figure 3.16 (a) shows a typical decision tree where the DM is interested in both design alternatives and in knowing the risk involved in choosing either. Such a tree is more common in business related decision making problems where information related to outcomes of decisions is scarce and cannot be assessed at arbitrary points in the decision tree.

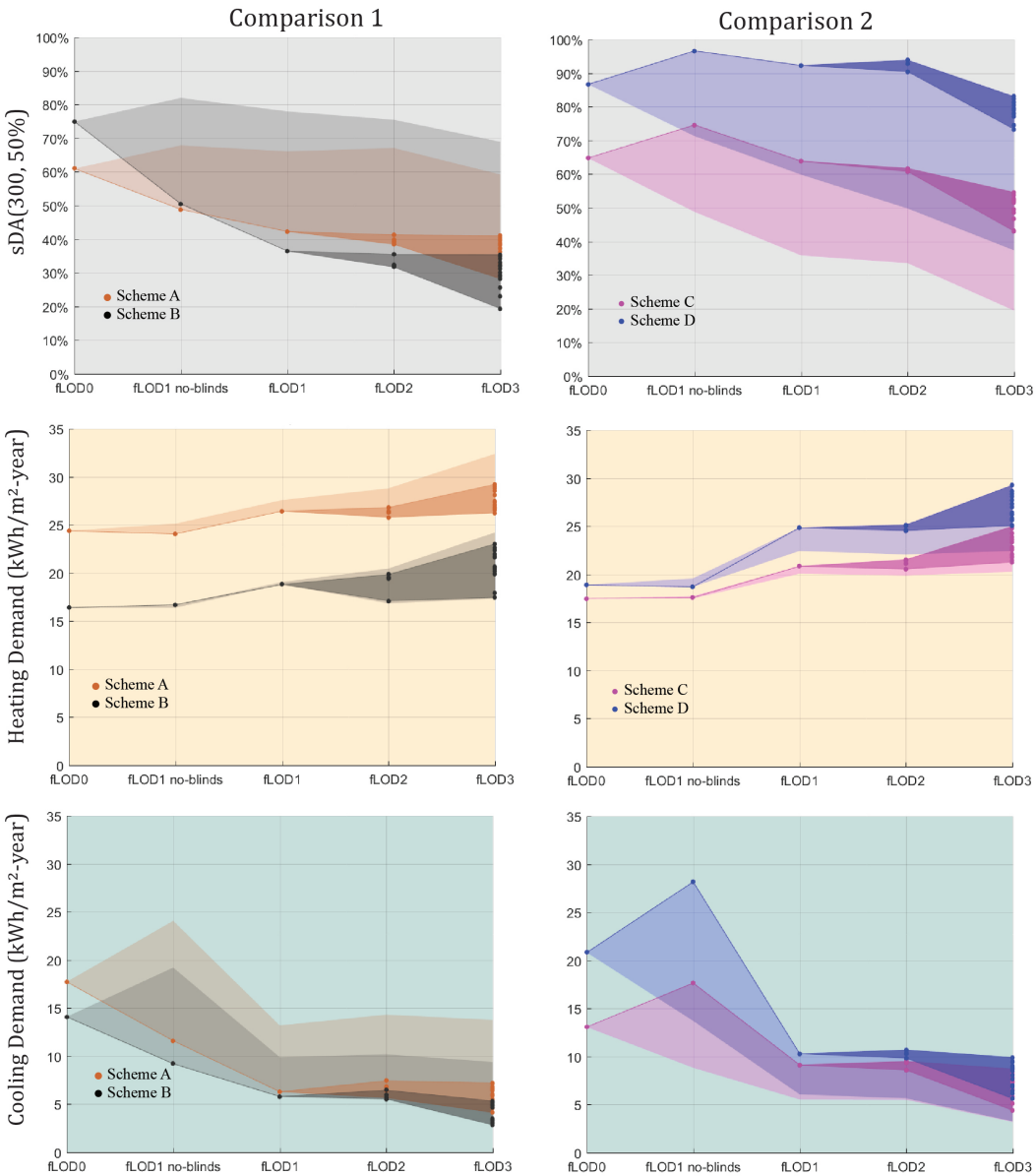


Figure 3.15 – (Left Column) Evolution of performance of design options A,B shown in figure 3 and 8 (Scheme A in orange, Scheme B in gray), on three metrics (1) sDA, shown on top (2) Annual Heating Demand, middle (3) Annual Cooling Demand, bottom. The highlighted regions show an evaluation of performance when 'low' WWR (20%) is decided upon by the designer at fLOD1.(Right Column) Evolution of performance of design options C,D shown in figure 8 (Scheme C in pink, Scheme D in blue), on three metrics in the same order as column on left. The highlighted regions show evaluation of performance when 'high' WWR (40%) is decided upon by the designer at fLOD1.

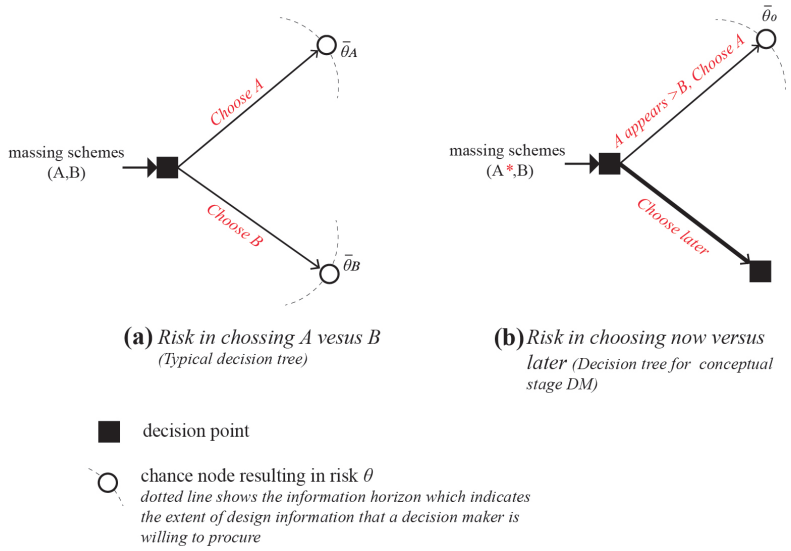


Figure 3.16 – Decision tree diagram for a conceptual design-stage DMs.

preferred massing-scheme.

The risk involved in the decision (A^*, B) is estimated based on the comparison between (A,B) using their respective high fLOD design scenarios. Su and Tung [2012] used a pair-wise comparison of possible performance outcomes under different scenarios. In the context of architectural design problems, it may be required to go beyond a strict one-to-one mapping or pairwise comparisons. Cross-comparison between some design variants maybe advisable if an equivalence between scenarios cannot be established.

For example, consider two fLOD3 variants for two massing schemes in figure 3.17. The two variants shown have both been developed using the same WWR, same balcony depth and same logic for choosing the prominent facades. The facades with high sky view factors are assigned higher WWR and balconies. However factors within the massing-scheme, i.e. building typology ('blocks' versus 'linear-bars') and orientation of buildings also influence design configuration at fLOD3. While figure 3.17 (b) shows a fLOD3 variant with 'outward' facing balconies, such a clear pattern is not seen in the corresponding variant for the other scheme in figure 3.17 (a). This induces the possibility that the designer may take a different design approach (e.g. method for choosing prominent facade) depending on the massing-scheme. Thus for comparing future facade design alternatives, a peer-to-peer comparison approach was taken. Figure 3.18 shows the difference between this approach and a pair-wise comparison between scenarios, diagrammatically.

A few local architects were consulted to determine design features that could be subjected to a strict pairwise comparison and those in which direct equivalence is difficult to establish. Five architects were asked to discuss a residential neighborhood project that they had recently worked on and the design alternatives that they considered during the process. The

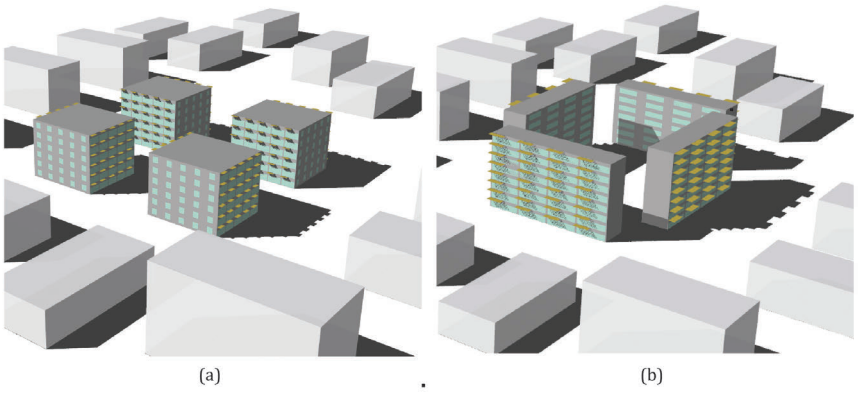


Figure 3.17 – fLOD3 variants of two massing schemes generated through the same grasshopper workflow that favors facades with high sky view factor to assign more glazing and balconies.

discussion topics, subject projects and all conclusions drawn from these meetings are described in Appendix A.1. The discussions suggest that orientation of windows is one of the factors considered when developing a massing scheme. Different massing schemes offer different opportunities in terms of window distribution and orientation. Similar orientation of windows cannot be achieved in two different massing schemes. An equivalence cannot be established between window orientation related facade design scenarios and scenarios and cross-comparison between scenarios with different window placement (in terms of orientation) may be allowed. The discussion revealed that WWR and balcony depth are fairly independent design decisions that are not related to the massing-scheme. In these cases only a strict pair-wise comparisons would be considered valid. An architect would not deem different balcony depths assigned to two massing schemes as comparable designs.

When the 48 high fLOD variants of the two massing-schemes (Ai, Bi) are compared, 192 valid peer-to comparisons are produced, allowing for cross-comparison between four orientation-related variants (48 X 4 =192). This leads to relative performance difference of the two given massing-schemes as a distribution with N=192. Figure 3.19 shows the outlay of all high fLOD scenarios. The paths to arrive at the high fLOD scenarios are also shown. An example high fLOD scenario of both the schemes are also shown.

The various relative performance evaluations that are possible at fLOD3 (N=192) are used to produce the probability distribution function of future relative performance difference between the two given massing-schemes. The expected opportunity loss (EOL) [Su and Tung, 2012] or risk is then expressed as:

$$EOL(A^*, B) = -\int_{-\infty}^{0} f\Delta(A^*, B)d\delta \tag{3.1}$$

where $EOL(A^*, B)$ is the EOL when A is the preferred massing-scheme after a conceptual stage

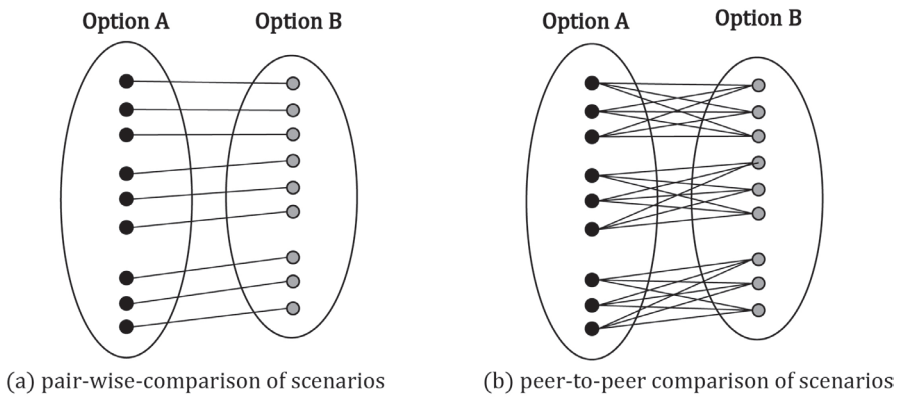


Figure 3.18 – (a) one-to-one-mapping in pairwise comparisons [Su and Tung, 2012], (b) example peer-to-peer where scenarios within specific groups can be treated as comparable peers to each other.

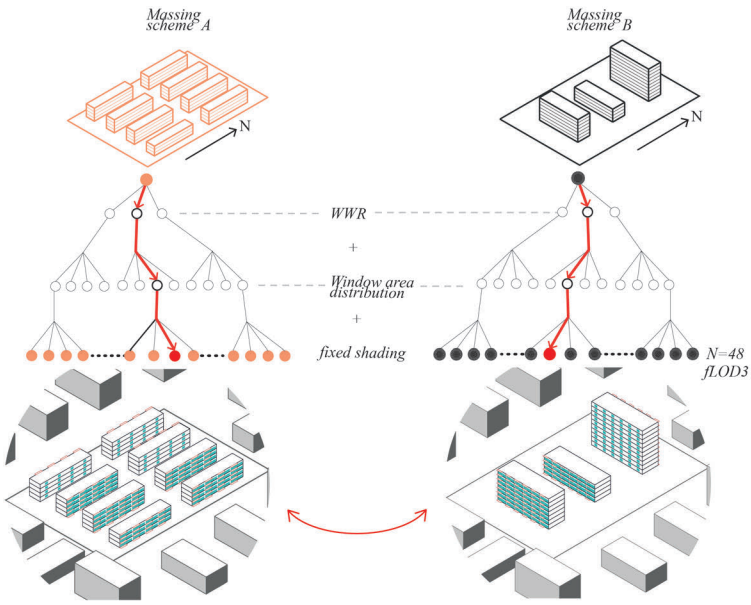


Figure 3.19 – Diagram showing number of design variants generated at each level of detail. Two example massing-schemes (SE, isometric view) are shown with one instance each, of fLOD3 variants. The fLOD3 variant shown (red dot marker) has 30% WWR, non-uniform distribution of windows to favor vertical surfaces with higher Sky View Factor. Simple projected balconies are assigned to facades with higher glazed area, and are 1.2 m deep, covering 50% of the horizontal length of the facade.

performance evaluation, and $f\Delta(A^*, B)$ is a probability distribution function (PDF) of the relative performance gain from design pairs formulated later in the design process.

Figure 3.20 shows an example comparison between a pair of massing-schemes. In this figure (left-column) only strict pairwise comparisons at fLOD3 are shown on three metrics (sDA, annual heating demand, annual cooling demand). On sDA, contradiction is seen in ranks that would be assigned to the two massing scheme design at fLOD3 (top left panel) even in strict pair-wise comparisons. The probability distribution function (PDF) of relative performance difference values at fLOD3 from the same pair of massing (top-right panel) shows the EOL (shaded region). The PDF is generated from the performance difference values from peer-to-peer comparisons. No EOL was found on heating demand and a small value of EOL is found on cooling demand.

In the following sub-sections show illustrative examples of what the PDF of the relative performance gain from future design scenarios could look like in more generic terms. Other forms of loss (other than rank-reversal) than those described by equation 3.1 are also presented.

Opportunity loss due to rank reversal

As mentioned earlier, the methodology presented in this thesis is based on the premise that a DM is able to conduct a BPS based performance evaluation at any given stage of design development. Anytime a BPS based relative performance evaluation is done, a ranking of the design alternatives can be done. If the DM abides by such a ranking, the risk of performance loss from yet to made, future design decisions can be assessed using the EOL risk metric. The EOL metric (equation 3.1) essentially estimates the risk of performance loss due to ranks being reversed. This is the type of opportunity loss that a conceptual design stage DM is likely to be most averse to, where the design alternative chosen at the early design stage, would very likely be rejected if the DM did the same comparison later on in the design design process.

The risk is assessed by enumerating future design possibilities and calculating the performance of each of the future design scenarios. Figure 3.21 shows an example PDF of future relative performance evaluations indicating high likelihood of rank reversal. The blue vertical line indicates conceptual stage performance evaluation (at fLOD0) on the x-axis. This evaluation allows the DM to establish ranks between the design alternatives. Future relative performance values are calculated, while respecting the ranks established at the conceptual design stage. All future evaluations that yield negative results are source of 'regret' and contribute to opportunity loss.

A pre-requisite to observing loss from rank reversal is that clear ranks need to be established between the design alternatives. However it is unlikely that a DM establishes ranks based on minor performance differences. For example Iversen et al. [2013] called 10% difference on sDA as appreciable as a matter of common practice. However, it is not clear if such a threshold is based on expected error in performance evaluations using simulation tools and/or due to

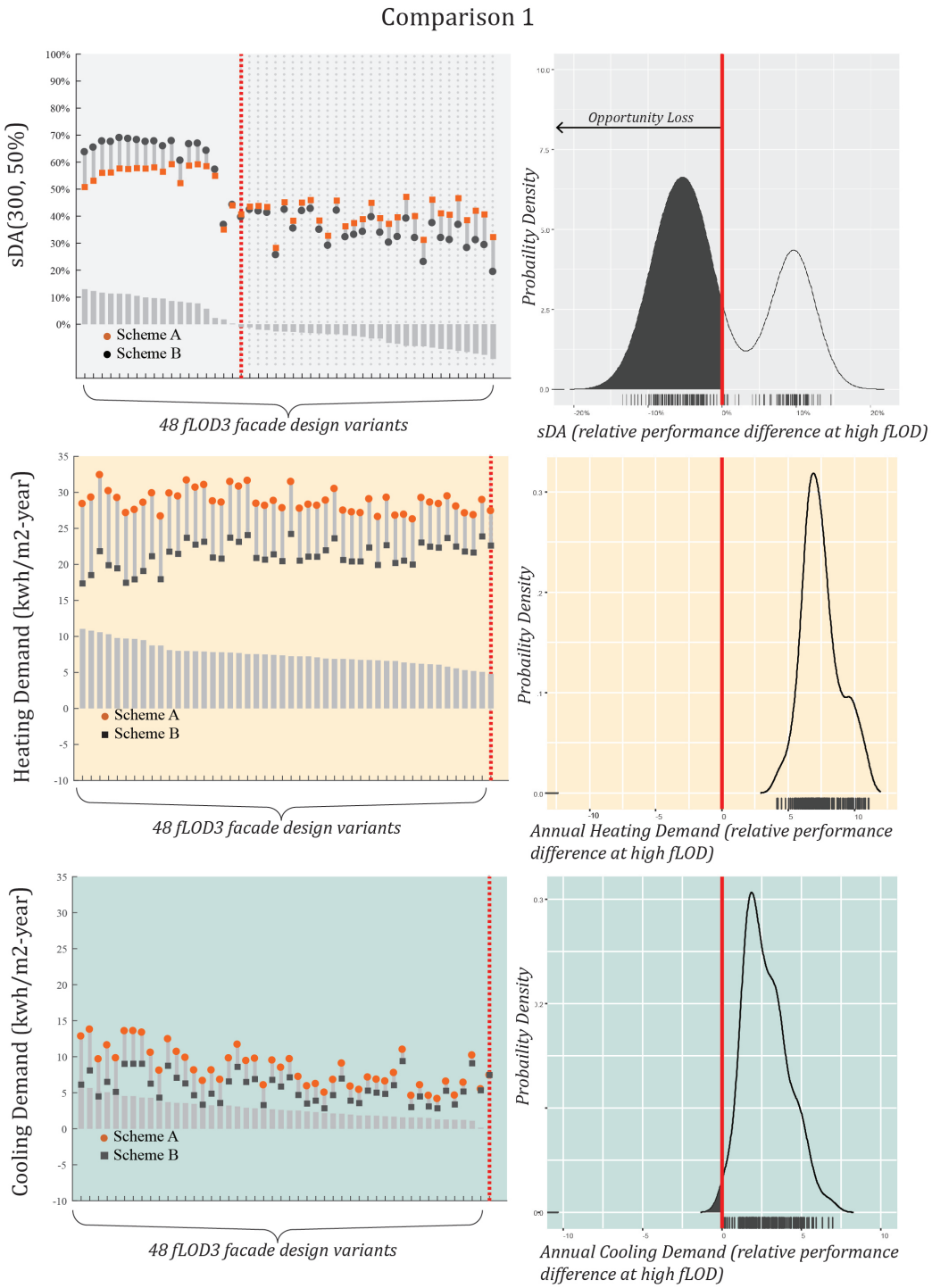


Figure 3.20 – (Left Column) Relative performance comparison for massing-scheme A,B at fLOD3 shown when strict one-to-one pairing is done for the 48 design variants at FLOD3. The comparisons to the right of the dotted line reflect opportunity loss due to rank reversal.(Right Column) Probability distribution curve derived from relative performance difference values at fLDO3. The shaded region under the curve indicates risk of performance loss.

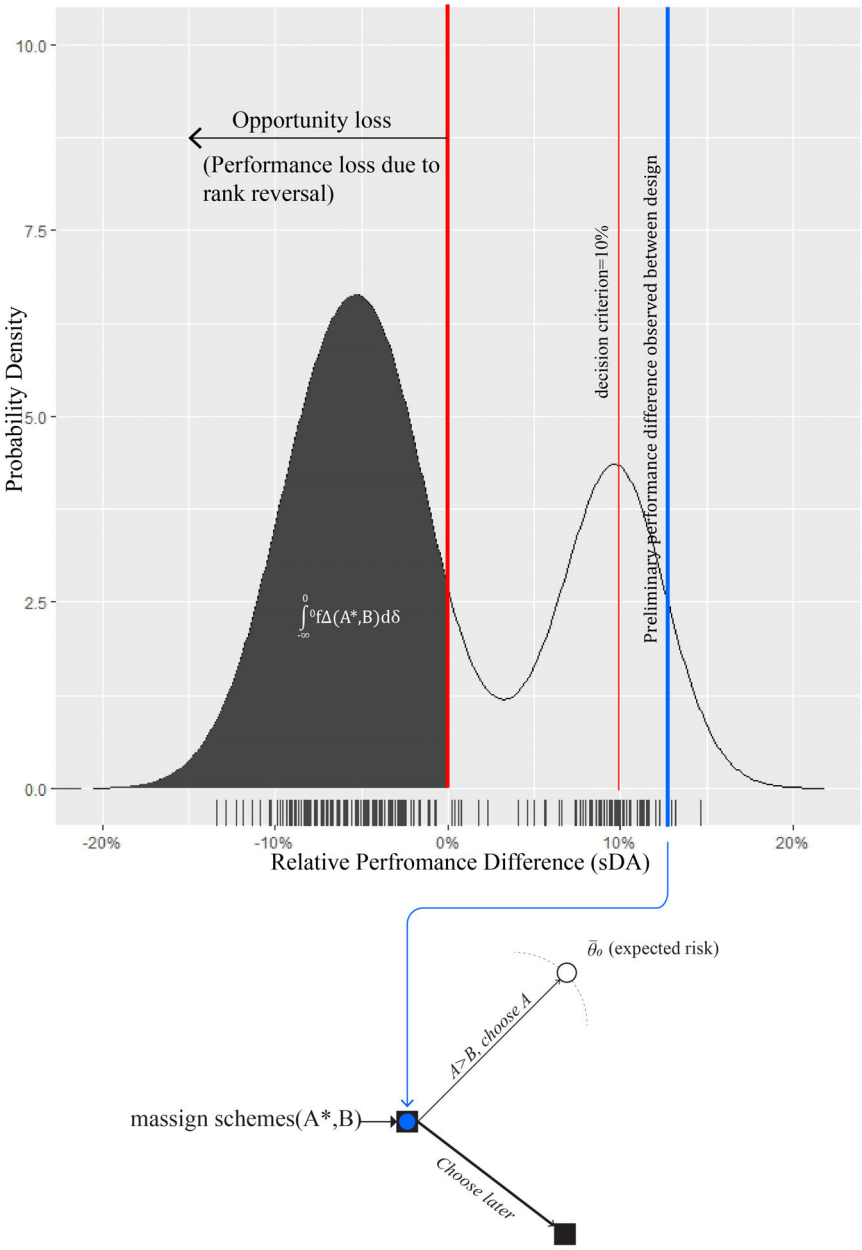


Figure 3.21 – Example distribution of relative performance where opportunity loss is observed due to rank reversal.

Table 3.7 – Possible values for decision threshold (dt) for various performance metrics.

Metric	Minimum performance differentiation needed for assigning ranks
sDA (daylight)	10% (sDA units) improvement in sDA is generally considered to be a minimum appreciable difference between design alternatives [Iversen et al., 2013]
Annual Heating Demand	2.8 kWh/m2-year (reduction in annual heating demand from code compliant (SIA 2016) to MINERGIE rating [Association MINERGIE, 2015]
Annual Cooling Demand	3.6 kWh/m2-year (commensurate active cooling energy requirement to advance from comfort Category III to Category II (EN15252) (See Appendix A.2 for detailed description on calculation of cooling demand based decision threshold)

uncertainty in design, construction and usage. Compagnon et. al. [2015] raised the issue of appreciable performance difference in a formal manner. In order to establish a threshold of meaningful performance difference, they suggested basing it on difference in performance that would be perceptible to the building occupant (e.g. indoor illuminance).

Given the multiple ways for establishing thresholds for performance difference for ruling in favor of one design alternative over another, this analysis parameters could be specific to the DM (based on his/her design goals/rating system being followed). This issue is not investigated further in this thesis. A survey of BPS users could be done to gain a better understanding of performance factors that would lead the DM to rule in favor of one design alternative over another especially at the early design stage. In absence of further knowledge on such thresholds used by DMs at the early design stage, an interim approach followed in this thesis to rely on defined performance levels built into building performance rating systems. The DM could treat the performance difference between various levels in building standards as relevant criteria for ruling in favor of one design alternative over another. This performance difference based threshold for making an early design choice is referred to as 'dt'. It is indicated in the example in figure 3.21 as a vertical red line marker. In this case, the DM considered a performance difference of 10% as sufficient to favor A^*, B. In case of heating demand evaluations, a DM could assign ranks in case the difference between two schemes is equal to the difference between a code compliant building and one that qualifies for an energy efficient building label such as MINERGIE [Association MINERGIE, 2015]. For cooling demand based comparisons, the energy input that would be required for reducing indoor temperature (during the cooling period) enough to comply with a more stringent comfort criteria (i.e. EN15251 category I instead of EN15251 category II) is suggested a significant performance improvement. The cooling energy demand improvement shown by a design alternative indicates its ability to bring about a notable improvement in indoor thermal comfort compared to the other design alternative.

Opportunity loss due to latent performance gain

Rank reversal is the most serious error in decision making where, due to insufficient level of detail, the findings at conceptual design stage are overturned later on in the design process. However, other forms of loss are also worth considering. For instance a DM may choose not to assign any ranks to the design alternatives if an appreciable performance difference (dt) is not observed between them at fLOD0. The decision maker would be indifferent and could choose either, in which case (Ai*, Bi) and (Ai, Bi*) would be considered equally likely to occur.

When the DM decides to not assign ranks to the design options at the conceptual stage (fLOD0) due to insufficient difference in performance, loss may still be incurred if the DM is failing to identify a better performing design solution due to low level of design information available at the time and at high fLOD one of the schemes delivers significantly higher performance (difference in performance at later design stage > dt). Under insufficient performance gain between (Ai,Bi) at fLOD0, the DM considers both (Ai*, Bi) and (Bi*, Ai) to be equivalent, and there is only a 50% chance that he/she will choose the massing-scheme that yields higher performance later on. This form of loss more than anticipated gains can be achieved later but would remain hidden or latent unless the performance comparison is done at a later design stage. Opportunity loss in this case could be expressed as follows:

$$EOL(A,B) = -(0.5 * \int_{dt}^{\infty} f\Delta(A^*,B)d\delta + 0.5 * \int_{dt}^{\infty} f\Delta(A,B^*)d\delta) \tag{3.2}$$

where EOL (A, B) is the EOL when A or B could be chosen with equal probability after a conceptual stage performance evaluation. $f\Delta(A^*,B)$, $f\Delta(A,B^*)$ are the PDFs probability of the relative performance gain from design pairs formulated later in the design process.

The latency effect could be interpreted as a condition where one design alternative delivers more that expected performance gain over the other. Isolating such conditions from the distribution of possible relative performance outcomes reveals loss when design alternatives that appear to be equivalent at an initial assessment. However later on in the design process, one design alternative is found to exhibit superior potential to deliver high performance.

An example of latency in performance gain is shown in figure 3.22. Like earlier, the blue vertical line indicates the performance difference observed between the two design alternatives at fLOD0 on the x-axis. At fLOD0, the performance difference between two design alternatives A,B is found to be insignificant (less than a predefined decision threshold). The combined PDF of $f\Delta(A^*,B)$ and $f\Delta(A,B^*)$ shows that (A,B^*) is likely to result in positive outcomes in most of the scenarios. However there is only a 50% chance that the DM will be able to choose the correct design alternative given that he/she is treating them as equivalent (no ranks assigned)

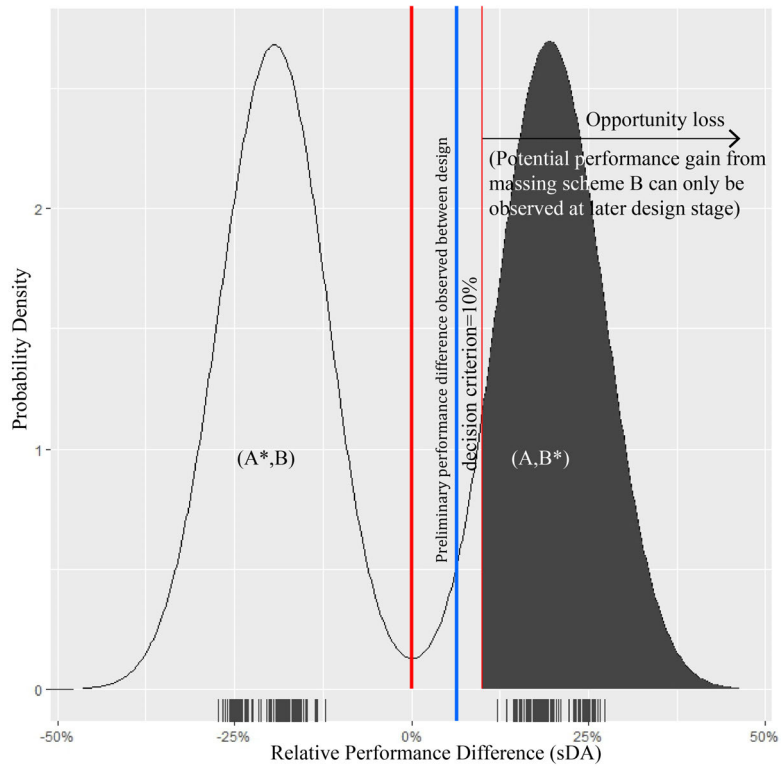


Figure 3.22 – Example distribution of relative performance where opportunity loss is observed due to latency in performance gain.

Opportunity loss due to insufficient performance gain

This form of loss does not indicate a wrong choice from the point of performance outcome of the design process, but one where the anticipated or expected performance gain, on the basis of which a preferred massing-scheme is identified (A*, B) is not achieved or realized. This is also called the lower partial moment (LPM) and is used as a risk measure when less than desired performance is achieved [Bawa and Lindenberg, 1977]:

$$EOL(dt, RPG, (A^*, B)) = -\int_{dt}^{\infty} (dt - RPG) f(RPG) f\delta \tag{3.3}$$

where $EOL(dt, RPG, (A^*, B)$ is the EOL when 'dt' is the amount of the relative performance gain that is desired, and RPG is the distribution of the relative performance gains when massing-scheme A is chosen as the preferred scheme. The distribution of relative performance at high fLOD in such a case could look like figure 3.23.

Expected Relative Performance Loss (ERPL): Depending on the outcome of this comparison, the EOL is calculated based on all applicable equations out of 3.1, 3.2, 3.3. The total loss present the sum total risk from all possible forms of EOL. **This joint value of risk is referred to as the expected relative performance loss (ERPL).**

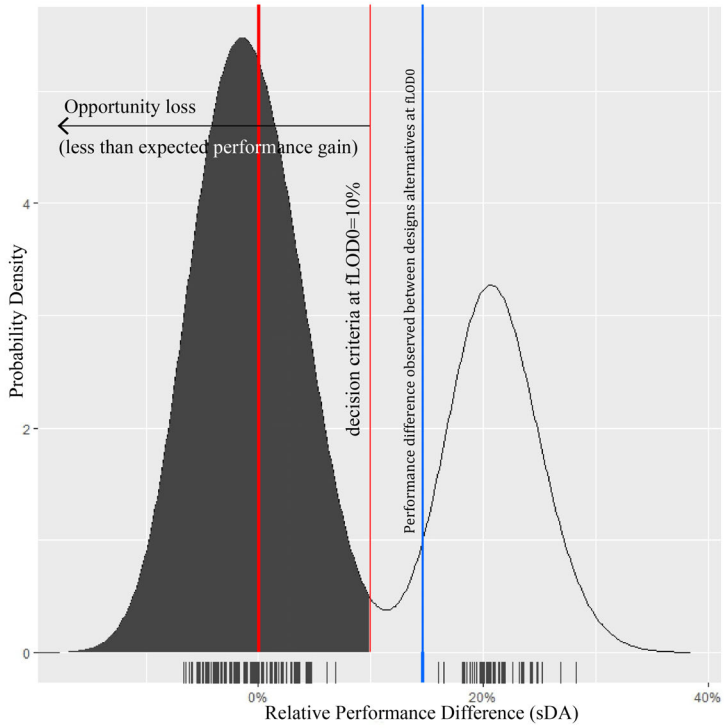


Figure 3.23 – Example distribution of relative performance at high LOD where opportunity loss is observed due to insufficient performance gain.

3.3 Analytical framework for the results

In order to understand the design conditions that induce risk, this thesis examines various aspects of the massing-scheme pairs that were found to be associated high risk. Various properties of the massing schemes that are known to have effect on one or more performance metrics evaluated in this work. In cases where detailed thermal and daylight simulations fail to provide reliable relative performance evaluations at fLOD0, it was tested whether certain properties of the massing scheme could be a contributing factor. The manifestation of scenarios at fLOD3 on a particular massing scheme is linked to its geometrical properties. This provides incentive to investigate relationship between the geometry of massing-schemes and the risk involved when their performance is compared at fLOD0.

The selected properties and the method for calculating these properties are described below:

- **Surface Area to Volume Ratio:** (Exterior surface area divided by enclosed volume) has been shown to be strong indicator of heating demand in buildings in cold climate [Pessenlehner and Mahdavi, 2003]. It is the total building surface area (including roof area) divided by the enclosed volume.
- **Mean Building Height:** (Building height weighted by area). Average height and site

coverage, at fixed density, are often found to be inversely related to each other. These factors were studied together by Chatzipoulka et al. [2016]; Cheng et al. [2006] and height was found to positively related to amount of irradiation received on facades. This variable indicates the height in terms of number of floors weighted by the floors area at each floor level.

- **Site Coverage Ratio:** Site coverage area has been shown to be inversely related to irradiation received on the building facades[Cheng et al., 2006] and indoor daylight access (sDA) [Nault et al., 2015]. Site coverage is the total building footprint area over site area.

- **Number of Buildings:** This refers to the count of number of buildings in a massing-scheme. This variable has been typically included in studies at the urban scale [Mohajeri et al., 2016] and is used to measure the fragmentation in the built mass. Given the significant variation in the subject set of massing schemes in the experiment (3-8 buildings) this factor has been included.

- **Passive Zone Ratio:** A passive zone is the part of the built space that can potentially be heated, cooled and or daylit in a passive manner [Baker and Steemers, 2003]. As a rule of the thumb, a perimeter zone is twice as deep as the floor-floor height and is called the passive zone. Passive zone ratio is the ratio of the passive area to the total building floor area. Ratti et al. [2005] found passive-zone-ratio more closely related to heating demand than surface to volume ratio for several Euorpean locations using the LT method [Baker and Steemers, 1996]. [11] Other studies show weaker relationship of passive-zone-ratio to passive heating and daylight related metrics [Nault et al., 2015]. Nevertheless, passive zone ratio is easy to understand and is often used a rule of thumb in the design process and has thus been retained for the analysis.

- **Complexity:** This refers to the total facade area over site area. Complexity is thus related to surface-volume-ratio and has been shown to positively correlated to annual irradiation on facades after controlling for density [Chatzipoulka et al., 2016].

- **Mean Outdoor Distance:** This variable measures the mean distance of open un-built space to the built masses. It is meant to indicate the mean distance between buildings. This factor was calculated by laying a 1m x 1m grid of points and calculating the mean distance of each point on the grid to the built masses nearest to it. Figure 3.24 shows the graphical output of the MOD calculation done in grasshopper [Chatzipoulka et al., 2016].

- **Directionality:** This indicator is the standard deviation of ground permeability in 36 directions weighted by site coverage. This is calculated by drawing radial, vertical sectional planes (at 10 degree intervals) through the the built mass of each massing scheme. The area of the built mass intersecting each plane is calculated and the standard deviation in these areas is reported as directionality. Buildings blocks placed along

[11] The LT method estimates the energy use of given urban form based on simulations carried out on a pre-defined 'test' room.

different orientations will have low directionality, while regularly spaced linear buildings (bars) all aligned in the same direction will have high directionality. This variable has been studied in multiple climatic contexts and increasing directionality has been found beneficial for irradiation received on building facades. [Chatzipoulka et al., 2016; Cheng et al., 2006]

- **Mean annual irradiation on vertical surfaces per unit area of vertical facade:** So far, many of the factors considered, have been found to be associated with high solar irradiation on facades. Total annual irradiation was calculated as well. It may appear that this factor may make several geometry related factors redundant, however all factors, including this one, offer partial information regarding the shape of the built mass and its impact on performance. The average irradiation per unit of facade area does not reveal equity in distribution of the irradiation on all surfaces but tells the overall level of irradiation received.

- **Mean annual irradiation on vertical surfaces per unit area of floor area:** This factor normalizes the irradiation available on the facades per unit of built floor area. Since the floor area of all the schemes in the experiment is the same, this factor allows for comparison of total irradiation received by various massing schemes.

- **Passive solar heating potential:** This and the following two factors assess the fraction of the building facade that is able to receive a certain threshold of solar irradiation (or daylight). It is anticipated that when the thresholds are met, indoor comfort requirements can largely be met by passive means. Irradiation thresholds are based on Compagnon [2004]. Since there is some degree of subjectivity involved in the calculation of the thresholds, three (low,medium,high) values were as follows 208 kwh/m2,178 kwh/m2,156 kwh/m2. A grid of points with a density of 1m x 1m was laid on the exterior surface of each massing scheme. Total annual irradiation received at each point was calculated and tested for compliance on each threshold. The fractions of grid points that meet the threshold is the fraction facade area that could be considered fit for a specific passive design strategy. Figure 3.25 shows the calculation of this indicator on a massing-scheme, where the orange colored points are the points found to be compliant with the threshold of 178 kWh/m2 (annual). The next two indicators are calculated in a similar manner.

- **Overheating avoidance potential:** Cooling period irradiation thresholds (201 kwh/m2,254 kwh/m2,301 kwh/m2) are based on Nault [2016]. These thresholds are meant to indicate need for overheating avoidance and the percent area of the facade is calculated using the same method as mentioned for passive solar heating potential.

- **Daylight potential** For this factor, threshold values of 5 klux,10 klux ,15 klux are used based on Compagnon [2004] to once again assess the fraction of the facade area that can be considered suitable for daylighting the adjoining interior spaces.

Table 5 in appendix A.6 shows a summary of the factors described above and range and mean values for the subject massing- schemes in the experiment.

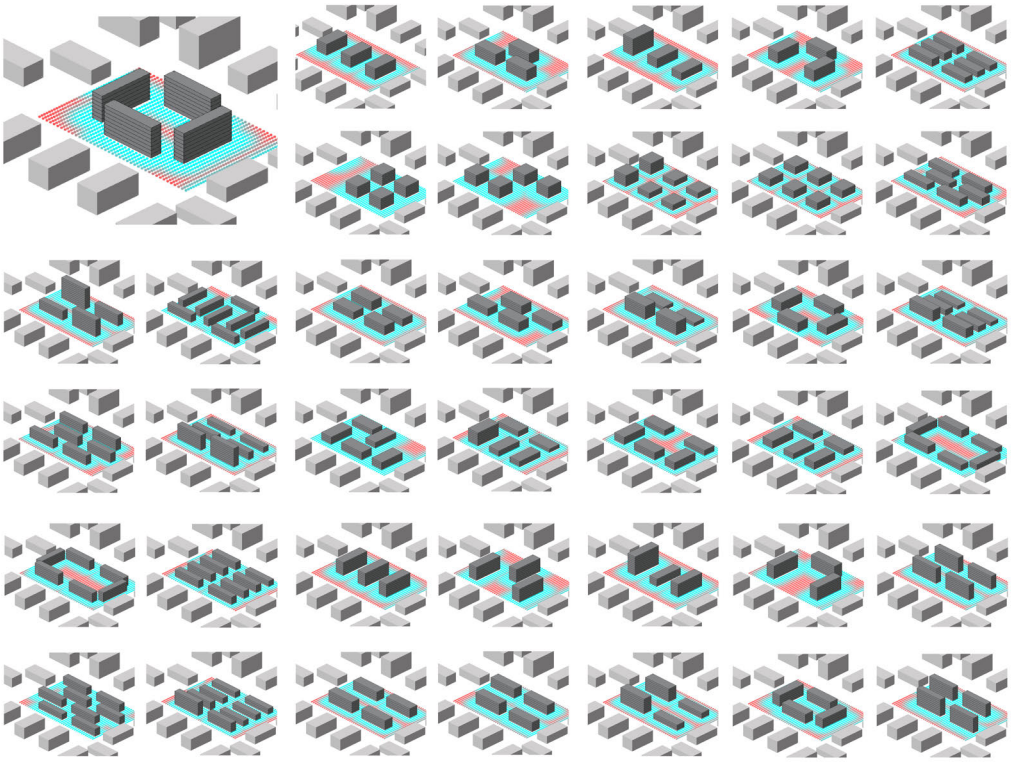


Figure 3.24 – Graphical output of calculation of mean outdoor distance of all massing schemes used in this experiment. Blue color indicates points close to built mass and pink points are farthest from built mass. The analysis is limited to the site boundary.

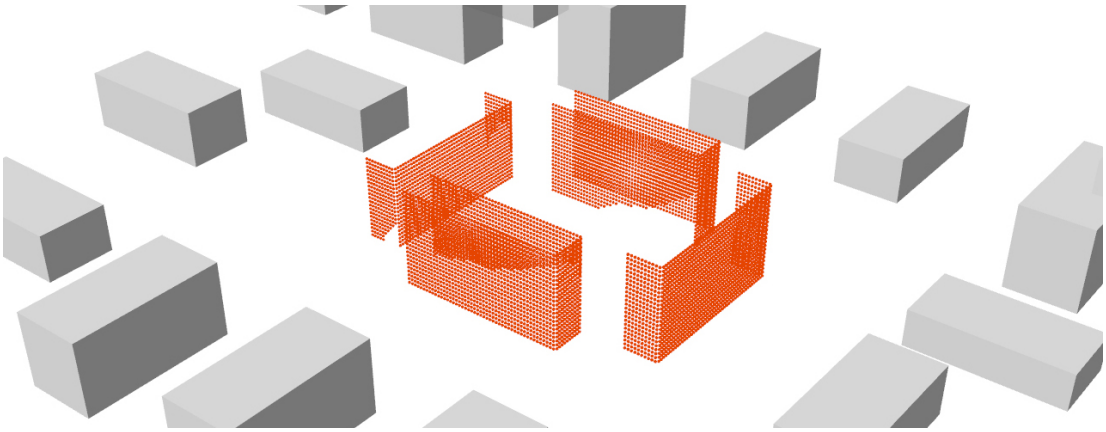


Figure 3.25 – Percent of facade area complying with irradiation threshold, example output from a massing-scheme.

The analysis strategy was to first visually examine the data using bivariate plots and also slicing the data set in various ways. Regression analysis would allow the examination of multiple factors simultaneously and also assess the degree to which the present of risk can be explained using the various massing-scheme related properties extracted in preparation of this analysis.

To identify geometrical/solar access properties of massing-schemes that are most closely related to the observed risk, a multi-factor analysis of variance (ANOVA) could be conducted. ANOVA relies on a model hypothesis and shows which model factors best explain the variance observed in the risk (dependent variable). Ordinary least square (OLS) linear regression is another analysis technique that allows us to compare the explanatory power of each individual factor and also the combined ability of multiple factors to estimate the risk, in case the risk is better explained using multiple factors. Given that OLS regression allows us to understand the effect of individual model factors (geometrical/solar access properties of massing-schemes), we shall explore this analysis technique. Final analysis strategy was left open to the nature of the data (risk) to be evaluated once the data set was obtained and preliminary analysis was done.

3.4 Summary

This thesis proposes ERPL metric for evaluating risk in conceptual design decisions. This chapter presents the methodology for calculating ERPL along with an experimental framework for estimating the prevalence of risk than can be expected in conceptual stage decision making. The chapter has three main elements 1) generation of 780 pairs of competing design proposals (neighborhood-scale massing schemes) from a set of 40 massing schemes (40C2=780) 2) generating future facade design scenarios for each of the massing schemes in the test set 3) development of the risk assessment method. Important features all these elements are as follows:

Facade design scenarios: The risk assessment relies on the generation of numerous future facade design scenarios (N=48) for each given conceptual stage design proposal. A grasshopper (parametric 3-D modeling tool) based workflow was developed for this purpose. The same workflow was found capable of applying the desired facade details on all massing schemes that were used. This was achieved using simple rules and trial and error mechanisms built into grasshopper scripts. For example, distribution of glazing on various facades while respecting constraints regarding maximum and minimum permissible WWR per facade required a trial and error sequence in the code to make sure all constraints are met. Further sophistication in code were not needed to generate the desired facade variants. Facade design variants were also generated in their intermediate states (fLDO1-no blinds, fLOD1 (with blinds), fLDO2). The ability of these intermediate fLODS to reduce risk is presented in Chapter 5.

Massing-scheme proposals: These were developed manually based on data drawn from recently built Swiss residential buildings to be used for experimental study of risk in conceptual stage decision-making. All schemes in this set have a built density (total built area/site area)

of 1.0 and align with projects built at this density (in Switzerland) on several respects (e.g. maximum and minimum height of projects at density of 1.0). Within the observed ranges in various building dimensions, several design possibilities were created. Chapter 4 presents the results from pairwise comparisons of these massing schemes based on performance evaluations at fLOD0.

Risk Assessment method: The ERPL risk metric presented in this chapter covers three types of conditions of loss (rank reversal, latency effect and insufficient gains) and presents a combined estimate of performance loss to a DM. [12]. This risk metric builds upon the existing risk metric EOL [Su and Tung, 2012] and includes further adaptions for a sequential design process. Key additional modifications include:

1. **Structural changes to decision tree** to evaluate adequacy of design information contained in early design BPS model, rather than the performance potential of each design alternative.
2. Pair-wise comparison of future scenarios, replaced by **peer-to-peer comparisons** as one-to-one equivalence cannot be established between some design scenarios.
3. **Inclusion of loss function due to delayed discovery of performance gain**. Such a function is meaningful when using the decision-tree as mentioned in point 1 above.
4. **A method for enumerating architectural design scenarios** using parametric modeling techniques. The method introduces design details in a nested manner to control model level of detail.

The experiment described in this chapter is intended to present a generalized view of quality of decision making (indicated by risk) when only the massing design geometry is available.

Figure 3.26 shows how the three main aspects of the risk assessment (methodology for generating the subject set of massing schemes, generation of future facade design scenarios and estimating the risk) are related to the results and additional explorations presented in the subsequent chapters.

Chapter 4 presents a summary of the results using all three elements from chapter 3 (second row in figure 3.26). More specifically in this chapter, the risk that the DM takes on by relying on the relative performance evaluations obtained at fLOD0 is shown. Statistical analysis is also done on these results to show conditions under which risk is induced.

While chapter 4 presents results for a DM who exclusively makes decisions at fLOD0, in chapter 5 (illustrated in the third row in figure 3.26), the risk in decision making that would be observed

[12] ERPL in its current form is flexible enough to be used at other design stages to evaluate other other competing choices as long as there are relevant outstanding design details to be added. Its use is only illustrated at the conceptual design stage

if the DM made decisions at higher fLDOs is examined. Also results are presented for different types of DMs making decisions at fLDO0.

Figure 3.26 (last row) shows two explorations that examine the ERPL metric's potential to support decision making. ERPL includes two dimensions of loss - its likelihood of occurrence and its magnitude. A DM may also want to know if the risk is low due to a low magnitude or low likelihood of loss. In such a case, the DM wants to know the distribution profile of loss. The ERPL metric in the form presented in this chapter does not support such an informational need. The question arises whether it is sufficient for a DM to receive a binary feedback on risk. Also, is ERPL, in its current form, useful for a DM who not only wants to know the risk, but also needs to decide the future course of action? To answer such questions, two lines of inquiry were initiated. These scientific inquires try and understand the potential of ERPL in providing decision making support. First, the degree of information beyond ERPL value that will be useful to expose to the DM is evaluated. This endeavor is structured as an opportunity to deliberate on how the metric, along with some additional information, could potentially help a DM understand the future implications of decisions made at the current fLOD. Second, ERPL is compared to other robustness measures to bring out potential advantages and disadvantages in a case study based decision making environment. These two topics are covered in Chapter 6.

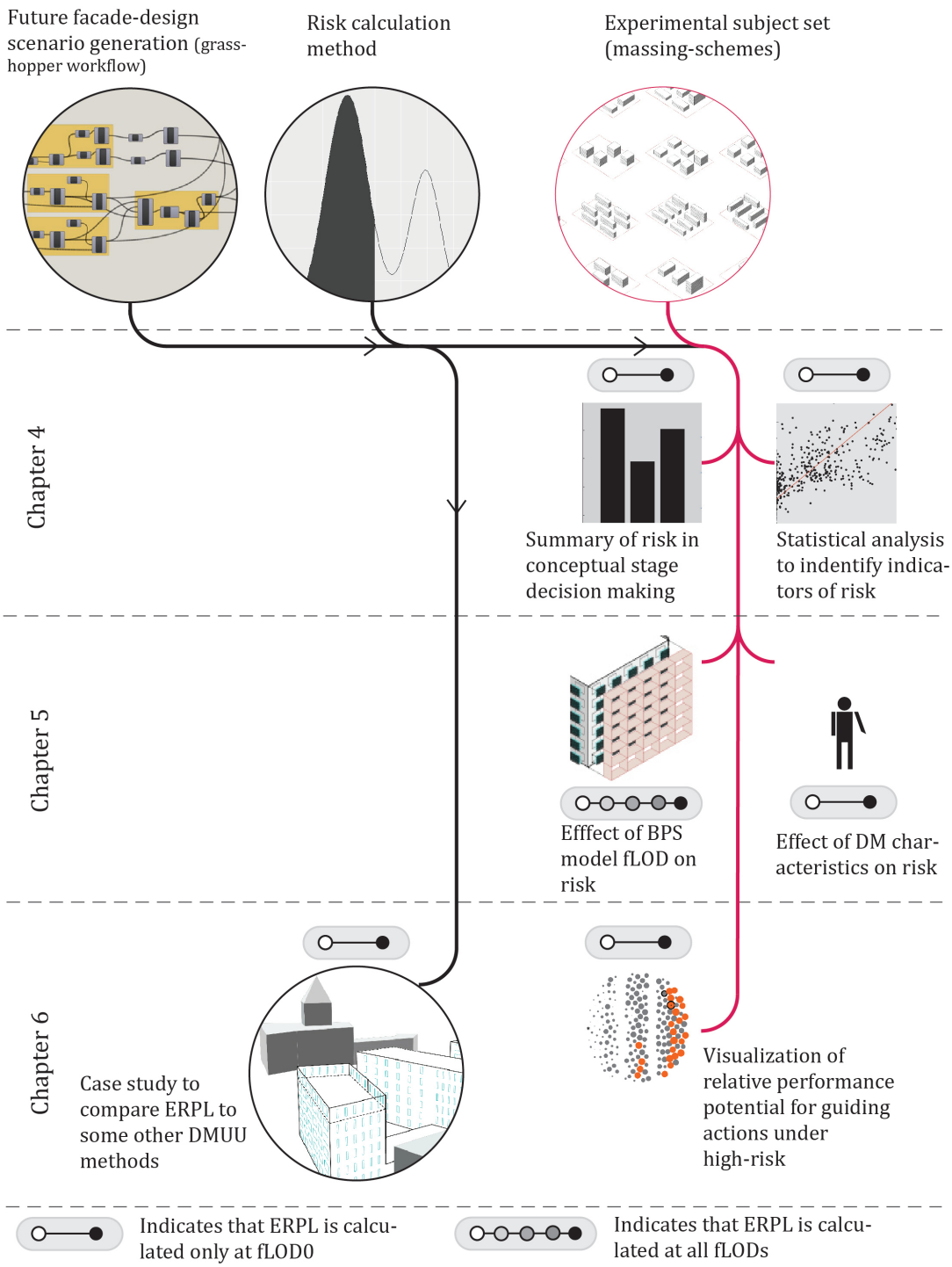


Figure 3.26 – Summary of explorations done in the thesis with the risk assessment method presented in this chapter.

4 Risk of Performance Loss in Conceptual Stage Design Decisions

This chapter presents the risk undertaken by a conceptual stage decision maker, who as a decision making practice relies on BPS evaluations obtained at fLOD0 to choose between two massing-scheme. Figure 4.1 indicates the scope of this chapter where two massing schemes are compared at a time (full set of schemes presented in section 3.2.1). When the DM is confronted with the decision to choose at low fLOD (fLOD0) or later (at higher fLOD), the DM in each case chooses the path highlighted in red (i.e. to choose right away). The risk that such a DM will encounter is presented here. Given the number of comparisons that were conducted in this manner (N=780), the results indicate the potential for risk in decision making that can occur at the conceptual design stage when adopting the simplest decision making process (within the scope of this thesis). The variation of risk with higher fLODs is discussed in the following chapter.

The performance evaluation received at fLOD0 and the resulting decision that a DM would make is compared to the performance evaluations that could be seen at the highest fLOD. fLOD3 (highest modeled fLOD in the experiment) proxies for the *'actual'* design. In other words, the highest modeled fLOD is treated as the experimental ground truth for assessing reliability of decision making at fLOD0.

The choice at fLOD0 is made separately on each of the three metrics (sDA, annual heating demand, annual cooling demand) and multi-criteria decision making is not addressed in this chapter. The DM ranks the massing schemes pairs based on one performance metric at a time and the risk observed in decision making based on each metric is reported separately. This allows for comparison in reliability of decision making achieved on different metrics at fLOD0. Multi-criteria decision is an important aspect of building design shall be addressed in chapter 6.

Also, in the results presented in this chapter, the DM is characterized by certain assumptions and is referred to as the baseline DM. Three characteristics for such a DM have been assumed: a) minimum performance differentiation threshold b) maximum acceptable loss and c) maximum acceptable chance of unacceptable loss.

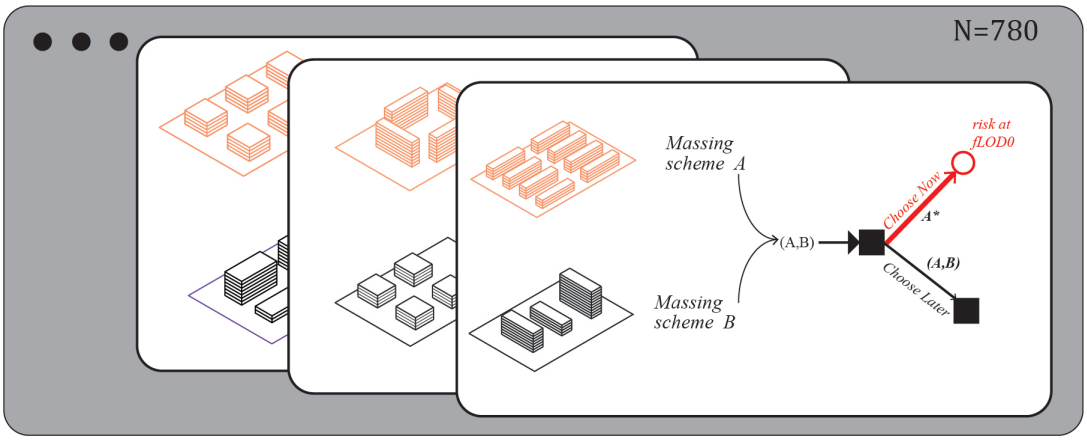


Figure 4.1 – Experiment to assess reliability of decision making at fLOD0 at neighborhood scale. This chapter focuses on results where the DM follows the approach indicated in red (choosing between design alternatives while ignoring the uncertainty in facade design details).

4.1 Observed risk in conceptual stage decisions

Several facade design scenarios at the highest fLOD (N=48) for each massing scheme design alternative in a pair-wise comparison result in several possible relative performance evaluations (N=192). These relative performance evaluation values were shown for an example comparison in section 3.2.3 along with the subsequent calculation of risk for the pair of massing schemes. The process is repeated between all possible pairs of massing schemes (N=780) from the pool of subject schemes. For illustration purposes, figure 4.2 shows the probability distribution of relative performance values at high fLOD for 30 such comparisons (out of a total of 780 comparisons that were done). Relative performance value is shown on the x-axis. Area under the curve to the left of the zero marker indicates possible opportunity loss. The occurrence of opportunity loss also depends on the decision that would have been made at fLOD0 in each case. In this section the results from all pairwise comparisons of massing schemes are presented in a collection manner. These results address one of the chief goals of this thesis which was to estimate the reliability in decision making that can be achieved at the conceptual design stage. While subject to the experimental limitations, these results indicate the potential reliability in decision making that can be achieved using conceptual stage BPS models.

The average risk in 780 sDA based evaluations at fLOD0 was found to be 1.12%. It was found to be 0.25 kWh/m2-year and 0.20 kWh/m2-year in heating and cooling demand evaluations respectively. These values appear to be quite small and insignificant (for example, in comparison to the decision threshold (dt) values). However these represent the average risk from all of instances of decision making (N=780) and further within each comparison an estimation of loss from 192 possible relative performance outcomes. These values are thus expected to be small, as no loss is found in a large number of cases. It is more relevant to examine the distribution of observed risk values which shows the extent of risk that could be found based on the comparisons that were done.

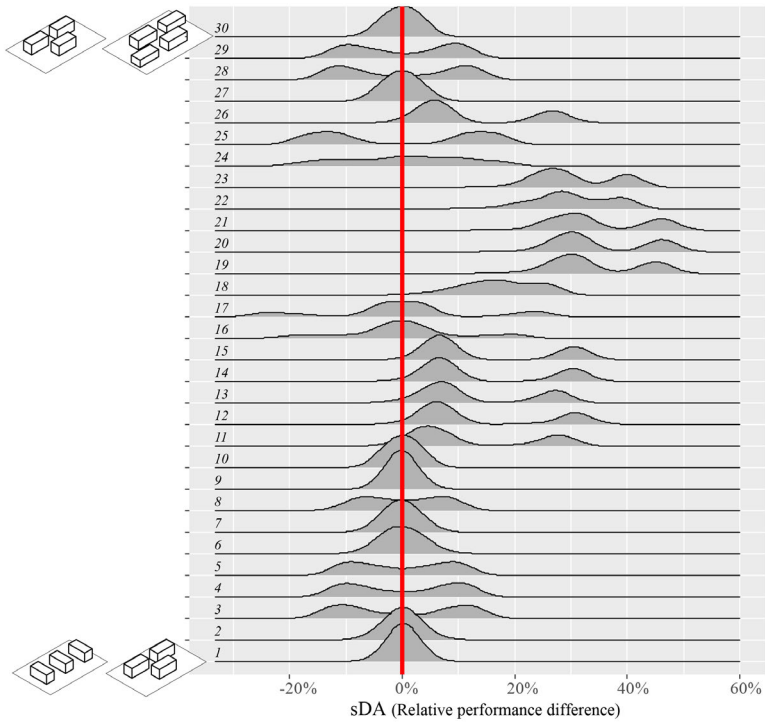


Figure 4.2 – Distribution (shown as PDF) of relative performance values (A-B) at high fLOD (fLOD3) 30 neighborhood comparisons. Area under the curve to the left of the solid vertical line indicates possible Opportunity Loss.*

Figures 4.3, 4.4 (top and bottom panel) show the distribution of risk(y-axis) by magnitude of the risk (x-axis) for each metric from all the comparisons done in the experiment. The source of performance loss is indicated by color. The highest risk value was of the order of 9.5% points on sDA and was cause by the latency effect. Rank reversals cause an risk of upto 6.5% points on sDA. Risk due to insufficient performance gains is also found in sDA assessments. In daylight evaluations the biggest contribution to risk comes from the latency effect. That is, in such cases the DM would have dismissed the performance difference observed at fLOD0 (< 10% on sDA), but in several such comparisons, one of the massing-schemes is found to deliver consistently higher performance later in the design process.

In heating demand evaluations, in almost all cases where any loss (and as a result risk) occurs, it is due to the latency effect. That is, loss occurs mostly in cases when at fLOD0 the performance difference between two massing-schemes appears to be negligible or insignificant (<2.6 kWh/m2-year on heating demand) but at high fLOD one scheme consistently delivers higher performance (>2.6 kWh/m2-year performance improvement over the competing scheme). **In heating demand evaluations, it may not be sound practice to disregard small performance differences seen in fLOD0 evaluations**. Higher fLODs may reveal whether the performance difference between design alternatives is indeed negligible. However, absence of other forms

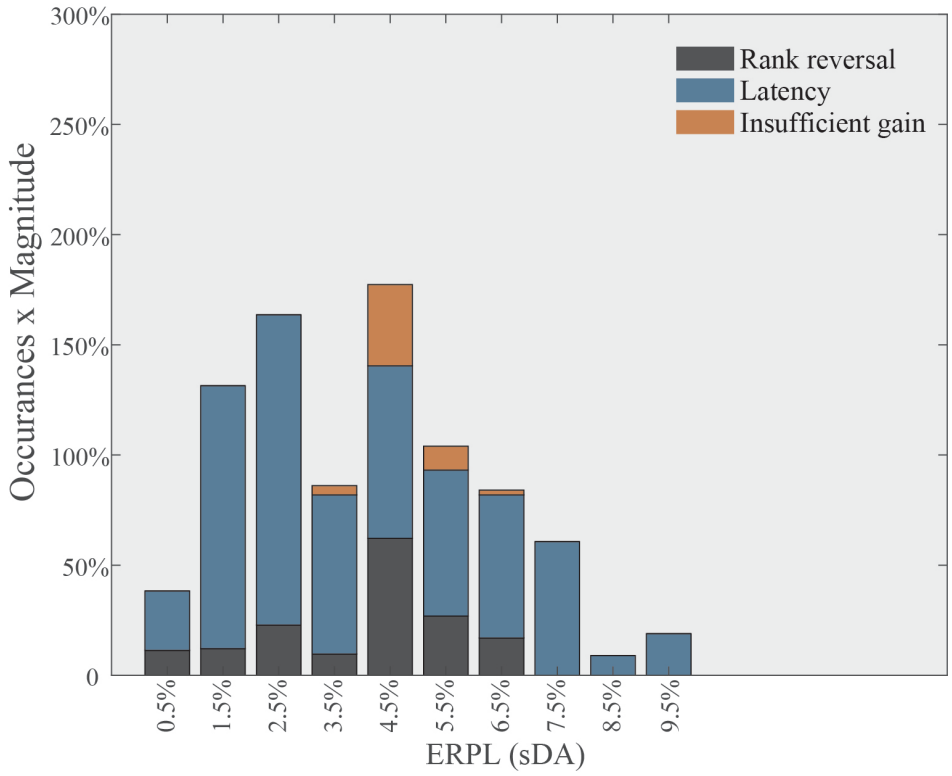


Figure 4.3 – Risk(ERPL) of performance loss in sDA assessments at fLOD0 observed in 780 comparisons. Nature of loss indicated by color.

of loss (rank reversal and insufficient performance gain) in heating demand evaluations could be a result of the nature of the experimental set up. Heating demand (in the chosen climatic context of the experiment) is known to be driven by conduction based losses heat through the building envelope and ventilation needs. The configuration of the envelope from the point of conduction related losses remains unchanged through the fLODs (only facade scenarios with the same WWR are compared). If other methods of specifying window were used (e.g. window to floor area ratio), then other forms of losses could be seen. The ventilation needs are held constant and understandably, the same across all fLOD3 design scenarios.

Similar to the sDA metric, on cooling demand as well 4.4 (bottom panel), all three sources of loss contribute. In cooling based evaluations the highest risk values (the right-most bars in figure 4.4 bottom panel) are from the rank-reversals and insufficient performance gain. At the same time, latency effect though present, is not important to cooling demand evaluations given the lower incidence of high risk cases (8.3%) and the low contribution of latency effect to high risk (left-most bars in figure 4.4 (bottom panel)). It may be noted that the risk observed in the heating and cooling demand evaluations in the given context (Residential buildings in climate of Geneva, Switzerland) are further influenced by the prevailing building code requirements. If the building code thermal insulation requirements for opaque and glazed assemblies were higher, the performance values would be more sensitive to the facade design factors considered in this thesis and vice-versa.

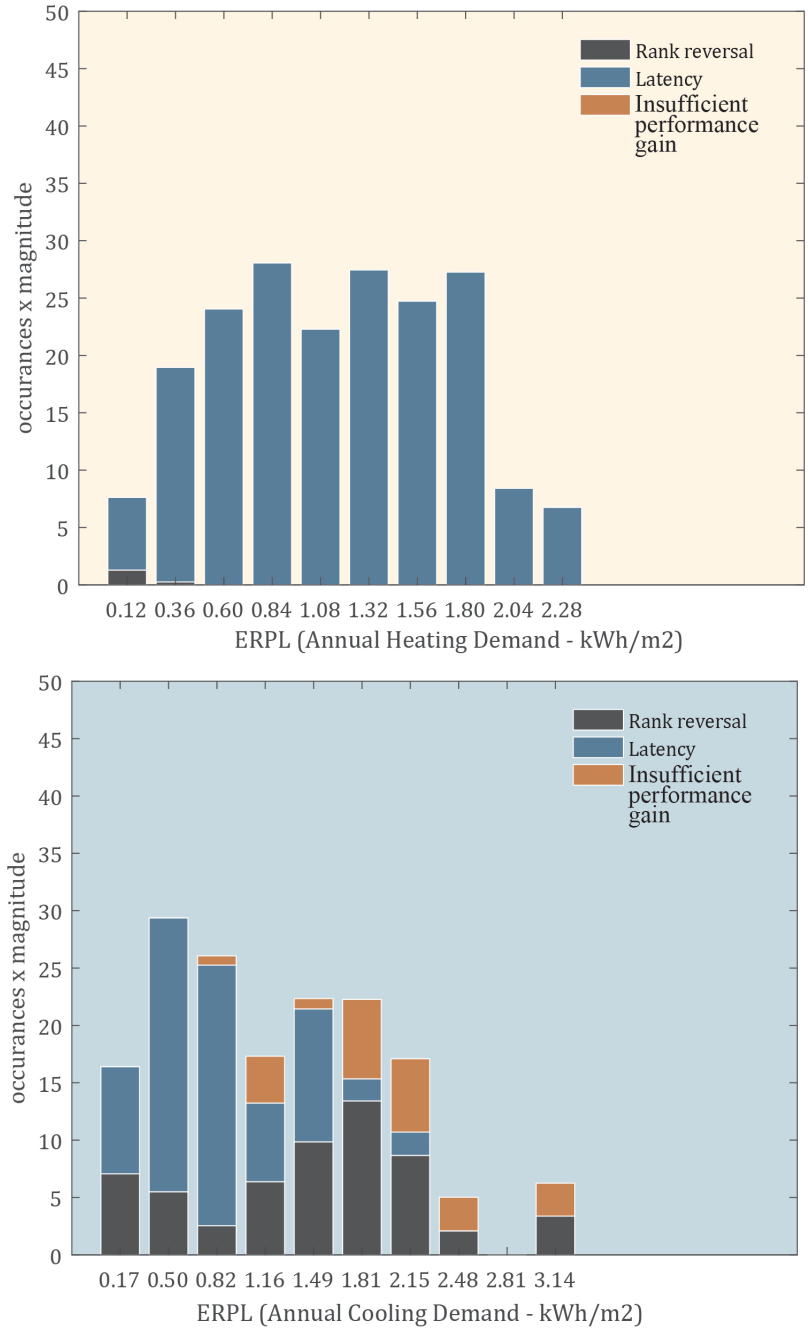


Figure 4.4 – Risk(ERPL) of performance loss in annual heating demand assessments at fLOD0 observed in 780 comparisons. Nature of loss indicated by color.

4.2 Relevance of observed risk to decision makers

Risk presents the loss that the DM can expect to incur when making a decision under uncertainty. Whether the DM actually ends in a state of loss or not is only revealed when the uncertainties are eliminated. So how should the DM act when uncertainties are still present and risk is found? Should the DM respond to all risk values? This section addresses such questions and re-presents the results shown in the previous section so that they can be distilled into inferences for DMs and policy makers concerned with robust decision making at the conceptual design stage.

4.2.1 Threshold for 'high' risk

The ERPL risk metric combines various sources of risk relevant to a sequential decision maker. However risk is probabilistic in nature and high values of risk do not imply that an incorrect decision will be made. Similarly, low risk does not guarantee a loss free decision. Thus the challenge in understanding implication of a certain risk value is that it represents a multitude of future possibilities. While ERPL provides an estimated loss value, the DM does not know the probability of following a particular design path (leading to a particular high fLOD scenario). The risk on various design paths varies but only the average risk from all paths can be known as the DM does not know which path he/she is going to follow.

Consider two DMs, DM1 and DM2 shown in figure 4.5(a) and (b) respectively. DM1 while deciding between two design alternatives A1 and B1 encounters risk. The risk is emanating from regret/opportunity loss being present in some fLOD3 scenarios, all of which branch from the same node (same fLOD2 scenario). If the decision maker was to ignore the risk and later result on the node 'n' then he/she runs a very high chance (three out of four) of ending in regret. While there are only three scenarios out of 48 modeled scenarios that lead to regret, the concentration of the regret on one set of scenarios is a potential threat. Now consider DM2 who is comparing a different pair of design alternatives A2 and B2 and also encounters risk from regret in three fLOD3 scenarios but the regretful scenarios are dispersed. If DM2 ends up on node 'n' of his/her tree of possible future facade decisions, he/she still has a fair chance of avoiding regret.

And yet DM1 and DM2 do not know if they will end up on the node 'n' of their respective decision trees. The only reliable evaluation they have of the situation is in the form of ERPL which weighs the regret by the probability of experiencing it. To meaningfully operate with such as decision tree, the 'normal form' of analysis Raiffa [1968] is followed. This method is usually employed when specific probabilities of downstream branches (future design paths) are unknown. In such a situation, a strategy e(s) is devised that would suit the needs of the decision maker. A simple rule/strategy is defined that sets a risk threshold value θ_0 for the risk at fLOD0 such that the possibility of going down on an extremely adverse future design path can be averted. The proposed strategy sets a risk threshold value for θ_0 such that the risk under every future design scenario at the penultimate stage before the final design scenario is

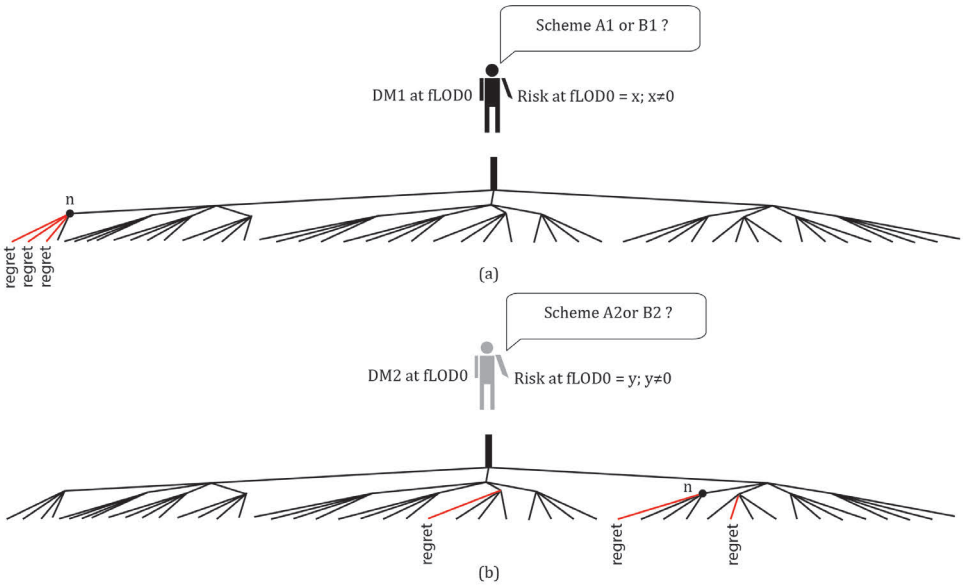


Figure 4.5 – Figure showing two DMs (DM1 and DM2) who encounter risk while making their respective conceptual stage decisions but the scenarios with regret at distributed differently.

arrived at, is less than a 50% chance of incurring an unacceptable value of performance loss. Table 4.1 shows the assumptions used in this thesis regarding loss that would be unacceptable to the DM and the maximum acceptable likelihood of incurring such a loss. It is assumed that the DM is not interested in eliminating loss completely (in that case, maximum acceptable loss would zero), but rather lower it to a safe limit. On sDA, as an example, the safe limit is assumed to be a 50% chance of being incorrect by a margin of 10% on sDA under every single scenario at fLOD2, the penultimate fLOD before fLOD3. This translates into a maximum permissible risk of 5% at fLOD2. Every fLDO2 scenario leads to 16 fLOD3 scenarios. A risk of 5% on sDA, as an example, could result from a loss of 10% on 8 (out of 16) scenarios beyond fLOD2. At fLOD2, a scenario that has a higher risk than 5% is called 'high' risk as the DM who arrives at this scenario has more than 50% chance of making the incorrect choice.

Based on this strategy, DM1 (figure 4.5(a)) would be called out as being in a high risk condition while DM2 (figure 4.5(b)) could be evaluated as being in a low risk condition.

This strategy was applied to all 780 instances of risk assessment resulting from each of the massing scheme comparisons. A common threshold that worked to identify most high-risk cases was identified from the experiment results using statistical analysis. This risk methodology for identifying the risk threshold is further illustrated in Appendix A5 in a simplified manner. The methodology was extended (beyond what is illustrated Appendix A5) to establish a risk threshold at fLOD0 based on risk observed at fLOD2 (penultimate fLOD). The resulting risk thresholds are shown in table 4.2). If the risk at fLOD0, when comparing two massing schemes exceeds these thresholds, then the DM must reconsider his/her relying on performance evaluations obtained at fLOD0 and take some remedial action like increasing the fLOD.

Table 4.1 – Assumptions related to maximum acceptable loss and chance of incurring such a loss.

Performance metric	Maximum acceptable loss	Maximum acceptable probability of unacceptable loss
Daylight (sDA)	10% (percent points on sDA)	50%
Annual Heating Demand	2.8 kWh/m2-year	50%
Annual Cooling Demand	3.6 kWh/m2-year	50%

Table 4.2 – High risk threshold for various performance metrics.

Daylight (sDA)	Annual Heating Demand	Annual Cooling Demand
2.1%	0.7 kWh/m2-year	0.9 kWh/m2-year

4.2.2 Impact of risk at the conceptual design stage

To assimilate the risk observed in the pair-wise comparisons is presented below using a series of risk indices. These indices present a second-order view of risk in decision making at fLOD0. While ERPL estimates risk in a single instance of decision making, these indices presented here show the overall chances of encountering high risk of loss at fLOD0 in the experiment containing 780 comparisons between massing schemes pairs. These indices address the dual dimension of risk, i.e., 1) the chance of an undesired outcome 2) the resulting loss from the undesired outcome. Some of these indices also rely on the risk threshold for a DM at fLOD0 (table 4.2) that used to identify instances of high risk. In such cases, ignoring the risk value would result in unreliable decision making.

1. **Relative error:** This is the risk per conceptual stage design decision (made at fLOD0) divided by average performance gain anticipated when the decision was made. It is a unit-less index (expressed in %) [Cabannes et al., 2018]. The index can be used to evaluate the risk relative to the gain that the DM hopes to achieve by conducting performance evaluation at fLOD0.

2. **Average risk per high risk case:** This refers to the average risk in all conceptual stage design decisions (fLOD0) where the risk was found to be 'high'. In low risk cases, while opportunity loss does occur, it is low enough to be ignored. Its unit is same as the index on which the risk is being evaluated (e.g. % sDA, kWh/m2-year annual heating demand)

3. **Percentage of unreliable decisions:** This refers to the number of high risk cases that can be expected per 100 instances of early design decision making.

4. **Maximum observed risk:** This refers to the highest risk that was found in the experiment per metric. This indicates the extent of risk that could be encountered in decision making at fLOD0.

5. **Maximum observed Loss:** This refers to the highest loss that was observed in the experiment, per performance metric. ERPL is the average loss that a DM could expect. It accounts for the possibility that under certain facade design scenarios there is no loss. Here the highest loss that was observed in the experiment, that a DM could potentially suffer is reported.

Figure 4.6 shows results of the experiment on the five indices mentioned above. The relative errors were found to be between 32% and 20% (Figure 4.6 (a)). This statistic depicts the potential benefit from preparing simulation models at fLOD0 while accounting for the downside of making the decision too early and the resulting risk of performance loss.

Figure 4.6 (b, c) shows the overall chance of making an unreliable decision ('high' risk) and the average amount of risk in such a decision at fLOD0. 22% cases on sDA are found to be high risk. Thus, 1 in 5 comparisons on sDA are high risk and each unreliable decision involves an average risk of 4% (percent points of sDA). Part (d) of the figure shows that compared to the average risk of 4%, the maximum risk was found to be 9.52%. Figure 4.6 (e) shows the maximum loss that was observed in the experiment. The loss values shown in this figure indicate the worst case result from decision making at fLOD0 (For example, 33% on sDA).

15.1% cases on heating demand were high risk (Figure 4.6 (c)). On heating demand, a DM can expect to take a high risk decision 1 out of 7 evaluations. The risk in such a decision is expected to be 1.25 kWh/m2.year in annual heating demand. The maximum risk was found to be 2.5 kwh/m2-year.

On cooling demand, high risk decisions could be made one 8.3% of the time, or 1 out of 12 times. The risk in such a decision is expected to be 1.54 kWh/m2-year in annual cooling demand. The maximum observed risk was 3.23 kWh/m2-year. On all performance metrics the maximum risk is roughly two times the average risk of all high risk cases.

For a conceptual stage decision maker, average risk per high risk case and incidence rate of high risk case are more relevant. Average risk could be relevant to a policy maker who needs to understand the overall implications of risk in decision making at the conceptual stage decision making.

4.3 Relation of risk to observed performance difference at fLOD0

Since the possible source of loss (rank reversal, latency effect and insufficient performance gain) depends on the difference found between a pair of schemes at fLOD0, the experiment results were divided into sub-groups based on the performance difference observed at fLOD0.

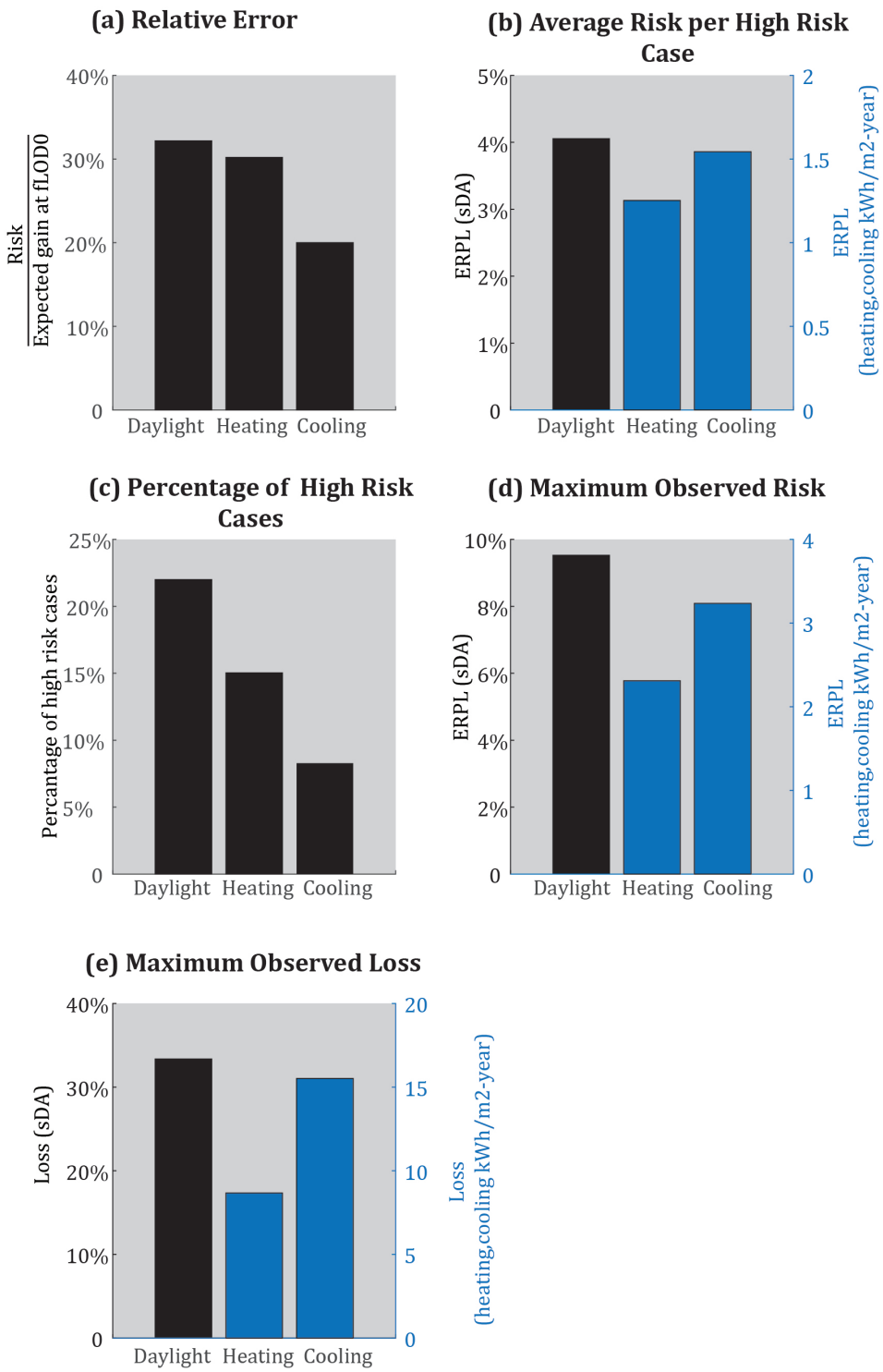


Figure 4.6 – Summary of results from the experiment; 780 comparisons at fLOD0.

Figures 4.7 (a),(b) - 4.9 (a),(b) show the data from the experiment divided into three sub-groups based on low (difference at fLOD0 <dt), moderate (2 x dt > difference at fLOD0 > dt) and high difference (difference at fLOD0 > 2 x dt). Part (a) of these figures shows the percent break up of high risk (risk > risk threshold), low risk (risk is not zero but < risk threshold) and finally no risk cases where the risk is equal to zero.

In daylight evaluations (see figure 4.7 (a), (b)), high risk cases were found irrespective of performance difference observed fLOD0. However, the percentage share and number of high risk cases drops significantly as the observed performance difference between two massing schemes grows. At a difference of less than 10% in sDA at fLOD0, which also a commonly held rule-of-thumb for minimum differentiation between design proposals on sDA [Iversen et al., 2013], the incidence of high-risk cases is very high (37% cases are high risk). This experiment thus supports the common notion that less than 10% difference in sDA at fLOD0 is unreliable, although this conclusion is drawn here on the basis of uncertainty in detailed design features. A significant number of high risk cases are found at higher differences in sDA as well. At difference of 10%-20% in sDA at fLOD0, 17% cases are still found to be high risk and thus performance difference values in this range do not guarantee risk free decisions. However, the number of high-risk cases diminishes significantly as the performance difference values approach 20%. In case of heating demand evaluations Figure 4.8 (a) and Figure 4.4 show that latency effect is the predominant source of risk in heating demand evaluations. Figure 4.8 (a) and (b) show that all high risk cases are limited to latency effect that occurs when the performance difference observed at fLOD0 is low (less than dt i.e. <2.6 kWh/m2-year). When the performance difference at fLOD0 is less than the decision threshold (<dt), there is more than 25% chance of latency effect. In evaluations based on cooling demand (see figure 4.9 (a), (b)), high risk cases were found to be more evenly distributed (in terms of percentage share) across the three subsets based on performance difference seen at fLOD0. In others words, high risk cases are found under low, moderate and high performance difference at fLOD0. Occurrence of risk irrespective of the amount performance difference observed at fLOD0 (small, moderate or high) suggests that risk cannot be ruled out or anticipated for given comparison. It needs to be estimated on a case by case basis if the performance difference at fLOD0 is less than 2 x dt or 6.9 kWh/m2-year.

In part (b) of figures (4.7 - 4.9) latency effect related high-risk cases are highlighted (performance difference at fLOD0 < dt) which were further analyzed. Under the latency effect, the performance difference between the two schemes grows as fLOD is increased. In such cases, a high performance scheme cannot be identified at fLOD0 as the performance difference is too low to make a definitive choice. However high risk of this nature is not found in all cases of such comparisons. Out of the sDA evaluations cases that could possibly suffer from latency effect, 36.2% were found to have high risk (Figure 4.7 (a)). This ratio was 26.3% for annual heating demand and 6.2% for cooling demand (4.8 (a), 4.9(a)). The number of risky pairs with latency effect on cooling demand is 25 (out of 780). In case of heating demand and sDA this number was 138 and 127 respectively.

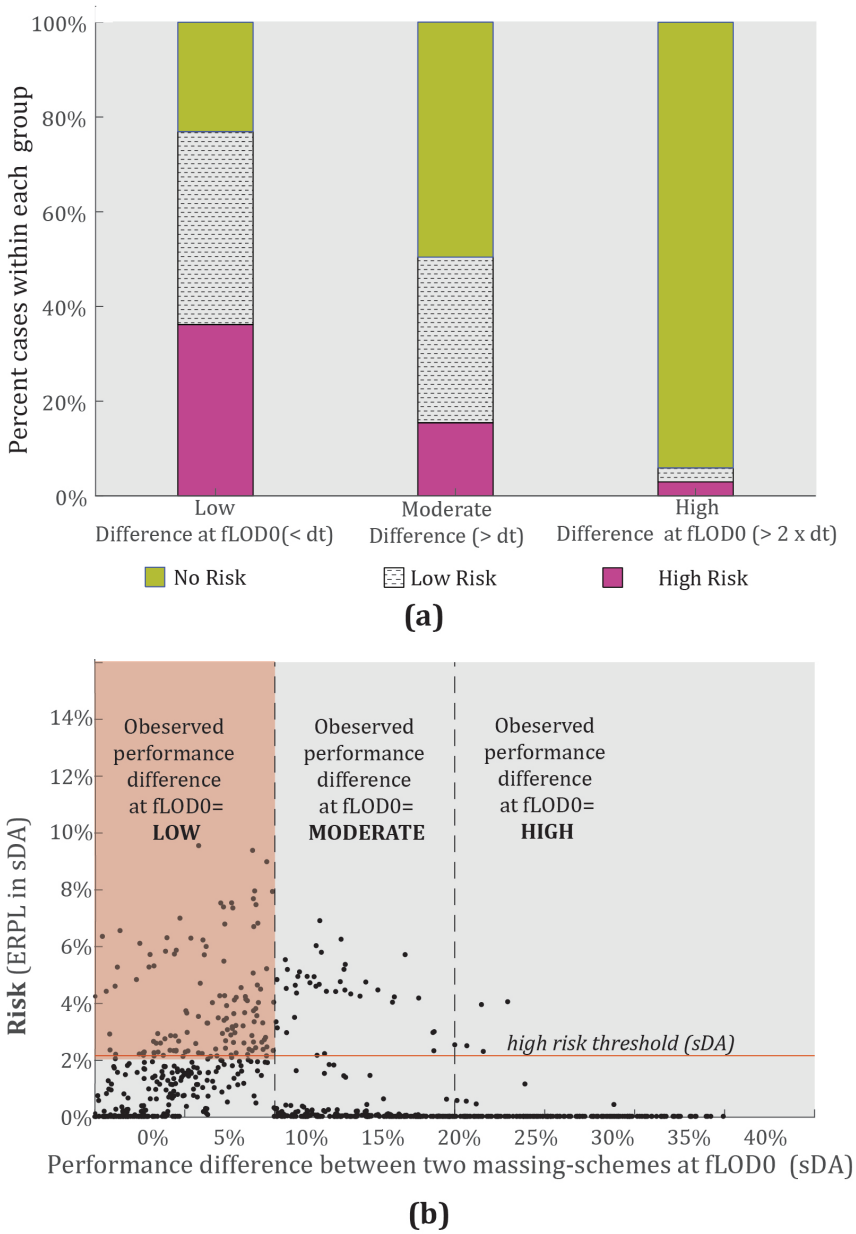


Figure 4.7 – Panel (a) shows percentage of risky cases by performance difference observed at fLOD0 on sDA. Panel (b) shows the underlying data behind the percent break-up shown in (a); data in highlighted region is further analyzed in the subsequent section.

Figure 4.10 shows this investigation for the sDA metric. Four geometry and two irradiation based indicators from table 5 were selected as a preliminary step. Building geometry indicators (Surface volume ratio (SVR), Passive Zone Ratio (PZR), Facade Complexity(CEX)), one site level geometry indicator (Site Coverage(SC)) and two irradiation based indicators (Total irradiation on exterior vertical facade surfaces per unit of indoor floor area (TIFA) and exterior surface irradiation/illuminance threshold based potential for passive performance) were calculated

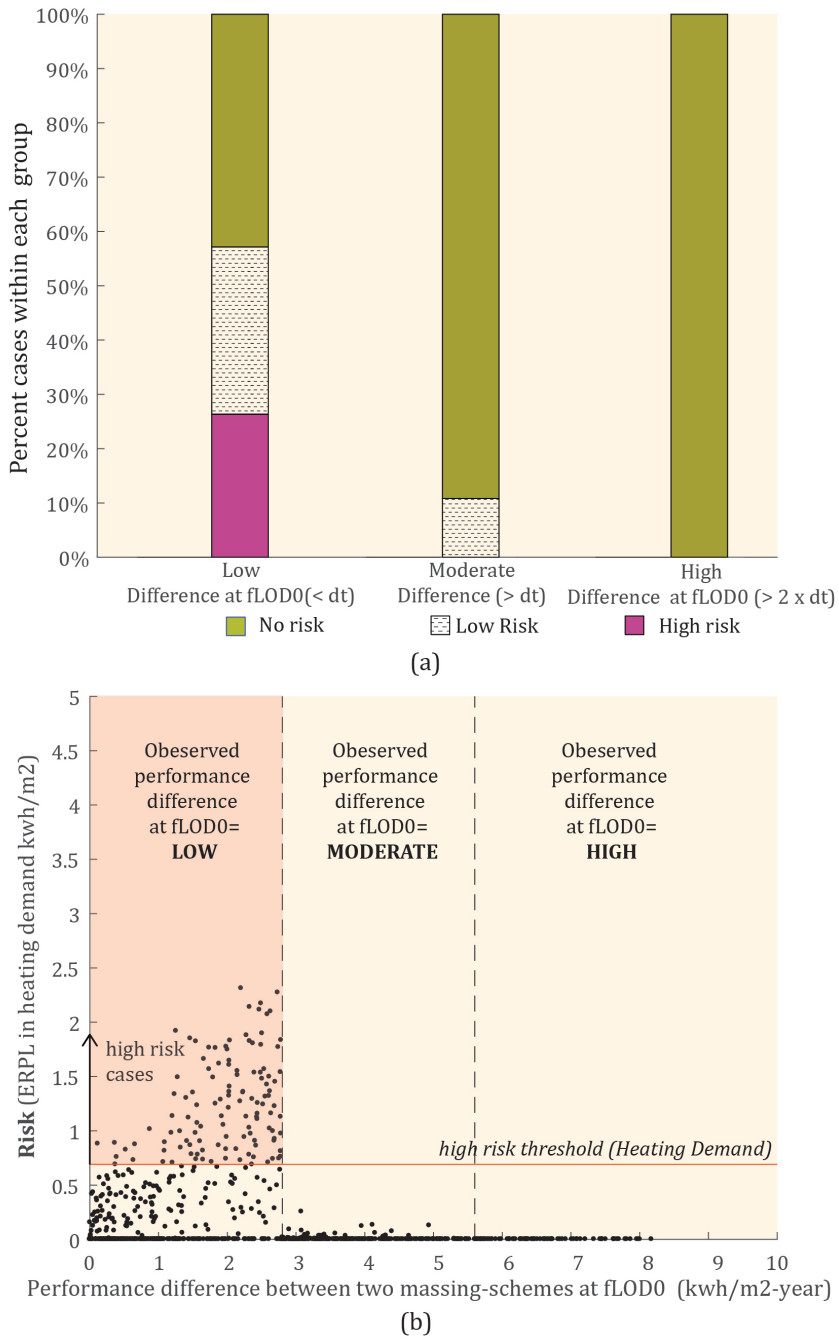


Figure 4.8 – Panel (a) shows percentage of risky cases by performance difference observed at fLOD0 on annual heating demand. Panel (b) shows the underlying data behind the percent break-up shown in (a); data in highlighted region is further analyzed in the subsequent section.

for the massing-scheme pairs suffering from latency effect. These short listed indicators are considered particularly important to one or more performance metrics for which risk was calculated (see section 3.3 for references to existing studies that highlight their importance).

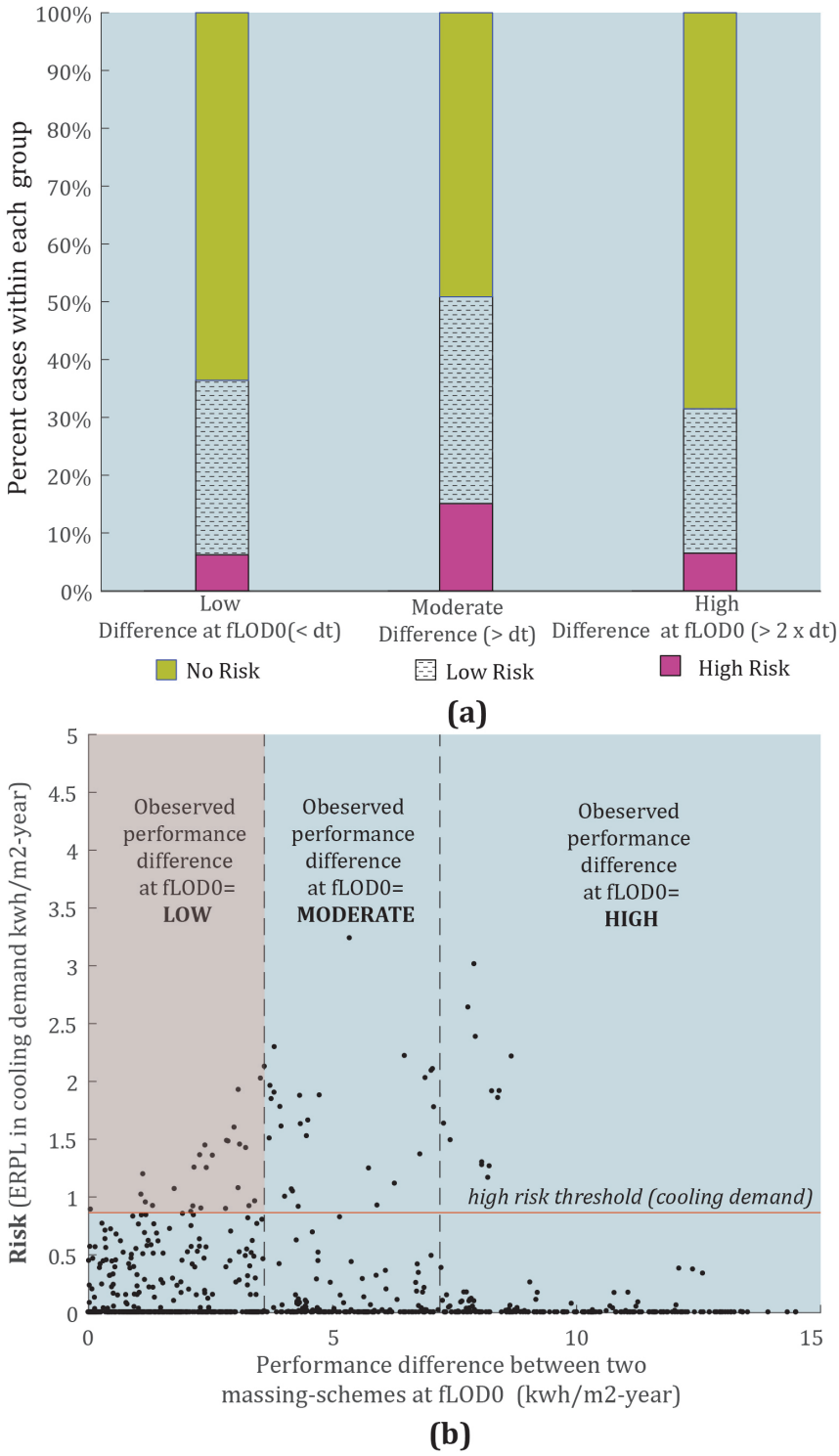


Figure 4.9 – Panel (a) shows percentage of risky cases by performance difference observed at fLOD0 on annual cooling demand. Panel (b) shows the underlying data behind the percent break-up shown in (a); data in highlighted region is further analyzed in the subsequent section.

Box plot in Figure 4.10 shows the differences in properties of the high risk and the no-risk cases. All pairs of box plots between which a significant difference in not found, are marked "n.s" (using a two tailed T-Test, 5% significance level). Significant differences are seen between high and no risk cases on surface volume ratio(SVR), passive zone ratio(PZR), facade complexity(CEX), site coverage(SC) and total annual irradiation received on facades per unit of floor area (TIFA). It was also seen that the difference in these properties on high risk was greater than no risk cases. That is, in all the no-risk cases, not only was the difference in simulated performance (sDA) low (at fLOD0), but, differences between such pairs of schemes were also small on other indicators such as SVR, PZR, CEX and TIFA. On passive daylight potential (P-Daylight) such a clear difference was not seen between high and no- risk cases.

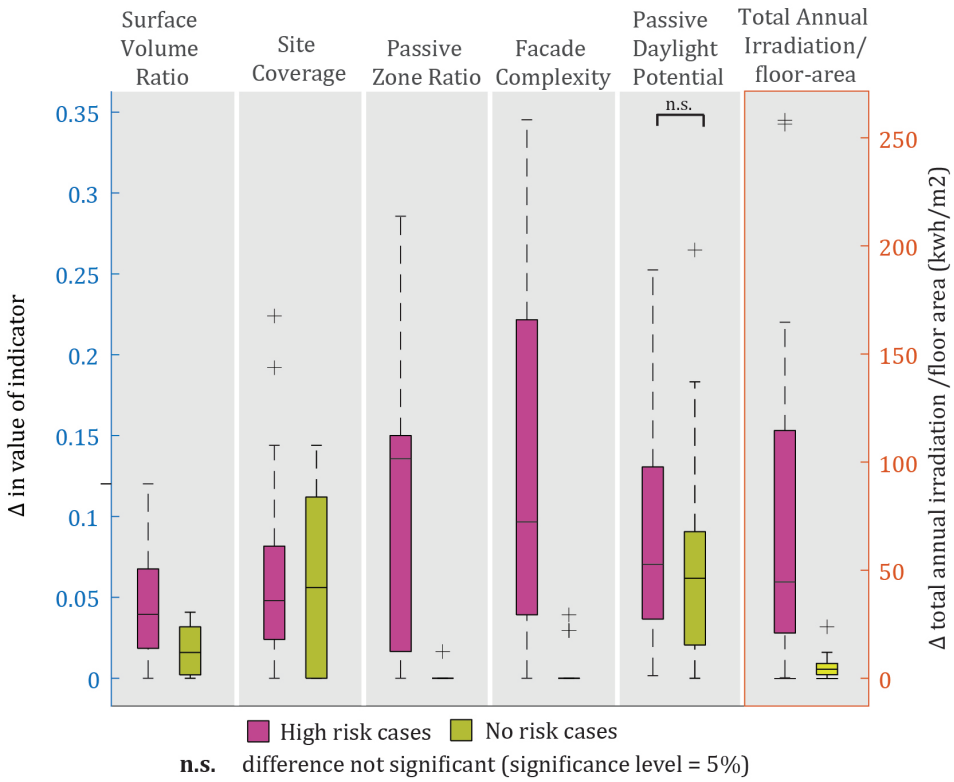


Figure 4.10 – Comparison of geometrical properties and other early design performance indicators of high risk and no risk cases in sDA based evaluations. Cases shown are pairs of massing schemes in which latency effect could potentially be found (see highlighted region in figure 4.7). Pink box plots indicate subset where the risk was high. Green boxes indicate subset where the there was no risk.

It may be further noted that SVR and CEX are geometrical properties that are directly related to the facade area of a massing-scheme. The denominator in both cases (volume and site area respectively) are the same for all the massing schemes. PZR indicates whether a building has sufficient exposed exterior surface in relation to floor area. High difference in TIFA in spite of low difference in sDA performance at fLOD0 indicates untapped potential that is not visible in simple box models with regularly spaced, uniformly distributed punched windows at fLOD0.

At high fLOD, as the window distribution varies along with different depth of balconies, that greater difference in performance is realized.

Similar trends are found in high versus no-risk cases in heating demand based risk evaluations. Figure 4.11 (top panel) also shows massing-scheme pairs that could potentially suffer from latency effect at fLOD0 in annual heating demand evaluations. The differences in the same set of indicators (SVR, SC, PZR, CEX, TIFA and P-Heat) are shown. Apart from SC and P-heat, on all other indicators, significant differences are found between high and no-risk cases. Figure 4.11 (bottom panel) shows the same analysis on cooling demand evaluations. Here differences in SVR, PZR, CEX and P-overheat were found to statistically significant between high and no-risk groups. However, in this case the difference in indicators values for the no-risk groups were high (as opposed to results for heating and daylight evaluations where the opposite was seen). Greater similarity between massing-schemes (lower difference on indicators) was found to induce high risk of latency effect. These plots however present a one-factor-at-a-time (OFAT) examination of possible relationship and do not permit a view into the co-relations that could exist between multiple geometrical factors. A more sophisticated approach is thus needed to understand the conditions under which high risk cases occur which shall be further explored in the following section.

4.4 Relationship of conceptual stage performance indicators to risk

In this section expands on exploration of conceptual stage performance indicators (e.g. SVR, PZR), as potential indicators of risk. This extended exploration is meant to provide additional means of anticipating risk. If some of the commonly used conceptual stage performance indicators are found to indicate risk as well, they could be used to inform an architect whether a risk assessment could be important. This exploration could address some questions such as - *are comparisons within a certain type(e.g. courtyard, linear bars) for massing-schemes more prone to risk?*. Statistical analysis methods can provide more robust basis drawing inferences regarding the conditions under which risk is likely to be found.

In following subsections, multiple plausible statistical (Ordinary Least Square (OLS) regression) models are presented that were tested for their ability to predict risk. Results from the model with the best fit are presented. Additional statistical modeling techniques were also tried to understand limitations found in OLS regression models.

4.4.1 Statistical modeling approach for risk prediction: linear regression

This section shall deal with the potential of statistical analysis to determine factors that could indicate risk of performance loss when comparing two given massing schemes. Linear regression is a popular statistical modeling technique. It is easy to understand the implications of a linear regression model and users of the regression model can develop a sense whether a factor is positively or negatively correlated with risk. First a set of Ordinary Least-Square

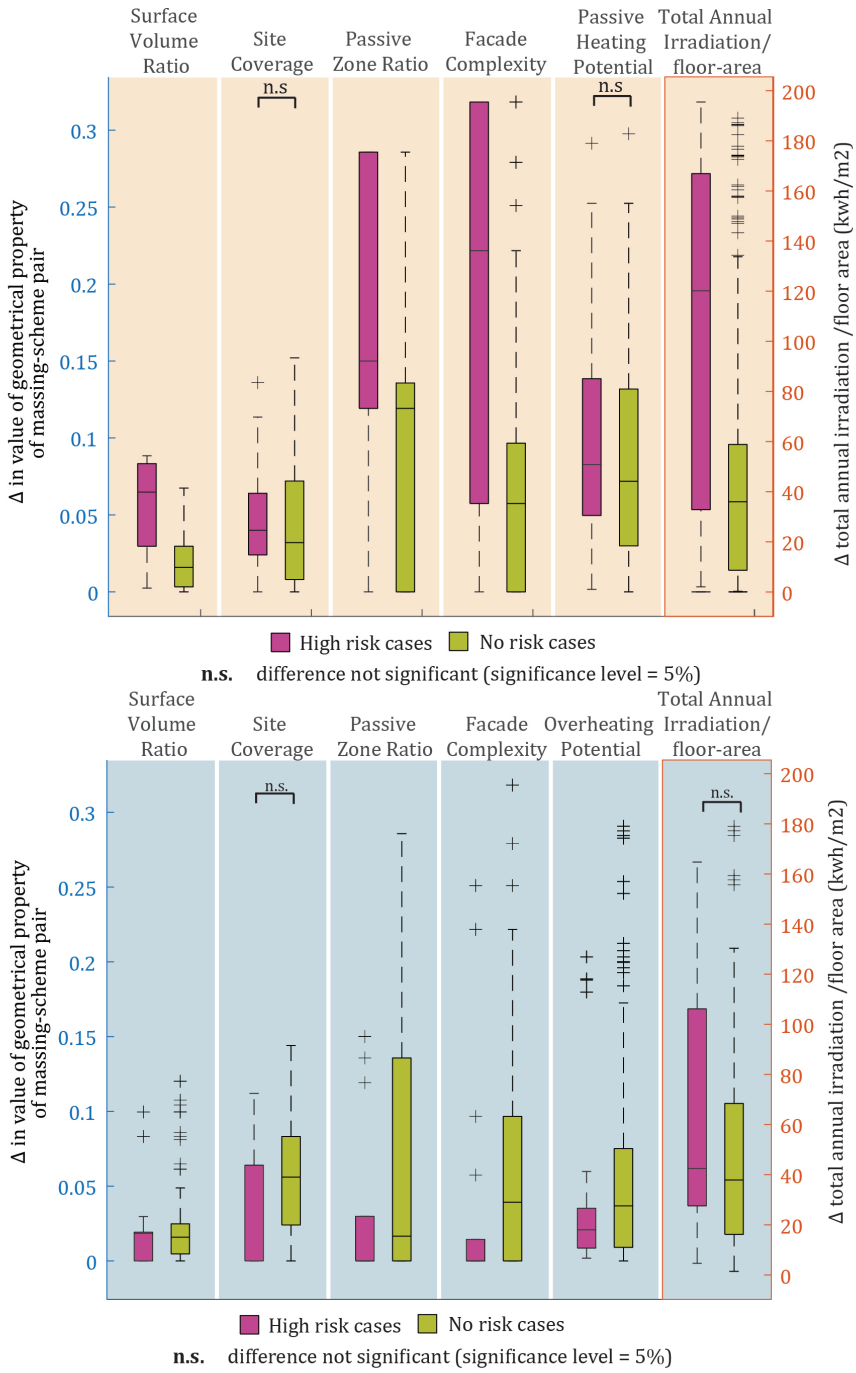


Figure 4.11 – Comparison of geometrical properties and other early design performance indicators of high risk and no risk cases in heating demand (top panel) and cooling demand (bottom panel) based evaluations. Cases shown are pairs of massing schemes in which latency effect could potentially be found (see highlighted region in figure 4.8). Pink box plots indicate subset where the risk was high. Green boxes indicate subset where the there was no risk.

(OLS) regression models are examined. In OLS modeling, the dependent variable (Y = Risk) is shown as a linear combination of various factors (x = Typology, Passive Zone Ratio, Surface Volume Ratio, etc.). These factors are also termed as *"independent variables"*, *"determinants"* or *"regressors"*.

$$Risk = \alpha\beta_1 + \beta_2 x_{1,i} + \beta_2 x_{2,i} + ...\beta_k x_{k,i} + \eta_i \tag{4.1}$$

In this equation, i indexes an observation (a pair) from the sample, α denotes the mean value of risk (when all other regressors are zero), x denotes a regressor where its first subscripts (1,2,...k) accounts for the number of regressors included in the model. η_i is the error in estimation, i.e. the portion of risk which cannot be explained by the hypothesized models.

In this study, the unit of analysis is a pair of candidate design solutions (say, scheme-A and scheme-B). Initial graphical exploration of the data set (see scatter plots in Appendix A.5) indicated that the geometrical/solar access related property of each pair of massing schemes, could be represented in a few different forms. The following forms of the regressors (x)/properties of massing schemes were tried:

x = $(x_a + x_b)/2$ (average of pair)

x = $abs(x_a - x_b)$ (difference of pair)

x = $abs(log(x_a/x_b))$ (ratio of pair)

where x_a and x_b are geometrical property/solar access related values for schemes (A,B) being compared.

Several variations of the model were tested using step-wise regression method that tries to optimize the quality of model fit by adding and removing variables in several trials. All tested models included the average form of x, in combination with either (1) the difference form of x, or (2) the ratio form of x. The logarithm in the ratio form is taken to result in same value whether x_a is the numerator or x_b is the numerator . The OLS regression factors were divided into groups indicated by alphabets K,L,M,N,O,P in table 5. All the variables were not tested simultaneously as there is likelihood of high correlations between them. For example total annual irradiation received on the facades per units of facade area is labeled L and annual irradiation received on facades normalized by floor area is indicated by M. Yet it is useful to test variations of similar variables as it cannot be known in advance which variable could be a stronger predictor. The set K includes only geometrical indicators. These are easy and fast to calculate. L,M, are annual exterior irradiation indicators. Set N,O,P consists of threshold based indicators that require a consideration of desired indoor environmental conditions and occupant behavior considerations. This particular set of indicators is of high interest for this work as high performance on these of indicators (close to 100%), implies that most of the vertical facade area is suitable for passive design related strategy (e.g. daylighting). Higher performance on this indicator could imply high design flexibility and increased robustness to future facade design decisions [Compagnon, 2004].

The coefficients (α and β_1 to β_k) are estimated through an optimization algorithm that seeks to minimizes the model error. The algorithm identifies a set of the these parameters where the squared sum of errors are minimized. R^2 of a model specifies the proportion of variation in the dependent variable (risk) explained by the hypothesized model. Higher the R^2, superior the model fit. The best model (i.e. the one with the highest R^2 is reported in Figure 4.12. The dependent variable is the risk of performance loss in sDA based evaluation at fLOD0.

A large number of variables were found to be significant whereas several others were statistically insignificant. Some of the most significant regressors are PZR ratio–X-CEX ratio, CEX-avg (quadratic), PZR-ratio (quadratic), CEX ratio, SVR-avg, SCO-avg and TIFa, among others. More details regarding variables such as use in existing studies and method for calculating them are described in section 3.3. Significance of various regressors terms can be understood as follows:

- **A linear term (e.g. CEX-avg)** indicates direct correlation between the indicator and risk. For example, CEX-avg with a positive coefficient indicates that the risk tends to be higher when the average facade complexity (CEX-avg) of a pair of massing schemes being compared is high.
- **A quadratic term** implies a non-monotonic association between the risk and the variable. For example, a positive coefficient of the quadratic CEX-avg implies that upto a certain critical value of CEX-avg, the risk falls, but starts to increase in CEX-avg after the critical value is exceeded. Regression coefficients can lead to accurate estimate of these critical values.
- **An interaction term** indicates two factors that were found to be dependent on each other for their effect on risk. For example, the term PZR-ratio-X-CEX-ratio indicates that the effect of the CEX-ratio depends on the value of PZR-ratio and vice-versa.

Figure 4.12 (a) Indicates the standardized coefficients for the model that was found to have the best fit. The factors are arranged in descending order of the magnitude of coefficients (indicated by (*) markers). The figure also shows the 95% confidence interval for indicator's coefficient using colored horizontal bars. The wider the bar, wider the confidence interval and lower the precision in estimation of the effect of the given indicator. In table 4.3 shows linear, quadratic and interaction effects that had the biggest coefficients while being statistically significant.

Figure 4.12 (b) depicts the association between actual risk (X-axis) and risk in sDA predicted by the regression model (Y-axis). Given a model fit R^2 of 0.52, the two values of risk (actual and predicted) exhibit some degree of correlation. Also, the correlation is nuanced. Near the origin, where the actual risk is low, the predicted risk is symmetrically, distributed above and below the values of zero. At higher values of the actual risk, the predicted values tend to fall below the regression line. This implies that in high sDA risk scenarios, the OLS models tend to

underestimate the risk. The findings are similar when similar OLS regression models were run for heating and cooling demand risk assessments(shown in figures 4.13 and 4.14).

Table 4.3 – Indicators with the highest standardized coefficients (top 20%) that were also statistically significant.

	sDA	**Heating demand**	**Cooling demand**
Linear effects			
CEX-avg	8.59	-2.49	1.46
PZR-avg	-1.78	3.10	**-1.97***
SVR-avg	-4.62	-	1.76
SCO-avg	3.39	-0.94	1.33
NOB-avg	-0.79	0.71	-1.06
TIFa-avg	-0.68	-1.33	-
Quadratic effects			
CEX-ratio$\hat{2}$	5.82	-	-
PZR-ratio$\hat{2}$	8.06	0.72	-
Interaction effects			
PZR-ratio-X-CEX-ratio	**-13.66***	-	-
PZR-avg-X-PZR-ratio	-	**-3.38***	-
CEX-avg-X-PZR-ratio	-	-3.30	-

Facade Complexity (CEX); Passive Zone Ratio (PZR); Surface Volume Ratio (SVR); Site Coverage Ratio (SCO); Number of Buildings (NBO); Total Annual Irradiation/facade area (TIFa)

suffix **-avg** indicates average of variable for a pair of massing schemes

suffix **-ratio** indicates ratio between a pair of massing schemes

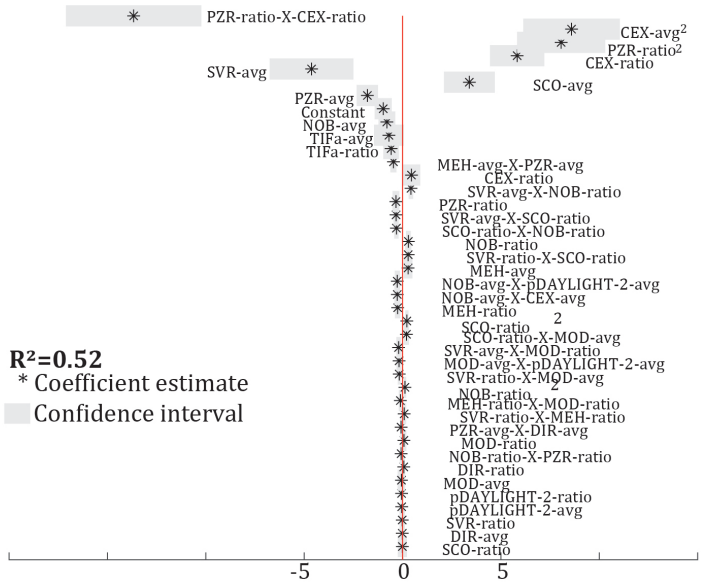
'X' between variables indicates interaction effects between two variables

(-) Insignificant or small standardized coefficient

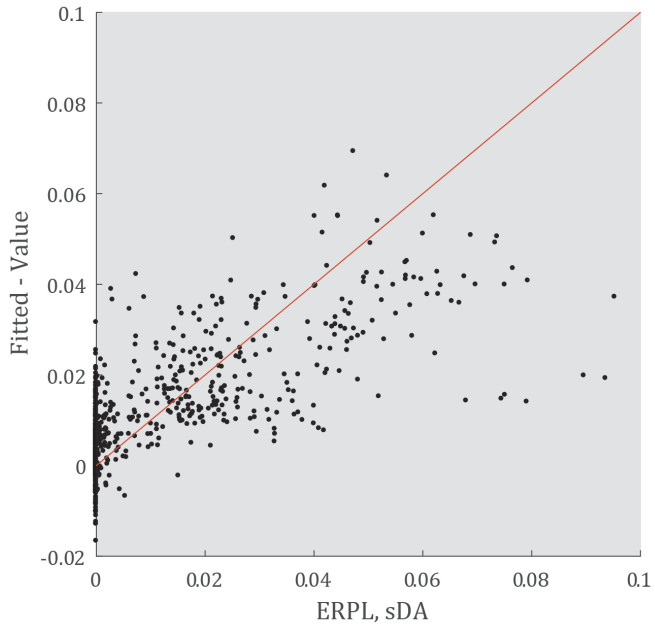
(*) variable with highest standardized coefficient

Performance potential indicators (pHEAT, pOVERHEAT, pDAYLIGHT) were considered indicators of high interest since they are based on the percentage of facade area meeting irradiation/illuminance threshold. High performance on these indicators could imply robustness to facade design related choices. In all models these indicators were assigned relatively smaller co-efficients (compared to simple geometrical indicators (PZR, CEX, SVR)) and were found to be insignificant through out in the quantile regressions. This is likely be the result of the density of the built massing schemes for the experiment (1.0) and high correlation with total irradiation received on the facade (TIFa) variable at this density. Total irradiation received on the facade (TIFa) per unit of floor area was a stronger indicator (than performance potential indicators) in the daylight and heating demand risk models, but less important than the geometrical indicators.

Several variables were statistically significant. However, the models reflect a modest fit (R^2 <52%). Figure 4.15 shows the distribution of the dependent variable (risk). The distribution of

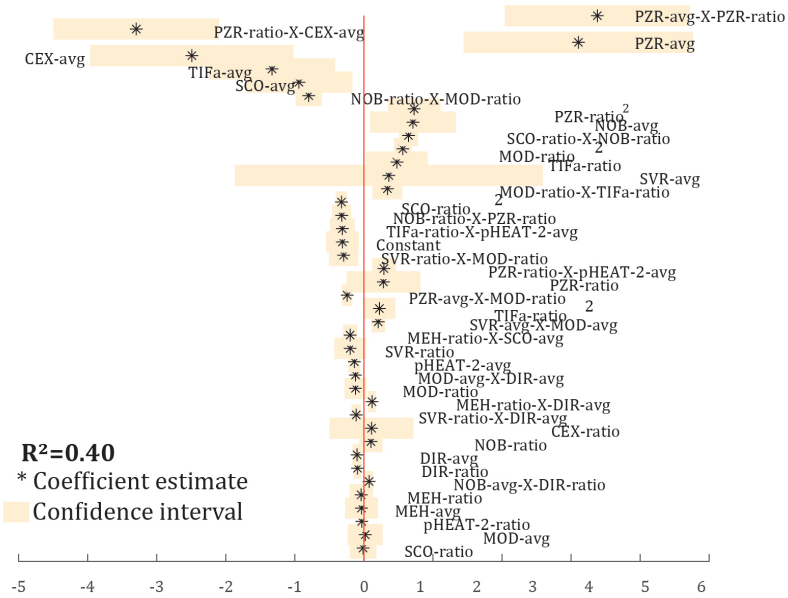


(a) Standardized Coefficients of linear regression model Y=ERPL, sDA

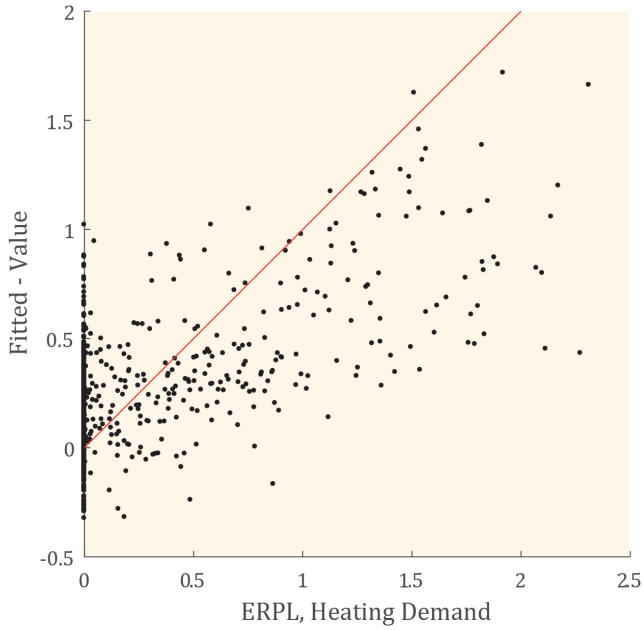


(b) Linear model fit

Figure 4.12 – Linear regression for predicting risk in conceptual stage decision making based on sDA (a) a shows model factor and respective coefficients ordered by absolute value of coefficients (not all are significant) (b) shows the over all model fit against actual risk value. the red line indicates perfect fit.

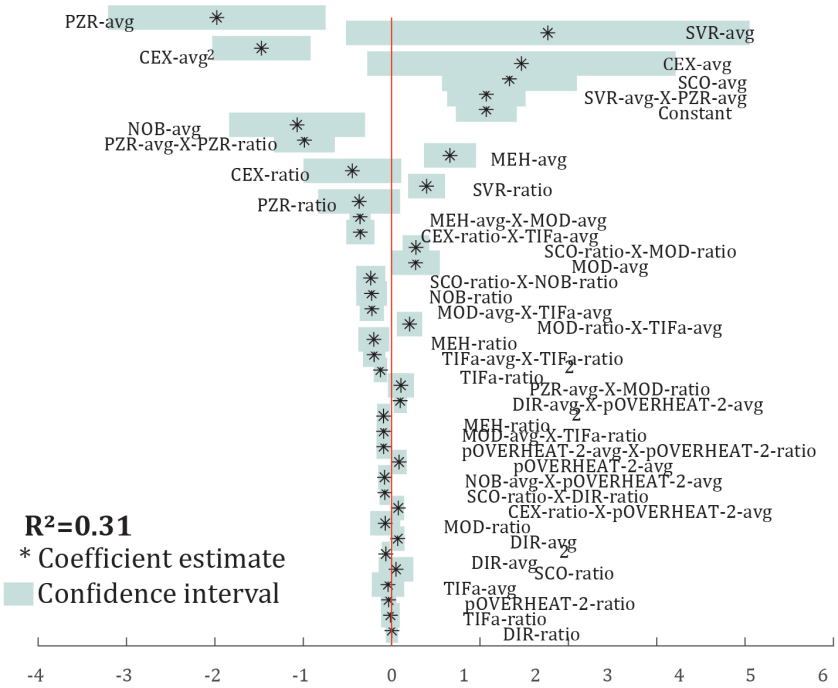


(a)Standardized Coefficients of linear regression model Y=ERPL, Heating Demand

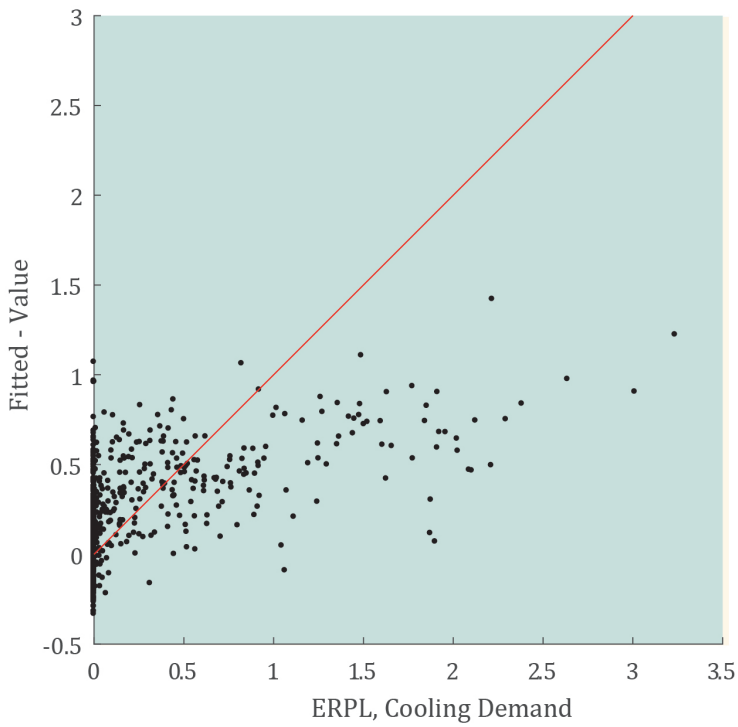


(b) Linear model fit

Figure 4.13 – Linear regression for predicting risk in conceptual stage decision making based on Heating Demand (a) a shows model factor and respective coefficients ordered by absolute value of coefficients (not all are significant) (b) shows the over all model fit against actual risk value. the red line indicates perfect fit.



(a) Standardized coefficients of linear regression model Y=ERPL, Cooling Demand



(b) Linear model fit

Figure 4.14 – Linear regression for predicting risk in conceptual stage decision making based on Cooling Demand (a) a shows model factor and respective coefficients ordered by absolute value of coefficients (not all are significant) (b) shows the over all model fit against actual risk value. the red line indicates perfect fit.

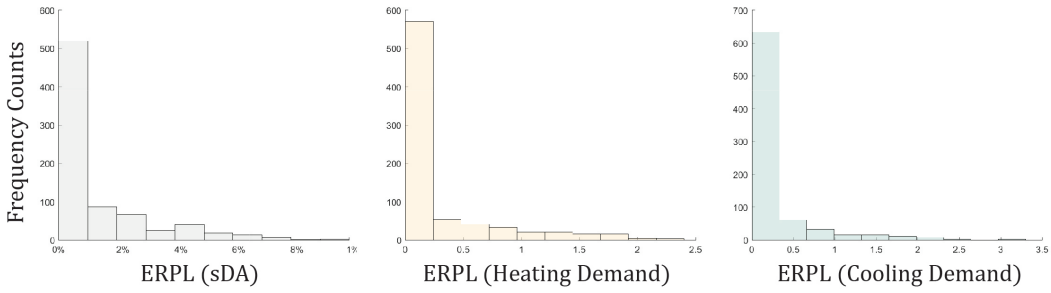


Figure 4.15 – Distribution of risk values in conceptual stage evaluations based on sDA, Heating and Cooling Demand.

risk is highly skewed on all three metrics. That is, the distribution is not symmetric around the most prevalent value and is skewed towards the zero value. For such distributions, there is value in exploring other methods of analysis. Given the modest fit achieved using linear regression and the skewed nature of the distribution of risk, quantile regression method was explored which could explain the deficiencies found in the linear regression model.

4.4.2 Quantile Regression

Unlike the OLS regression which models the average risk, quantile regression models, model specific quantiles of the risk. Quantile regression is considered appropriate when the dependent variable exhibits skewed distribution. In this subsection, the results of the quantile regression models are presented.

In quantile regression framework, the sample is divided across equidistant quantiles of the dependent variable (i.e. risk) and the corresponding set of regression coefficients is estimated at each of the quantiles by minimizing the absolute error. As such, the coefficient of a regressor may vary across the quantiles.

In Figure 4.16, the quantile regression coefficients for the sDA risk are presented in the *"Quantile Plots"*. Each graph corresponds to one of the coefficient estimates specified in the regression model. Across all the coefficients presented in the Figure4.16, the sample is divided across 10 equi-distant quantiles depicted on the X-axes. Y-axis depicts the mean of the coefficients. The grey-bands depict the 95% confidence interval of the coefficients. If at a particular quantile, the grey band includes the zero-horizontal line, the corresponding variable is considered statistically insignificant. If, however, the grey bands lie on a specific (and same) side (above or below) of the zero- horizontal line, it reflects statistical significance of the regressor in predicting that particular quantile (e.g. median) of the dependent variable. The horizontal solid line in black, the solid line in red and the dotted lines in red respectively depict the zero line, the average coefficient and the confidence interval of the coefficients for the whole sample (rather than for a specific quantile).

Consider the quantile plot for total annual solar irradiation per unit of façade area - average

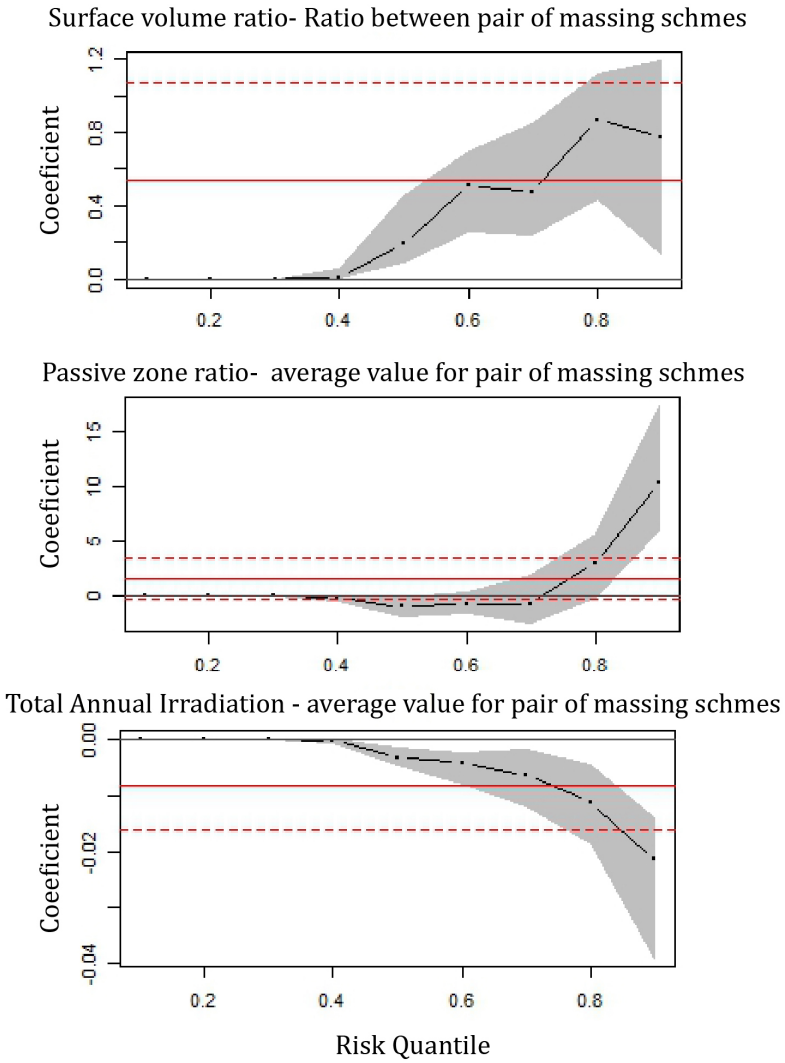


Figure 4.16 – Quantile regression for risk at based on 10 quantiles - daylight.

for example in figure 4.16 (plot at the bottom). Until the quantile 0.4, the coefficient is nearly zero and insignificant. However, the coefficient turns statistically significant and increasingly negative at higher quantiles. Similarly, average passive zone ratio is zero until the quantile 0.7, but become positive and significant at higher quantiles. All these factors which become statitically significant at higher quantiles, indicate a strong relationship with high risk values.

Figure 4.17 presents the quantile plots for risk in heating demand which also show a similar trend of change in coefficient at 0.4 quantile of risk, where variables such as ratio of number of buildings in pair of massing schemes, average passive zone ratio, ratio of mean-building-height of a pair of massing schemes and average passive-heating-potential become statistically significant. Average value of passive zone ratio and ratio of mean-building-height acquire negative coefficients meaning that in high risk cases, they tend to reduce the risk. On the other

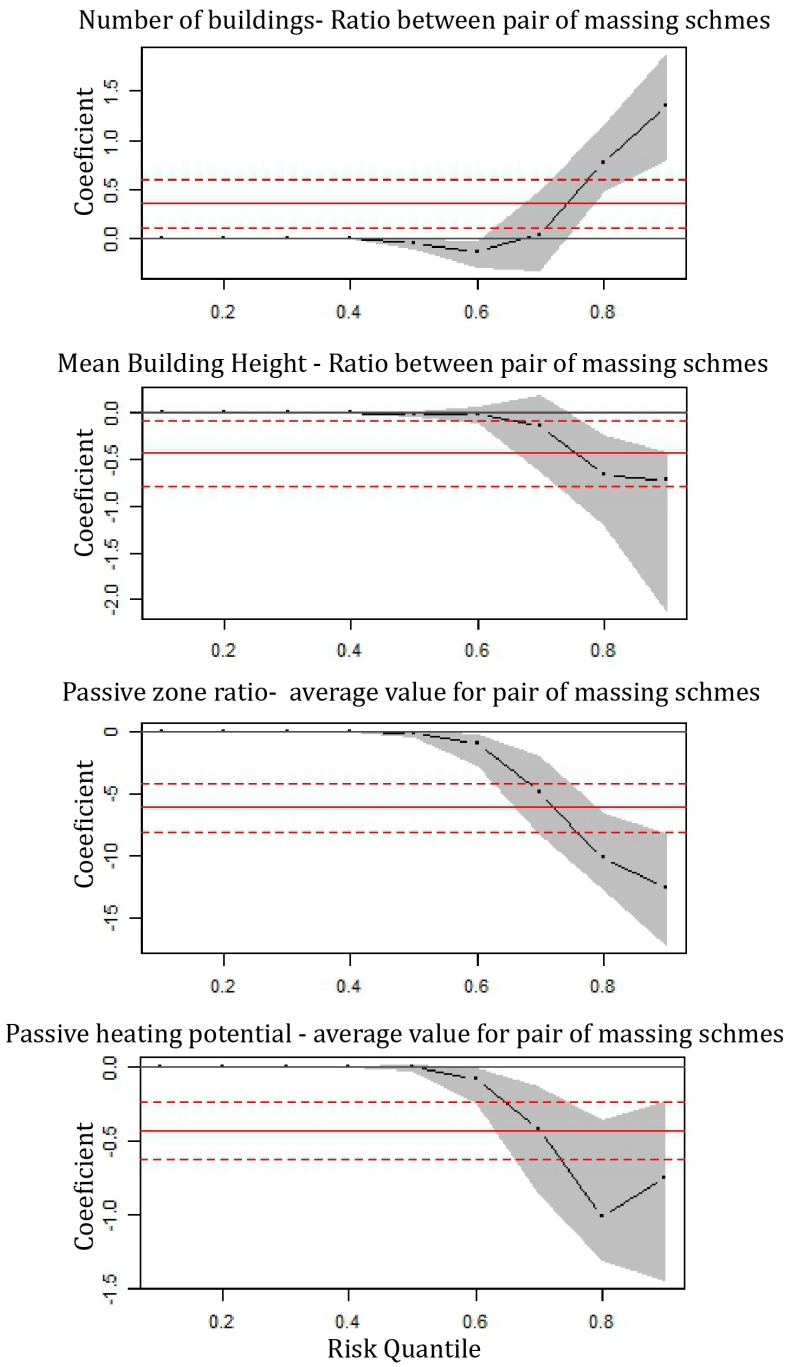


Figure 4.17 – Quantile regression for risk at based on 10 quantiles - heating.

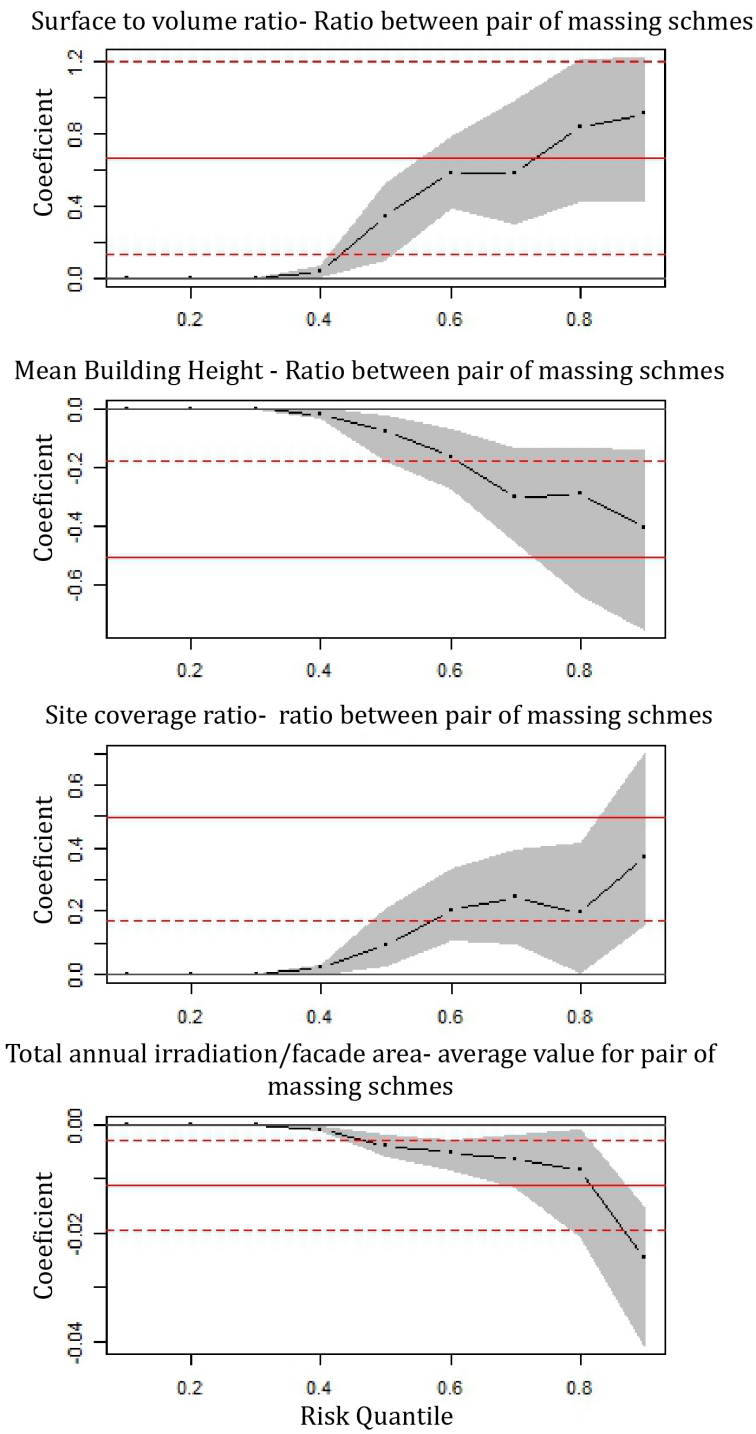


Figure 4.18 – Quantile regression for risk at based on 10 quantiles - cooling.

hand ratio of number of buildings has a positive coefficient implying that large difference in number of buildings in the massing schemes being compared Finally, figure 4.18 shows the quantile plots for risk in cooling demand assessments. On cooling demand ratio of surface-to-volume and site coverage induce risk but ratio of mean-building-height and average of total annual irradiation per unit floor area have a negative effect on risk in high risk conditions. Indicators that are not shown in figures 4.16 - 4.18 remained insignificant as per the quantile regression analysis.

4.5 Summary

In this chapter the results of 780 relative performance comparisons between pairs of massing-schemes at fLOD0 have been examined. The performance loss resulting from the decision taken at fLOD0 was calculated using possible performance evaluations that could be achieved under design scenarios at fLOD3. From these 780 comparisons, four metrics of risk in conceptual stage decision making were calculated. Some of these metrics help us examine the quality of decision making (i.e. relative error and percentage of high risk cases) at fLOD0 and other metrics allow us to understand the implications of decision making at fLOD0 (i.e. Average risk per high risk case and maximum observed risk). The maximum observed loss in all the comparisons was also reported in figure 4.3. The findings discussed in this chapter are summarized below under a few heads:

Reliability of evaluations across metrics: Difference in number of high-risk evaluations among the metrics shows that performance evaluations on different metrics at the same fLOD are not equally reliable. 1 in 5 (incidence rate of high risk =22%) evaluations at fLOD0 on sDA were found to be high risk. The same statistic for heating and cooling was found to be 1 in 7 (incidence of high-risk= 15%) and 1 in 12 (incidence of high-risk =8%) for heating and cooling demand evaluations. Design decisions made at the same fLOD (fLOD0) on different metrics would have different likelihood of encountering high risk. Relative errors were found to vary between 32% to 20%. This metric indicates to what extent relative performance improvement expectations that a DM has at the commencement of the design process could be met using a sequential design process. While the difference between sDA and heating demand metrics based on incidence of high risk was 7% (22%-15%), the difference between them in terms of relative error is 2% (32% - 30%). Thus high-risk could be encountered more often in sDA based assessment (compared to heating demand evaluations), but the overall losses are similar in extent compared to the expected gain.

Types of risk: All three sources of loss identified in the methodology for risk assessment (rank reversal, latency effect and insufficient gain) were found in daylight and cooling demand evaluations at the scale of the experiment. However in heating demand evaluations, only latency effect was found relevant. As mentioned earlier, this could be a result for the experimental setup and under different experimental conditions, other sources of loss could be found in heating demand evaluations also.

Magnitude of risk: Per 100 decisions made at fLOD0 based on sDA one can expect to loose an aggregate of 112% on sDA where the average risk is 1.12% per fLOD0 decision. The aggregate loss on 100 heating demand and cooling demand evaluations was found to be 101 kWh/m2-year, 21 kWh/m2-year respectively. The aggregate loss may not be of great interest to an individual DM as all decisions may not result in loss. It is anticipated the aggregate loss would be of interest to policy makers in knowing the implications of BPS model LOD for reliability in decision making.

Building level indicators of risk: Building level characteristics of massing schemes such as facade complexity (CEX), passive-zone-ratio (PZR) and surface to volume ratio (SVR) reflect the geometrical properties of the buildings that form a massing-scheme. These and other such properties that were found to be significant, may warrant attention from designers. Irrespective of the metric, the two most important regressors were passive zone ratio (PZR) and facade complexity (CEX). Both of these variables are building level geometrical descriptors(and not site or building placement related indicators). The linear regression analysis shows that if two massing schemes have high passive zone ratio or high facade complexity then the risk could be low (coefficient has negative sign). To further interpret such a finding, a building with a shallow plan will have a high PZR. These findings thus indicate that when two massing-schemes with shallow floor plans are compared, risk tends to be lower on daylight and cooling metrics. This finding was also corroborated by the quantile regression analysis for the daylight metric to some extent where the negative relationship between risk and passive zone ratio holds true for low risk values. Very high risk values (80th percentile and above) were found to come from pairs of massing-schemes with high-average passive zone ratio. Thus massing-schemes with very high passive zone ratios, when compared to each other, are not immune to risk. Please refer to table 4.3 for summary of important indicators.

Prediction of risk: Several OLS regression models were run to explain the variation in risk. An attempt was made to explain the ERPL variation in terms of the combined characteristics of a massing-scheme pair. Several variables were statistically significant. However, the models reflect a modest fit (R^2 <52%). Therefore, in their current state, these models cannot be used as strong predictors of risk. However, given that these models are derived from a controlled experiment (and confounding factors should be limited in number), the structural relationships presented above, between the indicators (such as PZR, CEX) and risk should be considered trustworthy.

5 Using Risk as a Determinant of Model fLOD

In this thesis, ERPL is presented as a risk metric to evaluate the quality of decision making at the conceptual design stage. A method for evaluating the risk as 'high or 'low' has been introduced for the DM to judge if the risk is indeed high enough for him/her to delay the choice between two massing-schemes until there is more clarity on one or more facade details.

The risk metric could thus be used to choose the fLOD at which a conceptual design decision should be made. Experimental results reported in Chapter 4 indicate that increasing the fLOD is not always neccassary and in a number of cases, decision made at the lowest fLOD (fLOD0) are reliable. In other cases, it may be prudent to increase the BPS model fLOD, re-evaluate performance and then make a decision between the massing-schemes. In this chapter, the possibility of having a generalized view on the ability of various fLODs to improve the decision outcomes is explored.

The sensitivity of the risk evaluation to two following aspects of the DM is also assessed:

1. The threshold (dt) used by the DM to assign ranks to design alternatives.
2. The DM's level of autonomy in making detailed design decisions beyond the conceptual design stage. The DM's autonomy in decision-making is assumed to affect the DM's priorities when determining the risk threshold.

Considerations such as these are an essential step towards a holistic discussion on the role that ERPL metric could play in improving decision making at the conceptual/early design stage. Figure 5.1 shows the scope of this chapter that includes all intermediate fLODS between fLOD0 and fLOD3 and characteristics of the DM.

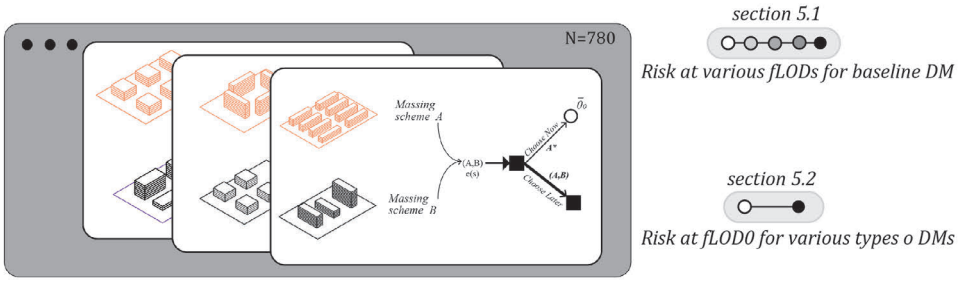


Figure 5.1 – Using experimental results to assess value of fLODs and impact on various types of DMs.

5.1 Role of fLODs in making better design decisions

5.1.1 Value of fLODs in decision making

Thus far, the risk of performance loss due to unknown design details at the conceptual design stage has been examined. Risk can also be interpreted as the possible loss that additional design details can help avert[1]. The ERPL metric can thus be used to assess the value of various design details in terms of the performance gain that they can help achieve through improved reliability in decision making. Value of design details is different from relative importance of design details assessed through sensitivity measures and factor screening studies (for examples, see Tian [2013]). While measures of sensitivity are useful in identifying important design factors, ERPL can be used to assess the design-gap that can be recovered by specifying certain design details. Unlike sensitivity coefficients, it is expressed in the units of the performance metric.

Figure 5.2 shows the aggregate ERPL from 100 (based on average risk observed in 780 comparisons) relative performance comparisons. ERPL is shown at each design path at each fLOD. The value of ERPL at each node shows the ERPL, if in all 100 comparisons the DM arrives at the given fLOD through that node. For example, if in all 100 cases the DM was to assign 40% WWR at fLOD1 (no blinds modeled yet), the ERPL would be 170%. sDA is the percentage of floor area of a project or building that meets a specific indoor illumination related threshold. The maximum value that can be achieved on this metric is 100%. The value 170% indicates a collective loss of percent points on sDA in 100 projects. This collective loss is presented in a form that could be of interest to a policy maker who can control decision making practices on number of projects (e.g. 100) and prevent a collective loss (e.g. 170% points). It was not meaningful to frame this discussion around the average value (1.7% points on sDA) as loss may not be incurred on every project. It may be noted that the ERPL does not always decrease with increasing fLOD, although overall it does decrease as the fLOD (design information) increases. ERPL is found to be subject to both, the fLOD and the design path followed. On

[1] Expected opportunity loss is also interpreted as the Expected Value of Perfect Information (EVPI) [Raiffa and Schlaifer, 1961] where ERPL is the difference in payoff from a decision made with current, versus the 'perfect information'. Perfect information here could imply a state where DM has eliminated all sources of uncertainty that can be overcome (e.g. uncertainty in design features)

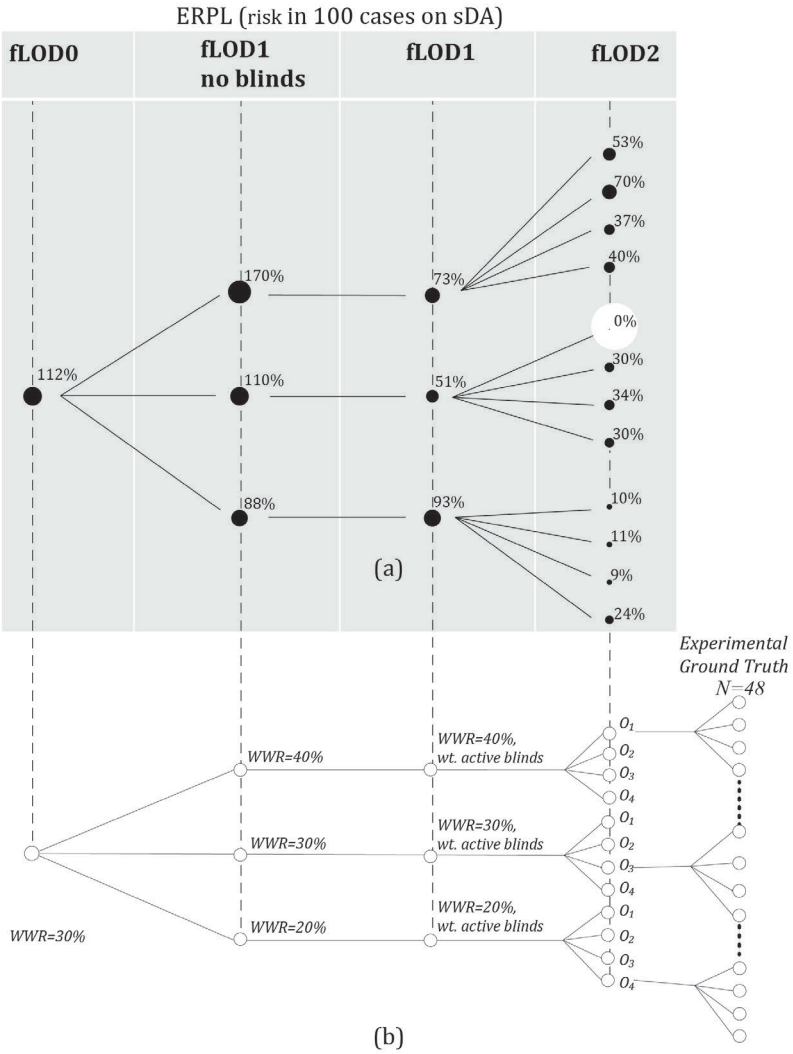


Figure 5.2 – ERPL or risk (% points on sDA) observed in 100 pair-wise choices between massing schemes at various fLODS and design-specification within each fLOD. Cumulative risk is presented for interpretation by a policy maker as the expected loss if DMs are stipulated to make decisions at specific fLODs. Zero ERPL indicates that further design details are not needed (top panel (a)). Bottom panel (b) indicates the corresponding fLODs and the design-specifications within each fLOD.

low WWR, it is found to decrease steadily. On one design path ERPL falls to 0. An ERPL of 0 indicates that further fLODs do not add any decision making value [2]. In other words, at low WWR specification, once the window distribution strategy is decided (fLOD2), a choice between massing schemes can be made. All further design decisions (irrespective of choice of balcony, no opportunity loss was found).

In thesis, the risk in decision making is reported in relation to the BPS model fLOD with the

[2] decision making value of a fLOD refers to the reduction in risk that can be achieved by specifying design details at that fLOD.

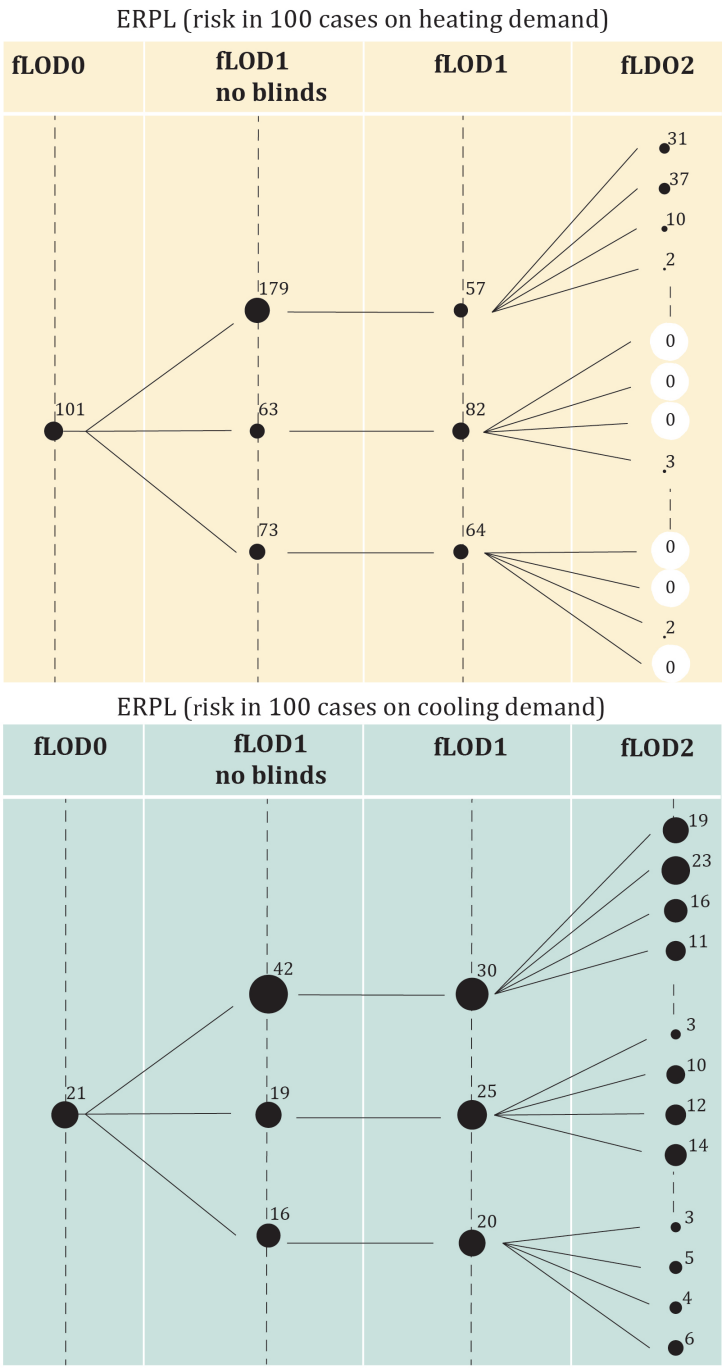


Figure 5.3 – ERPL or risk (kWh/m2-year) observed in 100 pair-wise choices between massing schemes based on heating demand (top panel) and cooling demand (bottom panel) at various fLODS and design-specification within each fLOD. Cumulative risk is presented for interpretation by a policy maker as the expected loss if DMs are stipulated to make decisions at specific fLODs. Zero ERPL indicates that further design details are not needed.

underlying presumption that higher the fLOD, lower the uncertainty and thus lower the risk from lack of knowledge regarding design features that are yet to be specified. This was not found to be the case with fLOD1-no blinds and fLOD1(active blinds included). Introduction of blinds in some cases (e.g. Heating demand evaluations at 30% WWR in figure 5.3 (top panel, middle branch at fLOD1-no blinds)) and also introduction of fLOD1-no blinds at high WWR 40% (see top most branch at fLOD1 no-blinds in figure 5.2) resulting in higher risk than the lower fLOD. As seen in the example presented figure 3.15, the ranks between alternatives are in 'transition' at these fLODs on daylight and cooling metrics. In future works, these two fLODS (fLOD1-no blinds and fLOD1(active blinds included)) could potentially be merged to avoid fluctuations in ERPL.

Similar observations can be made in figure 5.3, for heating and cooling demand assessments respectively where the ERPL value is found to decrease as the fLOD increases, however significant differences are also seen between design paths taken at a fLOD. ERPL is found to be both fLOD and design path dependent in cooling demand assessments. At fLOD2, the EPVI remains non zero in all cases indicating that even fLOD 2 may not be sufficient to avoid loss in risky cases.

5.1.2 When is low fLOD sufficient?

The experimental results reported in chapter 4 show that high risk was not found in several cases(section 4.1). This experimental data set was used to estimate the performance difference at fLOD0 beyond which the DM can safely make a decision without the need of any further design details or a risk assessment. Figure 5.4 shows the performance difference needed at fLOD0 in order to limit the probability of experiencing high risk to 5% (10% and 20% are also shown). This figure shows that if the baseline DM observes a performance difference of 18.3% on sDA at fLOD0 between two design alternatives, the risk of performance loss is minimal (probability of high-risk <5%). Typically, a difference of 10% (sDA points) is considered noteworthy at the early design stage [Iversen et al., 2013]. This industry practice appears to account for the expected relative error in simulation of point-in time illuminance values [Mardaljevic, 1999]. To avoid errors due deficiency of design information in daylight simulation models at fLOD0, additional (18.3%) difference may be needed.

On heating and cooling demand performance difference of 2.8 kWh/m2-year, 8.3kWh/m2-year can achieve the same effect. On these metrics no other criteria or best practice were found for comparison. These performance difference values could potentially be used as rules-of-thumb by DMs evaluating performance at fLOD0 could then gauge the the utility of engaging in a risk assessment. If the performance difference between two design alternatives are found to be high (>18.3 % on sDA, > 2.8 kWh/m2-year on heating demand, > 8.3kWh/m2-year on cooling demand) the risk assessment could be deemed unnecessary.

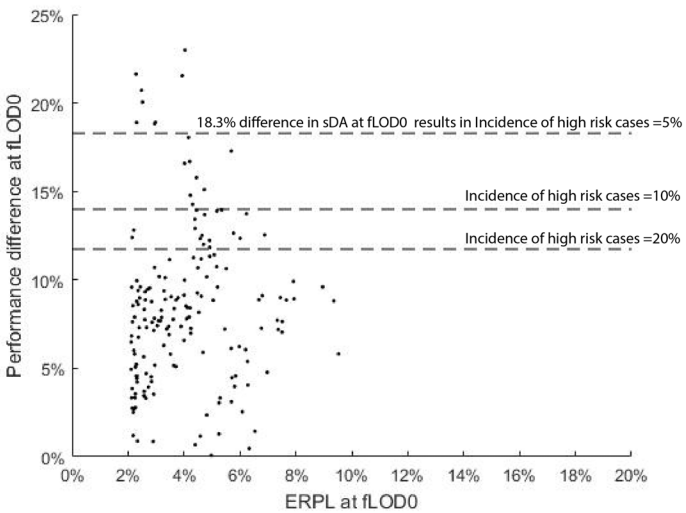


Figure 5.4 – Performance difference at fLOD0 needed to limit chance of encountering high-risk to 5%,10% and 20% (Given that dt =10%).

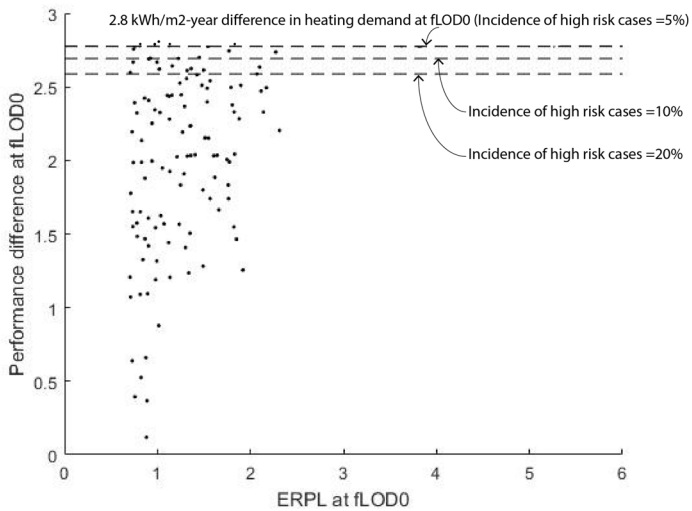


Figure 5.5 – Performance difference at fLOD0 needed to limit chance of encountering high-risk to 5%,10% and 20% (Given that dt =2.8 kWh/m2-year).

5.2 Determining user-specific early design stage model fLOD

Risk assessment (ERPL) or interpreting it as ERPL can facilitate understanding the degree of reliability of BPS models. However, simply quantifying the risk does not ensure that all DMs will use it in the same manner to prevent performance loss. High risk conditions prompt a DM to delay the design decision until more design information becomes available (e.g. WWR, balcony depth and type) or the BPS models' sophistication can be increased (e.g. add blind usage behavior). However any rational DM would like to limit their design, decision making

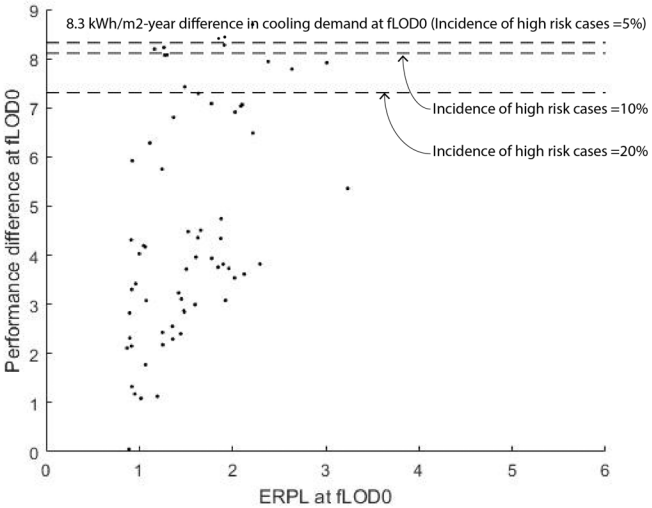


Figure 5.6 – Performance difference at fLOD0 needed to limit chance of encountering high-risk to 5%,10% and 20% (Given that dt =3.6 kWh/m2-year).

and modeling efforts by choosing a massing-scheme as soon as possible. Delaying the decision making increases the project work-load, but making a decision too soon can lead to loss of performance. These two opposing forces could affect a DM's tolerance for risk.

Several researchers working in the BPS domain, especially those related to use of BPS for design decision making have delved into the DM's beliefs and actions. Decision theory offers various constructs to aid decision making, however, in order to foresee the outcome of decision making, the attitude and preferences of DM have to be brought into the picture. For example De Wit and Augenbroe [2002] examined the effect of uncertain model inputs on the estimated hours of overheating and the subsequent decision on whether a cooling system must be included in the design. Instead of a point estimate of the overheating hours, inclusion of uncertainty in model parameters made it possible to estimate the uncertainty in number of overheating hours. This then becomes a classic problem of decision making under uncertainty (DMUU). For the same performance threshold of 150 hours of overheating, they then considered two DMs with different beliefs regarding acceptable probability and extent of exceeding this threshold. For the same design problem and same performance threshold (150 hours of overheating), they showed that the DM's individual levels of utility derived from could lead to different design decisions. Other studies (e.g. [Heo et al., 2012];[Kim et al., 2014]) have also found DM's attitude toward risk as important driver in decision making.

Bleil De Souza and Tucker [2016] proposed development of 'personas' of the BPS users, specific to certain design decision making contexts. These personas are meant to serve as figurative tools to investigate the interaction of users with BPS tools. In this chapter a similar approach is taken where some potential properties of the BPS based DM are suggested that could affect his/her attitude towards risk evaluation or affect the evaluation itself. These characteristics

were prompted by the decision making context (binary choice between design alternatives at low LOD) and the methodology presented to address the risk in decision making in this context. The characterizations of the properties of the DM, presented in this thesis, are speculative in nature and provide ground for further investigation.

5.2.1 DM's preferred minimum performance differentiation threshold

The first characteristic that is explored in this chapter is the threshold for assigning ranks at fLOD0. The analysis for sensitivity towards this threshold, builds upon the main experiment in Chapter 4 by considering alternative values for the threshold 'dt' that could be adopted by a DM. In column (1) of table 5.1, shows thresholds that could be suitable for a DM who is sensitive to performance changes on a given metric and is willing to choose one massing-scheme over another, if a modest improvement in performance is observed. In column (2), the decision criteria values are suited for a DM who is more discerning and would thus choose one massing-scheme over another only if significant performance improvement is seen. The threshold for such a DM is assumed to align with performance improvement needed to upgrade the evaluation of the project from a compliant level to high-performance level.

Figure 5.7 shows the occurrence rate of high-risk cases at fLOD0 depending on the decision criteria value found to be suitable by the DM. 'dt1' and 'dt2' in figure 5.7 indicate the decision threshold value from column (1) and (2) in Table 5.1. At dt1 (or at lower differentiation threshold) prevalence of high-risk in decision making is notable on all metrics. Adopting a higher, more stringent decision criteria makes the chances of high-risk in decision making (ERPL marked by bold solid line) negligible only for cooling demand. For all metrics, while adopting a more stringent threshold understandably lowers chances of rank reversals, it also leads to more latency-related losses, as more comparisons are considered to be equivalent in performance.

While overall risk may not always be eliminated by simply having a stringent decision threshold, a DM could choose their threshold for a given metric based on the type of opportunity loss he/she wants to avoid. For example, if a DM wants to avoid rank reversal on sDA, a minimum differentiation of 15.7% on sDA should be considered suitable for the conceptual stage (number of cases with rank reversal falls below 5%). On heating demand evaluations, the threat of rank reversal was found to be low for both dt(1) and dt(2) (<5% cases) (Figure 5.8). Once a DM observes an appreciable performance difference between two design alternatives (>2.6 kWh/m2) then simply choosing the higher performing form was found to be a reliable decision. Thus, a DM's decision threshold is found to be less critical on heating demand evaluations. On the cooling demand metric, it was found that higher stringency in threshold to be a successful strategy for lowering the overall risk of performance loss (Figure 5.9).

Figures 5.4 and 5.7 show the limits to improving the reliability of decision making on sDA by changing the decision thresholds. Say, for example, a DM finds that at fLOD0, the difference between two design alternatives is 16% on sDA. As seen in figure 5.4, this difference is not

Table 5.1 – Possible values for decision criteria for various performance metrics.

Metric	**1-Minimum performance differentiation for assigning ranks (high-sensitivity DM)**	**2-Minimum performance differentiation for assigning ranks (low-sensitivity DM)**
sDA (daylight)	10% (sDA units) improvement in sDA is generally considered to be a minimum appreciable difference between design alternatives [Iversen et al., 2013]	20% (sDA units) improvement needed change from acceptable to preferred space evaluation LEED v4 *
Annual Heating Demand	2.8 kWh/m2-year (reduction in annual heating demand from code compliant [SIA 380, 2009] to MINERGIE rating [Association MINERGIE, 2015])	4.3 kWh/m2-year (reduction in annual heating demand from code compliant [SIA 380, 2009] to MINERGIE-P rating [Association MINERGIE, 2015])
Annual Cooling Demand	3.6 kWh/m2-year (commensurate active cooling energy requirement to advance from comfort Category III to Category II (EN15252))**	6.9 kWh/m2-year (commensurate active cooling energy requirement to advance from comfort Category III to Category I (EN15252)**)

(*) see LEED v4 checklist - https://www.usgbc.org/resources/leed-v4-building-design-and-construction-checklist

(**) See Appendix A.2 for detailed description on calculation of cooling demand based decision threshold.

Table 5.2 – Number of times a DM with low versus high sensitivity to performance improvement would encounter high-risk at fLOD0.

Metric	High-Sensitivity DM	Low-Sensitivity DM
Daylight (sDA)	1 in 5	1 in 4
Heating Demand	1 in 7	1 in 10
Cooling Demand	1 in 12	1 in 44

sufficient to completely avoid risk. However this is sufficient (see dotted line in 5.7 (a)) to reduce chance of rank reversal. Also, the risk of latency effect is only applicable only if the DM chooses disregard the difference of 16% at fLOD0 and treats the two massing-schemes equally.

Table 5.2 shows how often the two types of DMs would encounter high risk at fLOD0. Wile on some metrics such as cooling demand, a DM how is less sensitive to relative performance gain can reduce chances of making poor decisions (unreliable decisions), on other metrics such as sDA and annual heating demand, being less sensitive to relative performance gain does not lead to higher quality (i.e. more number of reliable decisions with low risk) in decision making.

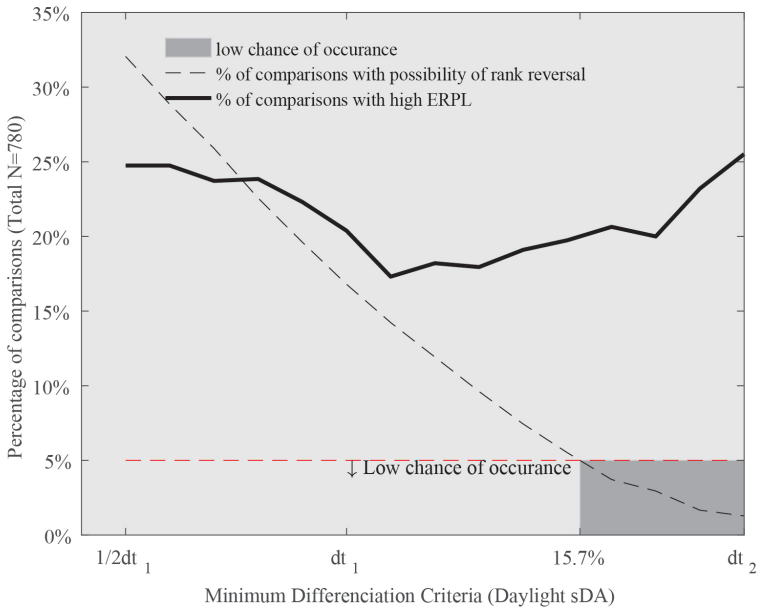


Figure 5.7 – Effect of decision threshold (dt) adopted by DM on risk of relative performance loss in sDA. dt2 indicates a low sensitivity decision maker who requires a greater performance gain between two massing-schemes in order to make a definitive choice. dt1 indicates a high sensitivity decision maker.

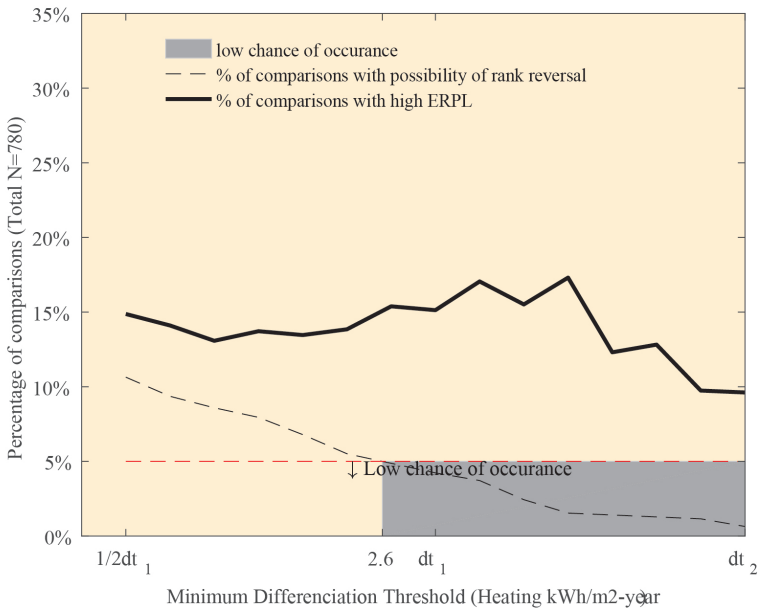


Figure 5.8 – Effect of decision threshold (dt) adopted by DM on risk of relative performance loss in annual heating demand. dt2 indicates a low sensitivity decision maker who requires a greater performance gain between two massing-schemes in order to make a definitive choice. dt1 indicates a high sensitivity decision maker.

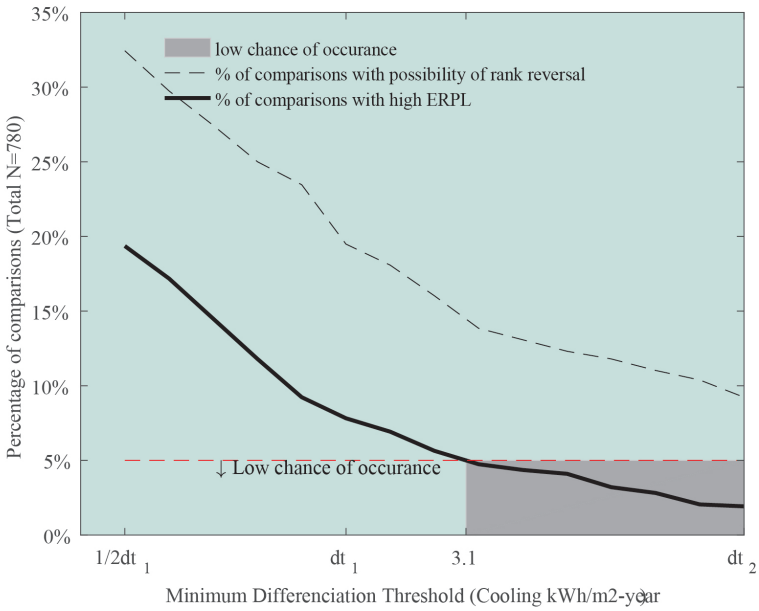


Figure 5.9 – Effect of decision threshold (dt) adopted by DM on risk of relative performance loss in annual cooling demand. dt2 indicates a low sensitivity decision maker who requires a greater performance gain between two massing-schemes in order to make a definitive choice. dt1 indicates a high sensitivity decision maker.

5.2.2 DM's autonomy in decision making

This work shows that low BPS model fLOD in some cases can lead to low reliability in decision making. However, the BPS model fLOD cannot be increased at will and depends on the degree of autonomy of the DM in decision making. Fragmentation in design scope of different design team members is considered an important barrier to adoption of energy efficient performance practices in building design [Metz et al., 2007]. For example, if the DM is commissioned to only work on the massing-scheme at the neighborhood scale and at a later stage of project development design teams may be hired produce detailed designs. If a DM has no control over design decisions beyond the massing scheme, the risk evaluation has little value for him/her.

In this section, a less extreme situation is considered where a DM who could potentially make additional design decisions to increase the BPS model fLOD, but is reluctant to do so. While he he/she is affected by some degree of fragmentation in scope of work but can increase the fLOD if compelling reasons are presented. Such a DM is considered to be working fragmented design process. In contrast, is a DM who is fully autonomous and has control of the design process and design decisions. These DMs could potentially have different priorities when establishing 'high' risk threshold at fLOD0. The implication of this issue are presented here on the experimental data set as a whole. Different risk thresholds, resulting from different priorities of the DM would lead to different incidence rates in high-risk. Here difference in incidence rate of high-risk does not indicate a change in the opportunity loss that occurs,

but simply how often DMs in working specific context (fragmented or autonomous) would categorize a certain value of risk as 'high' or 'low'.

Figures 5.2 - 5.3 show that distribution of risk at the tail nodes (at fLOD2) varies to a large extent. The DM does not know which design path will be followed in the future and thus the at fLOD0, the DM must rely on the average risk from all future facade design paths. As a decision making strategy, a threshold is drawn on the average risk at fLOD0 to try and identify cases with acceptable risk at all future design paths.

Drawing a threshold value on a continuous value (ERPL) for a binary classification (risk is 'high' versus 'low') is not trivial. For example, if the threshold is kept high, many risky cases may get classified as low risk. If the threshold is kept low, cases with minimal risk may get classified as high risk. A DM working in a fragmented design context may want to choose a risk threshold that helps in correct identification of cases that lead to unacceptable performance loss and avoid any false alarms (or false positives (FP)). Avoiding false alarms in identification of risky cases may not be large concern for an autonomous DM as they are themselves responsible for all design decision that need be made. The only concern for the autonomous DM is if facade related decision need to be concurrently with the massing, or not. Such a DM may be driven by the sensitivity of the threshold to identify all risky cases.

Measures of *accuracy* and *sensitivity* [Fawcett, 2006] from diagnostic statistics were used to identify suitable thresholds for these DMs. Accuracy measure is the ratio of true positives (TP) and true negatives(TN) identified by a test (in this case, the testing mechanism is the risk threshold) in a binary condition (risk is 'high' or 'low'). Sensitivity is the ratio of true positives identified by a test (risk threshold) out of the total number of thresholds.

Figures 5.10-5.12 show the thresholds that could be deemed suitable by the two types of DMs. DM working in a fragmented design context would prioritize high accuracy in identifying risky cases (avoid false positives). A DM working in an autonomous design context could be driven by sensitivity in detecting high-risk cases, where he/she prioritizes identification of all high risk cases even if that happens at the cost of a few low risk cases (false alarms or false positives). For all three metrics (sDA, heating demand, cooling demand) the DM working in a fragmented design context would prefer a higher risk threshold (see threshold cut-off lines in figures 5.11-5.13). Such a DM would thus categorize more cases as low risk and expedite decision making in higher number of cases. Table 5.3 shows how often each DM type would evaluate the risk as being high based on the risk thresholds shown in Figures 5.10-5.12

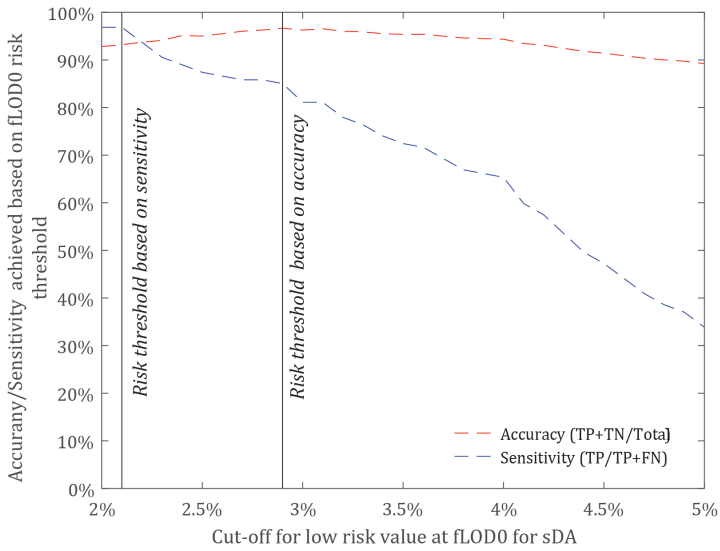


Figure 5.10 – Selection of threshold for high risk at fLOD0 for sDA evaluations based on Accuracy for non-autonomous DM and Sensitivity for autonomous DM (TP=number of true positives, TN = true negatives, FN=False negatives).

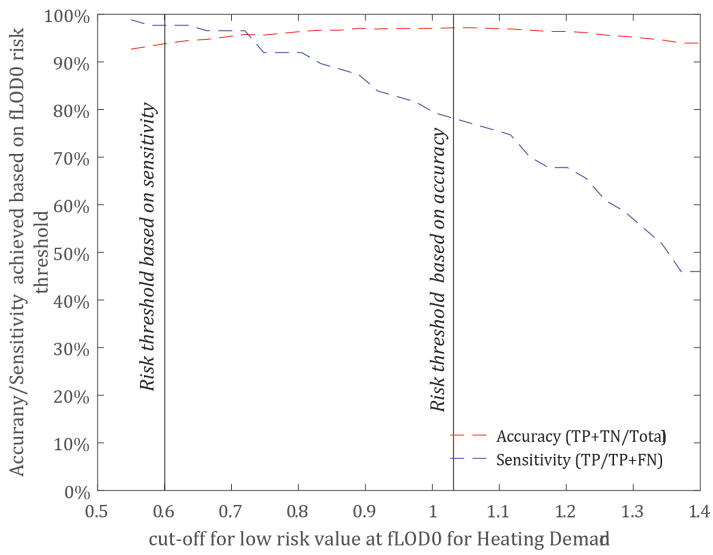


Figure 5.11 – Selection of threshold for high risk at fLOD0 for heating demand evaluations based on Accuracy for non-autonomous DM and Sensitivity for autonomous DM (TP=number of true positives, TN = true negatives, FN=False negatives).

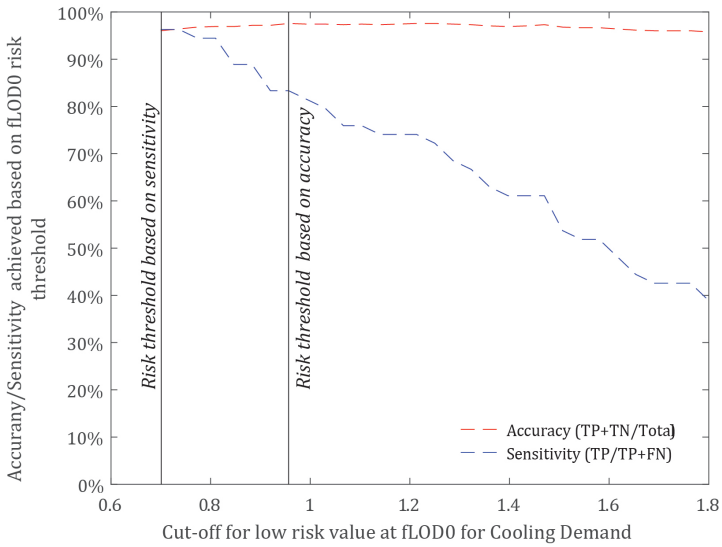


Figure 5.12 – Selection of threshold for high risk at fLOD0 for cooling demand evaluations based on Accuracy for non-autonomous DM and Sensitivity for autonomous DM (TP=number of true positives, TN = true negatives, FN=False negatives).

Table 5.3 – Number of times DMs in autonomous and fragmented design contexts would evaluate risk as being 'high' at fLDO0.

Metric	Autonomous	Fragmented
Daylight (sDA)	1 in 5	1 in 7
Heating Demand	1 in 7	1 in 11
Cooling Demand	1 in 12	1 in 14

5.3 Summary

In this chapter, the need for higher fLODs for decision making at the conceptual design stage was examined. Conditions where higher fLODs may not be needed were also identified. Conditions where the DM's characteristics can impact the need for higher fLODs were also examined.

Relationship of risk to higher fLODs: The reduction of ERPL from one fLOD to the next was found to be comparable to differences in ERPL within a fLOD. ERPL is found to be both design path dependent and fLOD dependent. That is, the value of design details that are yet to decided, depends on the design details specified at the current and prior fLODs. However combining fLODs into 'groups' or 'bundles' could bring a clear improvement to decision making quality from one bundle to another instead of increasing one fLOD at a time.

Increasing fLOD was found to reduce risk and on some design paths an exhaustive specification of all design details (upto fLOD3) was not needed to eliminate risk. In heating demand

evaluations, risk could be eliminated at fLDO2 in several facade design figure 5.3. Greater need for higher fLOD (higher than fLOD0) for conceptual stage decision making was found in daylight and cooling demand assessments, compared to heating demand. Risk was found to be distributed unevenly on various design paths that lead to different facade design scenarios. For example lower risk was observed in design scenarios related to lower WWR (See lowermost branch at fLOD1 in figures 5.2 and 5.3). However, lower risk observed in lower WWR cases should not lead a DM to choose low WWR. WWR and other design decisions should be made based on the performance goals established for the design. The risk metric ERPL is proposed purely as a criterion for selecting model fLOD for choosing between competing choices (massing scheme alternatives) and not for making detailed design decisions per se.

When does a DM need higher fLOD?: The need for higher fLODS was found to be minimal if the performance difference between two massing-scheme proposals is greater that 18.3% on sDA on fLOD0 (number of high risk cases less than 5%). The baseline DM (with dt=10% sDA) can proceed confidently and choose the higher performing scheme. If the performance difference between two design alternatives at fLOD0 is greater than 15.7% on sDA, then the risk of rank-reversals was found to be minimal (irrespective of decision threshold (dt) at fLOD0). Similarly, thresholds were established for performance difference at fLOD0 on heating and cooling demand metrics metrics as well (2.8 kWh/m2-year, 8.3kWh/m2-year for heating and cooling demand respectively). These thresholds on different performance metrics can be used to deliver similar degree of reliability in decision making (less than 5% change of high-risk) and could potentially serve rules-of-thumb for the conceptual stage decision maker.

Effect of DM's preferences and constraints on reliability in decision making: If a DM chooses a highly stringent threshold for decision making (e.g 20% on sDA) he/she heightens the latency effect but can lower the chance of rank reversal. By using a more stringent decision making threshold on sDA (e.g. 20%), the DM ends up not assigning any ranks until a large difference and this is not a very sound approach. A DM who is less sensitive to performance gain observed between design alternatives at fLOD0 regards an alternative better than the other alternative only if substantial difference is found. Such a DM is not necessarily making better (more reliable) decisions as he/she now would regard many pairs of massing schemes (with performance difference at fLOD0 being less than 20%) as being equivalent, and induces risk of latency effect.

The degree of autonomy in decision making was translated into how a DM could potentially draw cut-off values on acceptable risk at fLOD0. As expected, a DM with lower autonomy in case of a fragmented design process results would prefer higher thresholds for risk since he/she has limited powers to take remedial actions in case the risk is found to be high. In 7-8 % comparisons (depending on the performance metric), a DM working in a fragmented design process would regard the risk low enough to ignore where an autonomous DM would regard them high enough to take remedial action (table 5.3). While this result is quite intuitive in nature, the analysis presents the potential impact of fragmentation on the design process when a DM's autonomy in decision making is limited.

6 Managing Projects at Conceptual Design Stage Based on Risk

This thesis has so far suggested a risk evaluation method to assess the reliability of decision making at the conceptual design stage. In section 5.2 examined various predispositions and characteristics of the DM that could affect his/her evaluation of this risk. Once the risk has been evaluated (as 'high' or 'low'), the DM needs to take further action. This chapter examines what those actions could be and how risk evaluation could be used to guide those actions.

In section 2.2.1 several methods were discussed that have been used for ranking design alternatives based on BPS results. In this chapter, the ERPL risk metric is compared to some other commonly used mechanisms for ranking design alternatives in its ability to assist the DM. A case study was used for this purpose, where the objective was to identify themes and issues that could emerge from the use of the ERPL metric for decision making.

6.1 Guiding the decision-maker under high risk

As mentioned earlier, decision making is an action taking problem. That is, decisions imply actions. ERPL risk metric supports robust decision making between conceptual design alternatives, before taking any action. If a robust choice cannot be made (i.e. risk is 'high') then some remedial or supplementary actions must be taken before making a choice between the conceptual stage alternatives. If the risk is found to be 'low', then the DM could make the choice between the massing-schemes at the current level of design development and continue working further until new risky decisions need to be taken.

Madansky [1960] introduced the *"Wait and See"* problem on which the concept of Expected Value of Perfect Information (EVPI) [Raiffa and Schlaifer, 1961] is based. EVPI is used to quantify the value gained from reducing the risk. Building upon the assessment methods for risk, Madansky [1962] offers some useful actions or stratagems to a DM who finds a particular decision too risky:

1. **Delay the decision:** One strategy is to delay the decision until more information be-

comes available. More information implies that some of the prevailaing uncertainties can be eliminated. Reduction in uncertainty, can lead to reduction in risk.

2. **Seek flexibility:** Under uncertainty, flexibility offered by an alternative can also be a valuable asset. For example, flexibility in decision making can be realized if two or more alternatives deliver similar levels of performance on a particular metric. Such a situation offers flexibility in making a design choice based on other metrics. In this case, it would be useful to notify the DM that they can afford to be indifferent towards the choices they have at hand and that his/her choice is immaterial.

3. **Seek dominance:** Su and Tung [2012] translated this strategy by Madansky [1962] as a trigger to redesign and adjust the proposals so that a dominant solution can be found. In the context of this thesis, this strategy is not considered as it does not help the DM make progress but rather asks the DM to re-iterate through the design process. The DM may not find this strategy to be particularly useful or sound if all alternatives at hand are performing satisfactorily and they only need to identify the highest performer confidently.

4. **Re-evaluate risk:** This strategy calls for re-evaluating risk by including the decision maker's outlook for risk. One strategy under this purview suggested by Su and Tung [2012] is to take a pessimistic approach and choose the alternative that performs best in the worst situation. Thus the DM, as suggested by Su and Tung [2012] could refine his/her risk evaluation by identifying outcomes that they are highly averse to.

These strategies suggest that when a decision needs to be taken under uncertainty, reducing uncertainty is just one of the many actions that a DM can take. In the context of conceptual stage design, reduction in uncertainty implies 'fixing' design details that have not been decided upon yet. This requires additional work by the DM. It is thus expected that the DM would find other strategies discussed to be useful if found relevant in a given decision making situation.

6.1.1 Using ERPL to guide DM actions

Next, the ERPL metric's ability to support strategies discussed above is explored. The DM may choose to redesign the massing-schemes as a strategy to mitigate the risk, but the following three strategies (derived from the discussion above) allow the DM to continue working with the design alternative he/she has already produced.

1. **Delaying decision while trying to reduce uncertainty:**

   ERPL metric is intended to identify conditions of high risk and can inform the DM when it is advisable to reduce uncertainty in design features. However it does not tell the DM how to go about reducing uncertainty. Several techniques for efficient ways of reducing design uncertainty in conceptual/early design proposals have been explored recently

in the field of performance driven design (section 2.1.3): a combination of uncertainty analysis and sensitivity analysis is used to create an order to priority in which uncertainty from unknown features must eliminated (or reduced). Some recent studies were also able to produce a dynamic order of priorities that gets updated every time a DM 'fixes' a certain design feature (e.g.[Hester et al., 2017]). Decision making support of this nature cannot be achieved using the ERPL metric given the type uncertainty analysis method used (non-probabilistic) used to calculate ERPL. However the DM can take a trial and error approach to know which design feature to 'fix' first to lower the risk (e.g. [Basbagill et al., 2013]).

2. **Derive flexibility from a decision that is immaterial:**

   If all design alternatives are expected to deliver similar performance on a particular metric, this information may be well received by a DM who is trying to identify a design alternative that is superior on multiple criteria. The DM can then choose the best design alternative based on other metrics. This condition is addressed withing ERPL metric as the risk of *insufficient performance gain*(section 3.2.3) and can be reported to the DM. A significant risk of insufficient gain tells the DM that in a large number of future design scenarios, both design alternatives deliver comparable performance. Instead of feeling compelled to select the design alternative that appears to a higher performer at fLOD0, the DM is free to choose either or choose between the design alternatives based on other metrics.

3. **Refine risk evaluation:**

   As mentioned earlier, if the DM has a more refined understanding of his/her own attitude towards risk, the risk assessment can be revised in anticipation that the revised results will simplify decision making [1]. At the same time, even if the DM expresses an interest in a certain aspect of the distribution of risk (e.g. loss in worst case scenario), ERPL does not support it. More information regarding the extent and distribution of loss needs to be revealed to the DM so that the DM has an opportunity to form an independent point-of-view on risk apart from the ERPL metric.

The ERPL risk metric and the scenario based uncertainty analysis cannot provide information to the DM regarding the most effective way to reduce risk by reducing uncertainty in design features. But ERPL, by reporting loss due to insufficient gains, can inform the DM is he/she can enjoy the flexibility to choose the massing-scheme based on other metrics. ERPL cannot support the DM's interest in comparing the design alternatives under specific scenarios (e.g. worst case scenario). However the underlying loss values, used to calculate the risk, can be revealed to the DM. This could give the DM an opportunity to devise a risk mitigation strategy

[1] An individual's attitude towards risk is measured by economists and social science researchers using specialized surveys or incentivized games. The researchers have argued that survey takers and experimental participants may often not fully understand the complex questions and situations that they are asked to respond to [Dave et al., 2010]. It is thus unrealistic for a typical DM (whether an individual or a group of people making a collective decision) to be self-aware of his/her attitude towards risk.

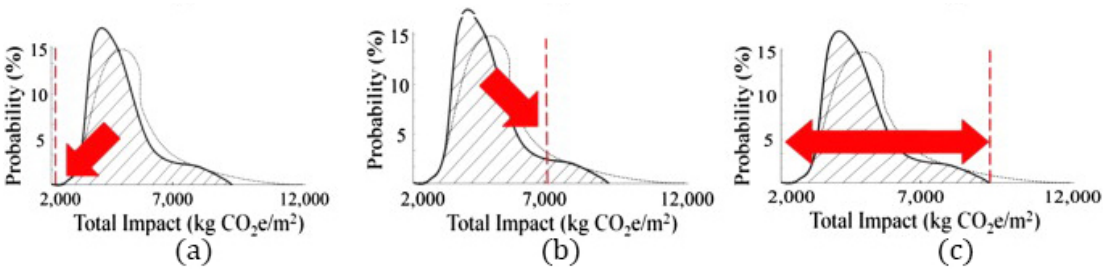


Figure 6.1 – Diagrammatic presentation of three different design goals (a) maximize chances of achieving high performance (b) improve probability of meeting a specific performance threshold (c) maximize design flexibility. Adapted from [Basbagill et al., 2014].

of his/her own. An an illustrative example, Basbagill et al. [2014] present design performance at the early design stage as a histogram and provide the user with feedback using the shape of the histogram (e.g. a single sharp peak versus a flattened one with a large spread). The shape of the histogram becomes especially informative once its shape changes in response to a design decisions made by the DM. For example, if the peak of the histogram shifts closer to the DM's design performance goals, the DM gets to know that the decision he/she just made, supports his/her performance goal. Figure 6.1 shows three kinds of goals that a DM may have (minimizing CO2 emissions, complying with a performance threshold or maximize design flexibility), all supported with the same visualization scheme.

Jusselme et al. [2017] used cluster analysis and then decision trees as a visualization tool to reveal to a DM, in a step-by-step manner, the number of design variants on future design paths that meet design performance targets (Figure 6.2). This visualization supports design flexibility and compliance with a a performance threshold. The design decisions that results in the biggest cluster with the most compliant variants is shown as the next step to the DM. This method is able to guide a DM down a path that at each step provides the greatest design flexibility in future design decision while maintaining performance compliance. This approach lays greater emphasis on design flexibility (flexibility in future design decisions). Data-visualization was further pursued as a possible means of giving the DM an opportunity to explore his/her interest in uncertainty reduction in the context of relative performance evaluations.

6.1.2 Bottom-up action guidance for DMs

Figure 6.3 shows a proposed graphical set up that could be useful to a DM who needs to decide future course of action after encountering high risk. As mentioned above, several strategies could be followed 1) reduce uncertainty to reduce risk 2) determine if the decision is immaterial 3) understand cause of risk to gauge relevance of risk.

The presented graphic (figure 6.3) has two main elements 1) a linear, continuous scale for ERPL that is specific to the metric that the DM is currently using or focusing on, and 2) a

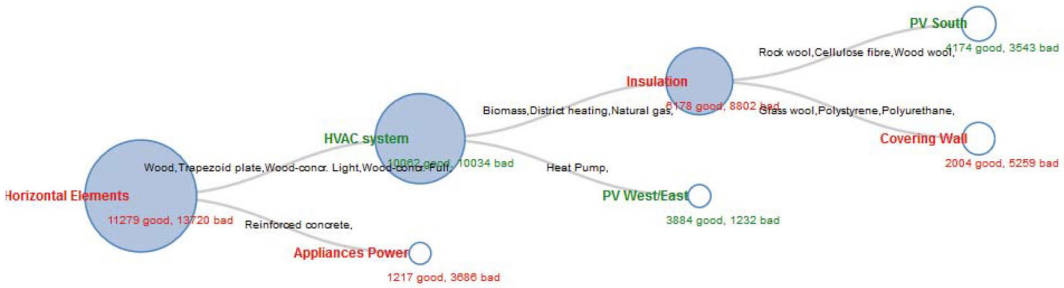


Figure 6.2 – Decision tree used to guide the DM down a path with most design flexibility [Jusselme et al., 2017].

collection of bubbles that show the spectrum (N=192) of possible future relative performance payoffs. The color of each bubble indicates the inference that the DM should draw from it (the bubble). The size of the bubble indicates the amount payoff which is the absolute difference in performance between two design alternatives being compared. The maximum, median and minimum payoff is also indicated using specific markers on the bubbles. The DM is free to focus on either the biggest bubbles (i.e. the biggest payoffs) or on the number of bubbles of the same color of bubbles (i.e. most likely future outcome).

Figure 6.4 shows the steps needed to arrive at the graphic in figure 6.4. The peer-to-peer future facade design-scenario comparison between two massing schemes results 192 values of possible future payoffs. If a payoff value is lower than expected or negative (i.e. the design alternative currently being rejected is better under a given scenario), then the design path leading to such a scenario is risky. Risky design paths are indicated in thick black lines in figure 6.4.

1. **Both schemes comparable:** All future performance payoffs that are smaller than the decision threshold(i.e. negligible for the DM) are shown as gray bubbles. If gray bubbles dominate (large in number), it gives an indication that the DM's choice between the two schemes is of low importance on the given metric. All gray bubbles contribute to the ERPL if the DM had clear ranks between design choices at hand when performance was simulated at the current LOD. Grey bubbles thus indicate potential loss due insufficient gain.

2. **A is better than B:** This indicates all future payoffs that are greater in magnitude than the decision threshold support design A. Irrespective of what design details the DM chooses in the future, if the DM can be sure that his/her choice will stand valid, this notification is given. Also, this is only indicated where the payoff is more than the decision threshold.

3. **B is better than A:** This is the same as the previous point, except it is indicated when design B delivers higher performance compared to A.

4. **fLOD is insufficient:** If the future design paths point towards different design choices

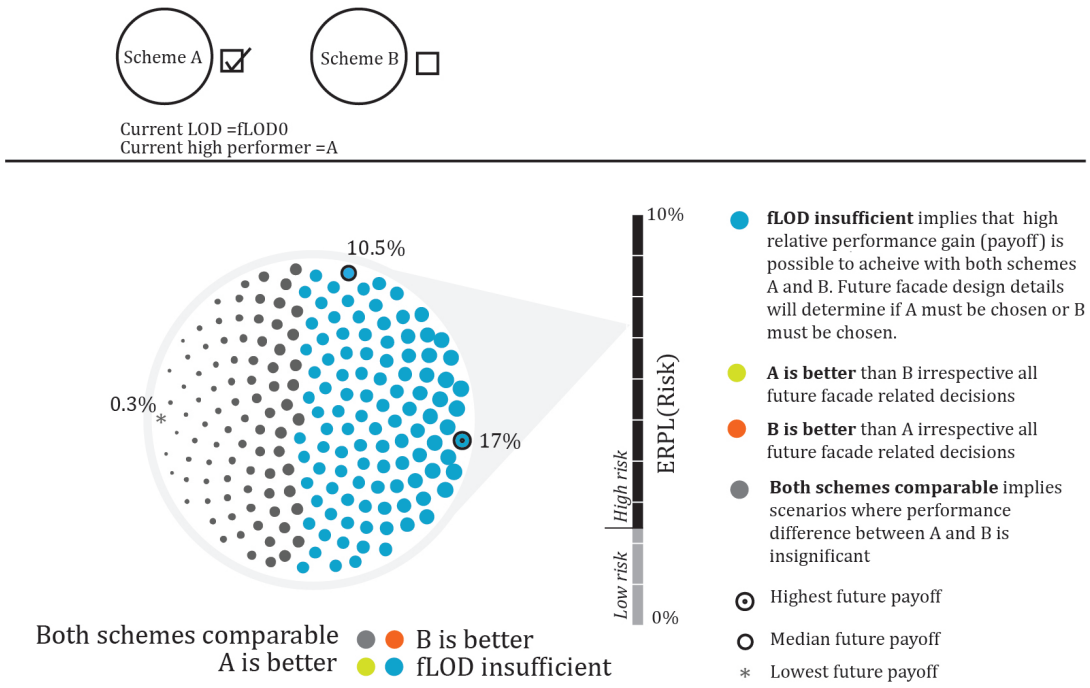


Figure 6.3 – Schematic diagram of visualization scheme showing risk of performance loss (ERPL) and a set of relative future performance outcomes. Each bubble represents a possible pairing of detailed design scenarios. Size of bubble indicates absolute difference between performance of two design alternatives at fLOD3.

(A,B) and a confident choice between (A,B) can only be made once the decision tree has been narrowed down by making more design decisions. In this case, the best course of action is to choose at higher fLOD and blue bubbles are shown for all payoffs whose alignment can only be discovered at higher fLOD.

6.1.3 Using action cues for different strategies

The information shown in figure 6.3 can be used in a few different ways by the DM to suit his/her decision making needs. To illustrate this, three different situations a presented below that a DM could find himself/herself in.

DM-1: DM-1 needs to identify a high performing massing scheme based on sDA. At fLOD0 the DM finds scheme A to be better compared to scheme B by 13.96%. However, in spite of observing a significant performance difference between two design alternatives at fLOD0, the DM finds that the risk of performance loss is high. Figure 6.5 shows the future payoff bubble plot for DM-1. While the ERPL value tells the DM that the risk is high, the bubble plot (figure 6.5 (i)) informs him/her that it is largely due to insufficient gain (gray bubbles). There is potential for rank reversal but is hidden under the blue bubbles. The presence of blue bubbles

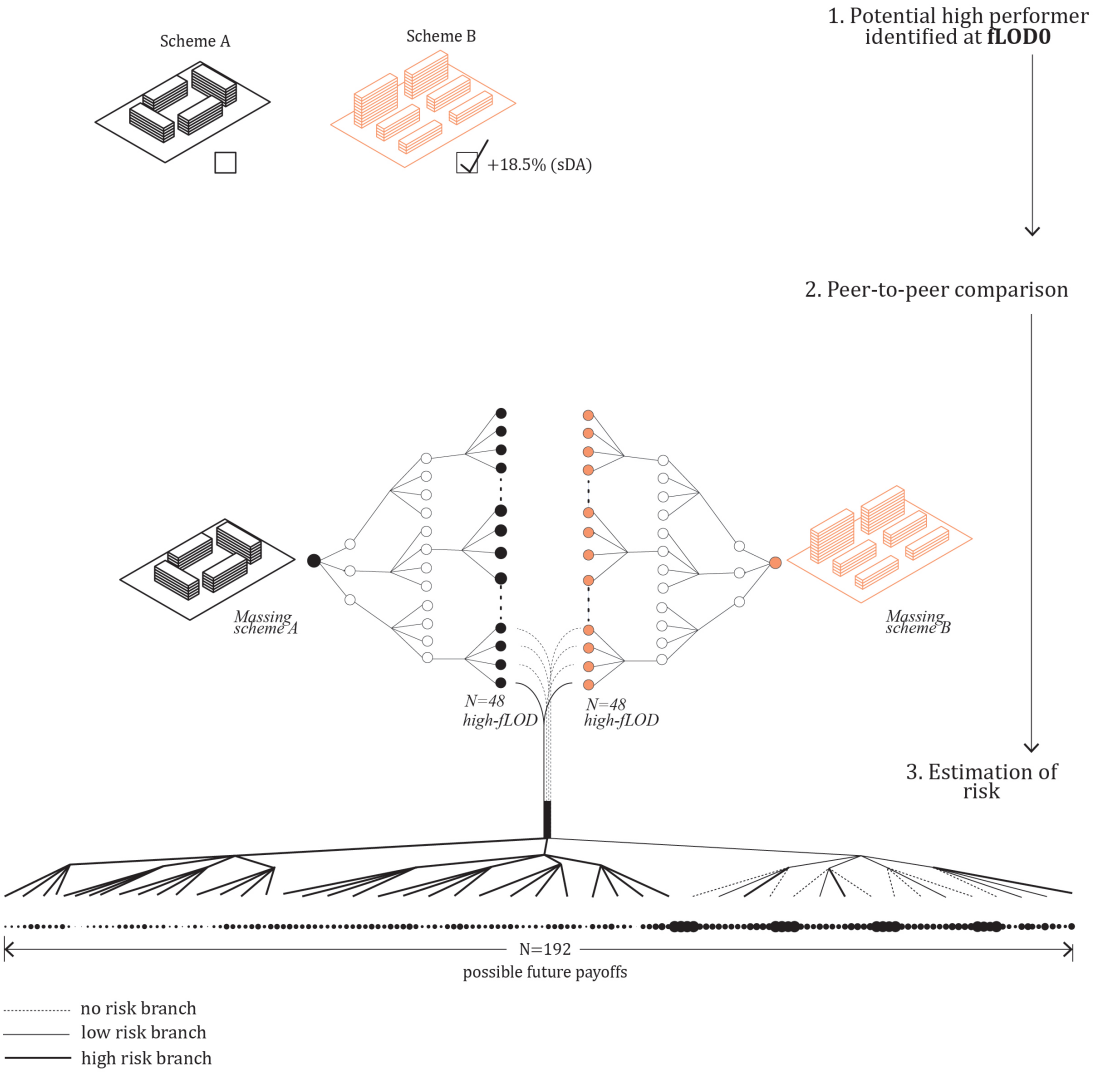


Figure 6.4 – Overall flow of information to arrive at bottom-up action cues 1) preferred design at current LOD is identified 2) peer-to-peer comparison between future design possibilities 3) populating decision tree with future performance difference values (also called future payoffs).

indicates that both A and B can deliver high performance but under different facade designs. DM-1 needs to increase the facade detail to know if A is a better choice or B should be chosen. DM-1 could group the payoffs by various design details to see which facade detail is needed to make a clear choice. Grouping the payoff by WWR results in clarity (figure 6.5 (ii)). By choosing the WWR, the DM can know which massing scheme should be chosen. The reverse path can also be taken by the DM, i.e. if the DM is inclined towards a particular massing scheme. Then he/she can choose the WWR accordingly.

DM-2: This DM also encounters high risk. Here the DM is not able to establish ranks between his/her design choices based on sDA because the difference between the two design alternatives is small (4.56% is less than the DM's decision threshold of 10%). Figure 6.6 shows that this is due to the latency effect (dominance of blue bubbles, indicates high payoffs) where in the DM is unable to establish ranks early on at fLOD0 and it is only at high fLOD that high payoffs can be observed (payoffs upto 19.9% are seen). Once again, at fLOD0 (Figure 6.6 (i)), the DM cannot know which scheme should be chosen as both schemes can deliver high performance. Once the payoffs are sorted by WWR, the DM finds that at 20-30% WWR the best choice is scheme A (Figure 6.6 (ii)). At high WWR, to achieve high payoff, more design details are needed. If the DM chooses 30% WWR further refinement in balcony depth can isolate conditions when scheme A can deliver high payoff.

DM-3 and 4: These DMs would like to resolve conflicts between different metrics and have a specific order of priority among metrics 1) Cooling Demand 2) Heating Demand 3) Daylight (Figure 6.7). Given the DM's order of priority, he/she first examines the future payoff bubble plot based on cooling demand. While the risk (ERPL) is just under the risk threshold, it could still be worth investigating. The bubble plot is dominated by gray bubbles indicating insufficient performance payoffs in the future. Upon grouping payoffs by WWR (figure 6.7 (ii)) and balcony depth (figure 6.7 (iii)) the DM-3 decides to investigate payoffs on daylight and heating demand metrics with 30%WWR and shallow balconies. He/she finds that with moderate amount of glazing and shallow balconies, both design alternatives can deliver comparable performance on daylight and heading demand metrics. Another DM (DM-4) in the same situation is interested in higher glazing ratio and deeper balconies. He/she finds that scheme B is more suitable (large number of orange bubbles) if cooling demand evaluation has higher priority and high WWR (40%) is desired but also implies a compromised choice on the daylight metric. With the facade design details preferred by DM-4, on the daylight metric, scheme A would be a better choice.

6.1.4 Discussion

In this section a graphical set-up was proposed for assisting a DM relying on risk metric ERPL to make conceptual stage design decisions. The ERPL metric is meant for a specific context of use that can prompt a DM that certain added measures are needed (e.g. increasing fLOD). However, ERPL is silent on what the DM should do next in case risk is high.

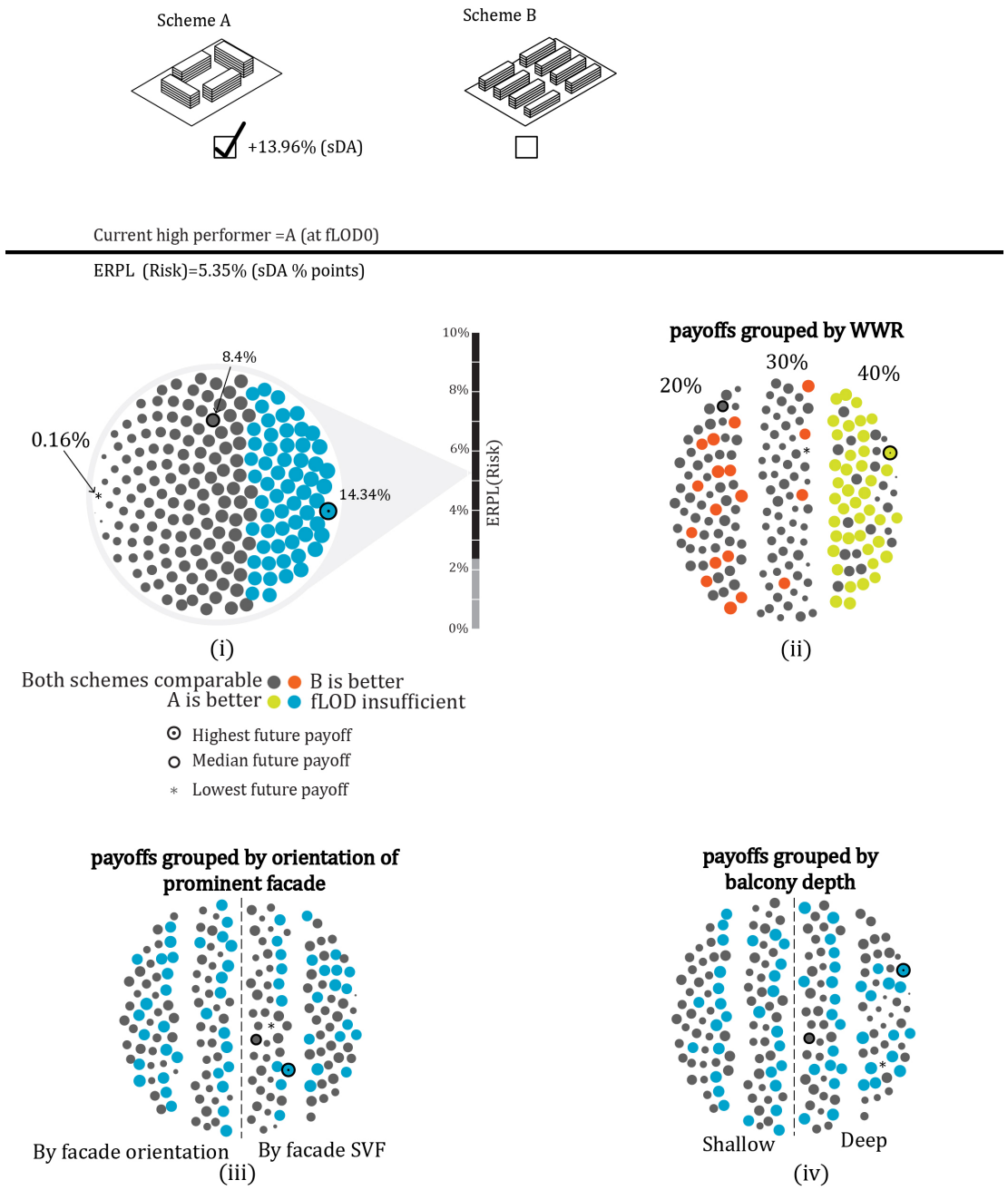


Figure 6.5 – DM-1 browsing through the future payoffs bubble plot to know if he/she would continue being in an advantageous position with his/her current preferred massing scheme (Scheme A) under various possible facade related decisions.

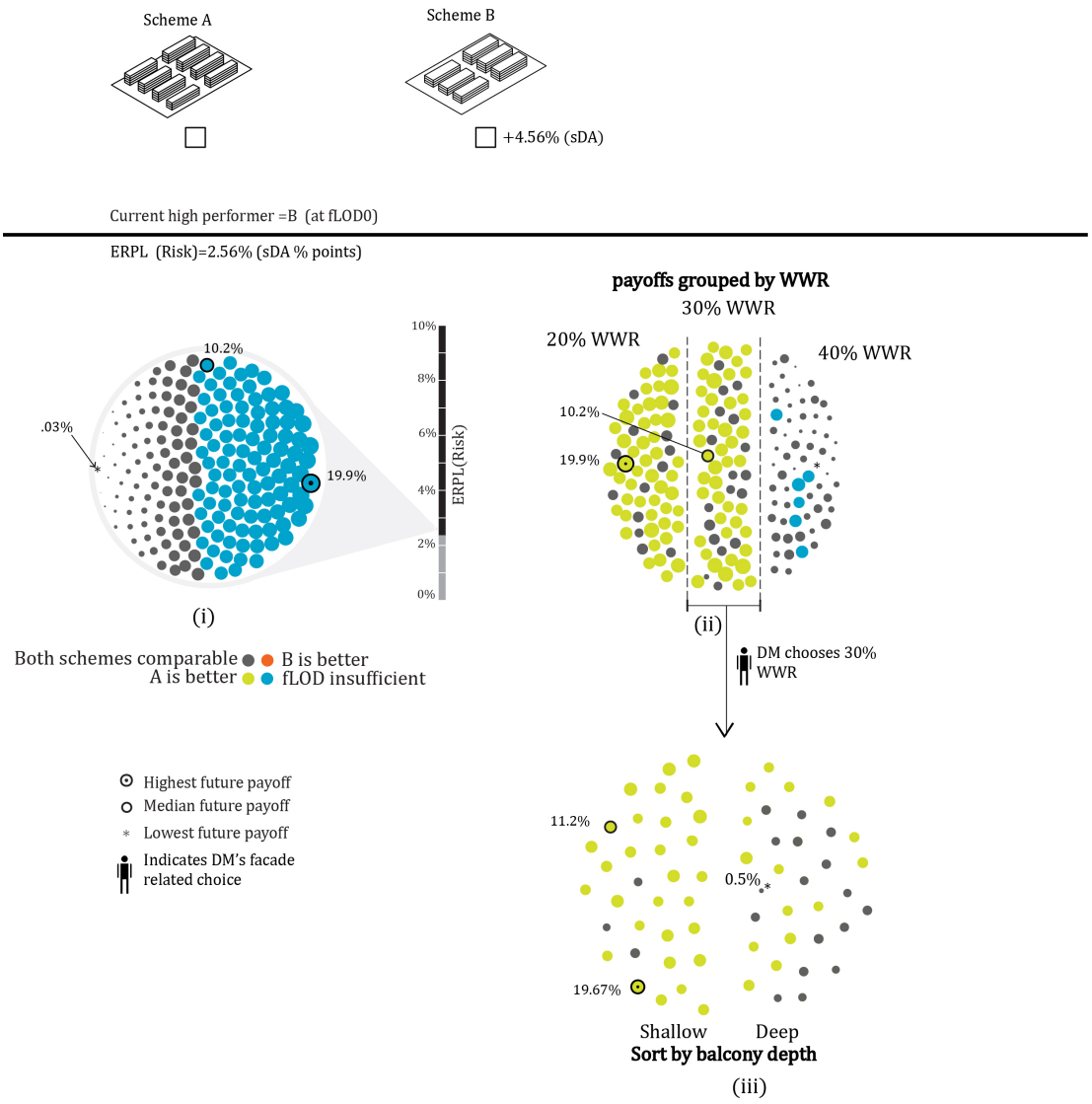


Figure 6.6 – Example showing DM-2 who is unable to assign ranks at fLOD0 due to low performance difference observed between design alternatives. He/she increases fLOD to isolate conditions where a confident choice can be made.

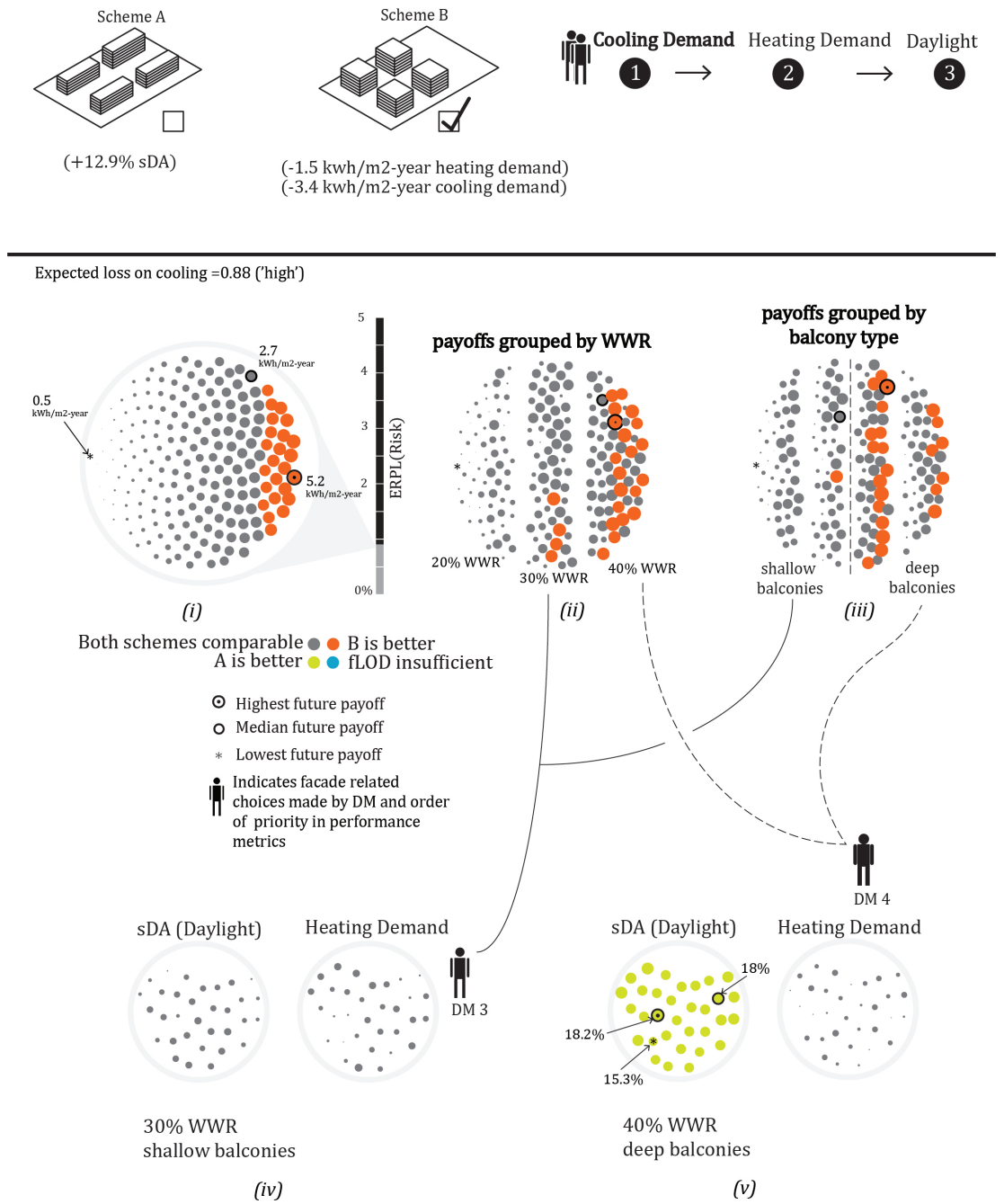


Figure 6.7 – Example showing DM-3 and DM-4 who are working in a multi-criteria decision making environment. DM-3 is able to resolve conflict in choice through his/facade related choices while DM-4 is not.

The bubble plot based graphical scheme helps the DM extend his/her understanding of risk beyond the ERPL value. Other design performance visualization schemes ([Basbagill et al., 2014], [Jusselme et al., 2017]) have been discussed above that are meant to guide a DM who is making design decisions under uncertainty while pursuing absolute performance goals for a single design alternative. The ERPL metric and the related visualization proposed in this chapter is meant for relative performance evaluations. In the proposed visualization scheme, to reduce uncertainty and risk, influential facade design features(s) can be identified by grouping the bubble plot in different ways. Thus, based on the nature of distribution of the bubbles when grouped by a design feature, the DM can know if that particular design feature can help in making a clear choice between massing-schemes.

The visualization scheme allows the DM to focus on one metric (e.g sDA, heating demand) at a time and leaves the DM free to decide whether he/she wants to focus on achieving high payoffs (bigger bubbles) or achieving a certain nature of outcome (outcome indicated by color). The visualization also informs the DM if his/her choice matters which is useful in a multi-criteria decision making environment.

The ERPL metric and the graphical-set up are meant to assist the DM in choosing the superior massing-scheme in a robust manner while committing to the fewest possible facade details. However, the DM must keep in mind that this analysis does not account for any absolute performance thresholds (e.g. 75% sDA) that the project may need to achieve. Thus facade decisions must not be used purely as a mechanism to reduce risk. Any absolute performance goals designated for the project must also be examined. For example, in the case of DM-2 discussed above, if the DM chooses 20 to 30% WWR, his/her choice of massing scheme A is justified. But in that case, can the absolute performance goals be realized even though the superior scheme for that range of WWR is chosen? Such situations could be found in many evaluations. Findings of this thesis reported in the previous chapter show that risk tends to be lower at low WWR (see figures 5.2, 5.3). However, that is not a sound justification to choose low WWR. Thus the use of **ERPL and the graphical set up is best suited for prioritizing actions** (e.g. increase FLOD by deciding WWR), rather than making specific design decisions (e.g. WWR needs to be 20%).

6.2 Risk as a robustness measure for ranking design alternatives

Several methods for robust ranking of design proposals were discussed in section 2.2.2. Different robustness methods have been found to result in different ranking of design alternatives under uncertainty in BPS assessments (e.g. [Rysanek and Choudhary, 2013], [Nikolaidou et al., 2017]). To resolve such a condition, the DM could choose a robustness method that is best suited to his/her characteristics such as risk-aversion. However, it can be difficult for a DM to identify his her preferences towards risk (see discussion in section 6.1.1). The ability of various methods to fill potential informational needs that a DM are compared.

Four types of robustness methods were discussed in section 2.2.2 that my be used to assess

robustness in ranking of design alternatives, namely, expected value methods [Wald, 1950], higher-moment methods, regret based methods and satisficing methods. Expected value methods [Wald, 1950] and higher-moment methods were selected for further comparison to ERPL.

Other regret based methods such as Opportunity Loss [Savage, 1951] and Expected Opportunity Loss (EOL) [Su and Tung, 2012] have partial methodological redundancies with the ERPL metric. Expected opportunity loss (EOL) is mathematically the same as the risk of rank-reversal incorporated within ERPL. Satisficing methods were not used in this case study as ERPL metric is meant exclusively for relative performance evaluations and not for comparisons to a specific performance level.

In this section ERPL is discussed in terms of its three constituents to exploit its full potential in informing the DM regarding future performance of the design alternatives, with the three parts being: 1) **risk of rank-reversal** (same as EOL) 2) **risk of latency effect** which helps the DM know that two design alternatives which appear comparable, may not remain so once more design details are developed 3) **risk of insufficient performance gain** which lets the DM know of the likelihood that the design alternatives will deliver similar performance even though they appear different in conceptual stage evaluations.

6.2.1 Choosing a winning design competition entry: a case study

The ERPL metric is meant to inform a DM regarding risk in relying on ranks observed from performance evaluations at fLOD0. Thus ERPL reports the risk in a given choice (e.g. (A*,B)) or ranks. The highest incidence rate of such risk was found in sDA based evaluations (1 out 5 found to be high risk). Thus a DM looking to make 5 relative performance comparisons (between pairs of design alternatives) or more can expect to encounter high risk comparisons.

Design competitions are a popular means of soliciting design solutions and often results in a large number of design proposals being generated for a given design problem. Also, a competition at a large neighborhood scale or urban scale could be conducted in multiple stages, where conceptual proposals are invited in the first stage. The winner(s) is/are chosen based on the potential observed in the conceptual level proposals. Such a design development process provides an opportunity to examine the efficacy of risk-based performance evaluation, when future design development paths are not available for inclusion in performance evaluations.

Existing example projects where 4 or more design alternatives had been proposed were searched for, so that 6 pair-wise (4C2=6) combinations would be needed to rank them. This would heighten the likelihood of finding high ERPL in atleast 1 or more comparisons on atleast the sDA metric. More number of design alternatives would be needed to find ERPL on heating and cooling demand metrics (10-15 design alternatives).

The city of Nidau and Bienne, Switzerland initiated such a design competition in 2013 [2] in order to revive a water front area of the city which had been an industrial site. Town planning designs were required to be developed on a site of 312,300 m2 with a target built area of 130,000 m2. The selection criteria for this project were (a) urban and landscape qualities, (b) profitability or economic viability and (c) sustainable development (environment and social aspects). 25 teams responded to this competition. The jury identified eight noteworthy design proposals, out which five were short-listed for further detailed design development. A preliminary winner (CiteLac scheme in Figure 6.8) was also announced and the design team received a prize award of 100,000 CHF. The second round of competition between the five finalists was completed in 2019.

As part of this thesis, 3D models were developed for the eight most promising designs identified by the jury [3]. Evaluating the design proposals for the whole site was not considered feasible. For comparison, the massing-schemes evaluated earlier (Figure 3.8) had a total built area of 15,000 m2 each, while proposals for this competition were 130,000 m2. Generating and handling the geometry at this scale at the desired level of detail would have required a significant shift in the modeling workflow developed earlier (section 3.2.2). A specific parcel of land was identified within the competition site from which neighborhood scale design proposals could be carved out. Four (out of eight) design proposals were found in which a clear and identical site boundary could be drawn to create neighborhood scale design proposals. Within this site some adjustments were made to harmonize the built density within the neighborhood for a more equitable comparison (built density adjusted to 1.73 which was the average density for the four proposals). Figure 6.8 shows the original proposal layouts. Figure 6.9 shows interpretation of these proposals after limiting the area of evaluation to a single parcel and fixed density of 1.73.

6.2.2 Establishing ranks using risk based assessment

This section further builds on the four design proposals identified for further analysis in this case study. First, BPS models were developed using the inputs which are typical for the conceptual design stage evaluation (assumed to be commensurate with fLOD0 BPS models, see 3.6 and Appendix A.3). The four proposal are evaluated and ranked on three performance metrics (a) sDA (b) annual heating demand (c) annual cooling demand.

Using typical conceptual design stage BPS model inputs (corresponding to fLOD0), the four design proposals were evaluated and simply arranged in order with the aim of ranking them. Figure 6.10 shows that performance results obtained for the four massing schemes on the three performance metrics. The performance difference between the various design alternatives is small.

[2] https://www.agglolac.ch/

[3] Plans and elevations were available on the competition website. Some examples massing schemes for this project were also used in Chatzipoulka et al. [2018]

Figure 6.8 – Selected competition entries for the vision AGGLOlac competition for the city of Bienne, Switzerland.

For example on sDA metric the performance difference between the top three design alternatives is 6.4%. The difference between Strandboden (ranked 2 as per fLOD0 models) and Marais scheme (ranked 3 as per fLOD0 models) is less than 1% on sDA. These difference are much smaller than performance differences that a DM would potentially need (refer to table 3.7 for possible relative performance difference thresholds) in order to rank these alternatives. The difference between the best and worst performer was found to be 1.06 kWh/m2-year on heating demand and 2.80 kWh/m2-year on cooling demand. These performance difference values for the massing schemes being evaluated here are not likely to help DM as no scheme emerges as the dominant one on all metrics. Citelac scheme, which is the best performer on sDA and heating demand, is 3rd on cooling demand. Also it is not known how stable any of these rankings are, given the nascent stage of design.

Next, I shall examine if adding more design details can help establish ranks between these design alternatives more confidently. Several design scenarios were developed for each massing-scheme at high level of design detail (fLOD3) as described in section 3.2.2. These scenarios represent design possibilities that could arise if the design process was allowed to progress on all design schemes.

Figure 6.11 shows one (out of 48) high fLOD (fLOD3) design scenarios that were generated for each of the massing schemes. Figure 6.12 shows a detailed view of two high fLOD scenarios each resulting from massing schemes CiteLac and Laridae. Three values of WWR (20%, 30%, 40%) and four possible combinations in selection of a prominent facade are used to build facade design scenarios. As defined earlier (section 3.2.2) choice of prominent facade is based on orientation or on the sky view factor (SVF) of the facade. Balconies of varying depth are then attached to the prominent facade. All fLDO3 variants also include active shading devices.

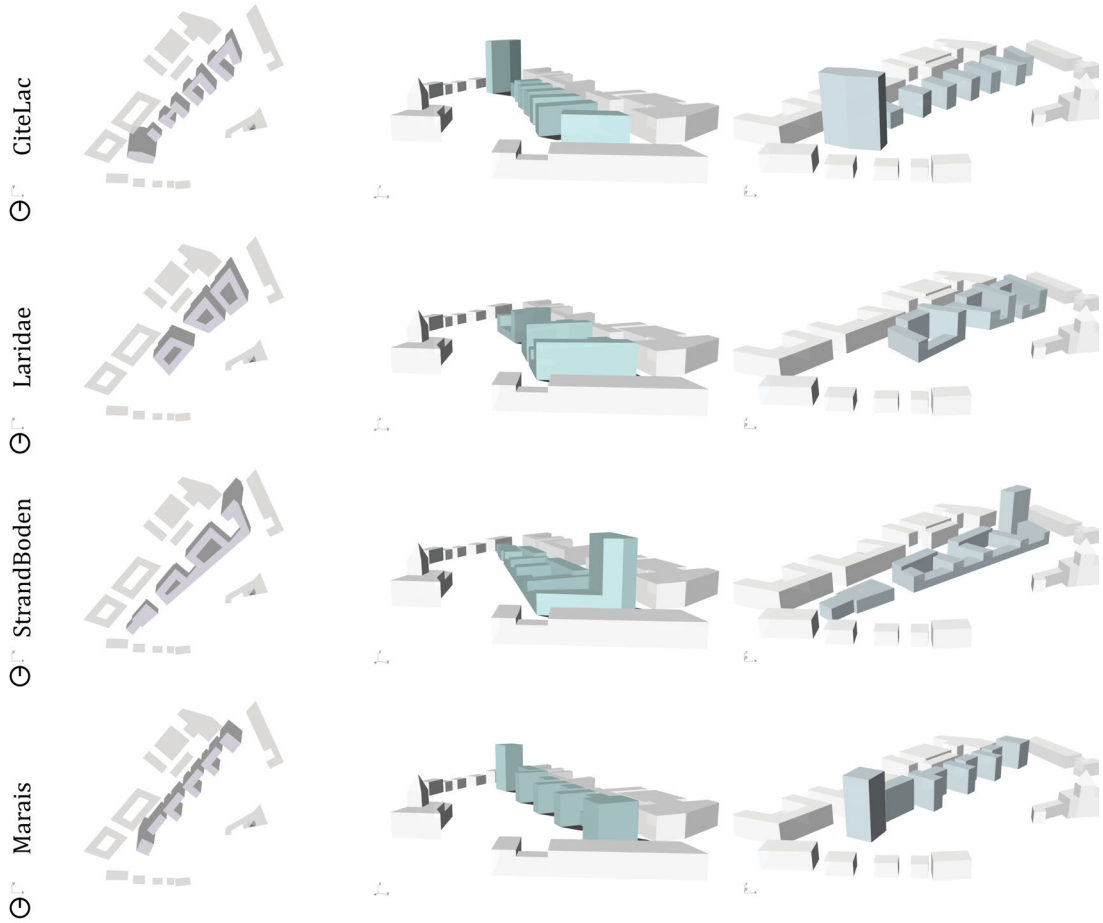


Figure 6.9 – Selected parcel out of the competition entries, all brought to the same development density of 1.73.

Figure 6.13 presents box-plots depicting the performance values that can be anticipated in the massing designs at fLOD3. On sDA, as seen in results obtained at fLOD0, the mean and median values of the top three performers are found to be between 1-2% of each other. Marais scheme is found to be best performer on heating demand. However, it shows negligible performance improvement over its competitors (improvement of 1.4 kWh/m2-year over worst performer) if the mean performance is compared at fLOD3. Here the DM may conclude that heating demand is not an important criteria for assigning ranks. The average performance on cooling demand is also found to be quite similar (difference between mean of best and worst design is 0.8 kWh/m2-year in annual cooling demand). The DM may also decide to exclude cooling demand as a criteria for ranking the design alternatives.

Figure 6.14 shows the evolution of performance of all the four variants from fLDO0 to fLOD3,

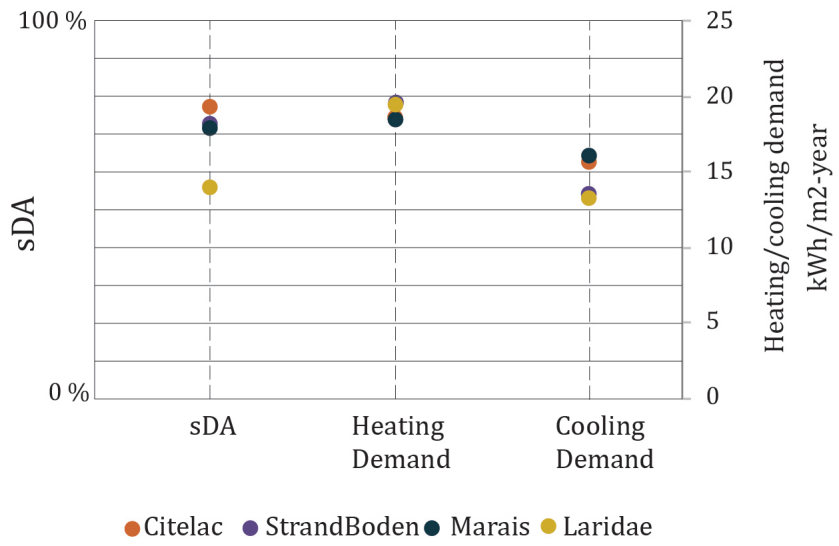


Figure 6.10 – Ranks of design proposals in the first stage of the design competition.

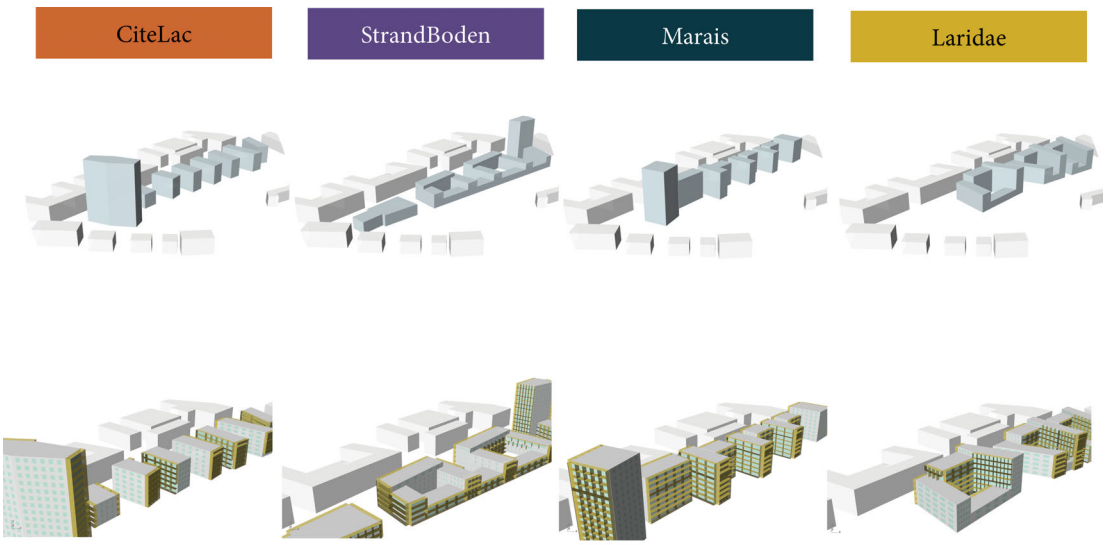


Figure 6.11 – One out of 48 high fLOD variant shown for each massing-scheme.

also showing the performance at intermediate levels of detail. This plot is useful understanding 1) the overall trend of performance values (from low fLDO to high fLOD) 2) degree of overlap in performance among the design alternatives and 3) the range in the performance values of each design alternative at fLOD3 (although the range of performance values of each design alternatives is better understood though the box plots). From modeling the design alternatives at fLOD3, and simply plotting the data, it may appear to the DM that heating and cooling

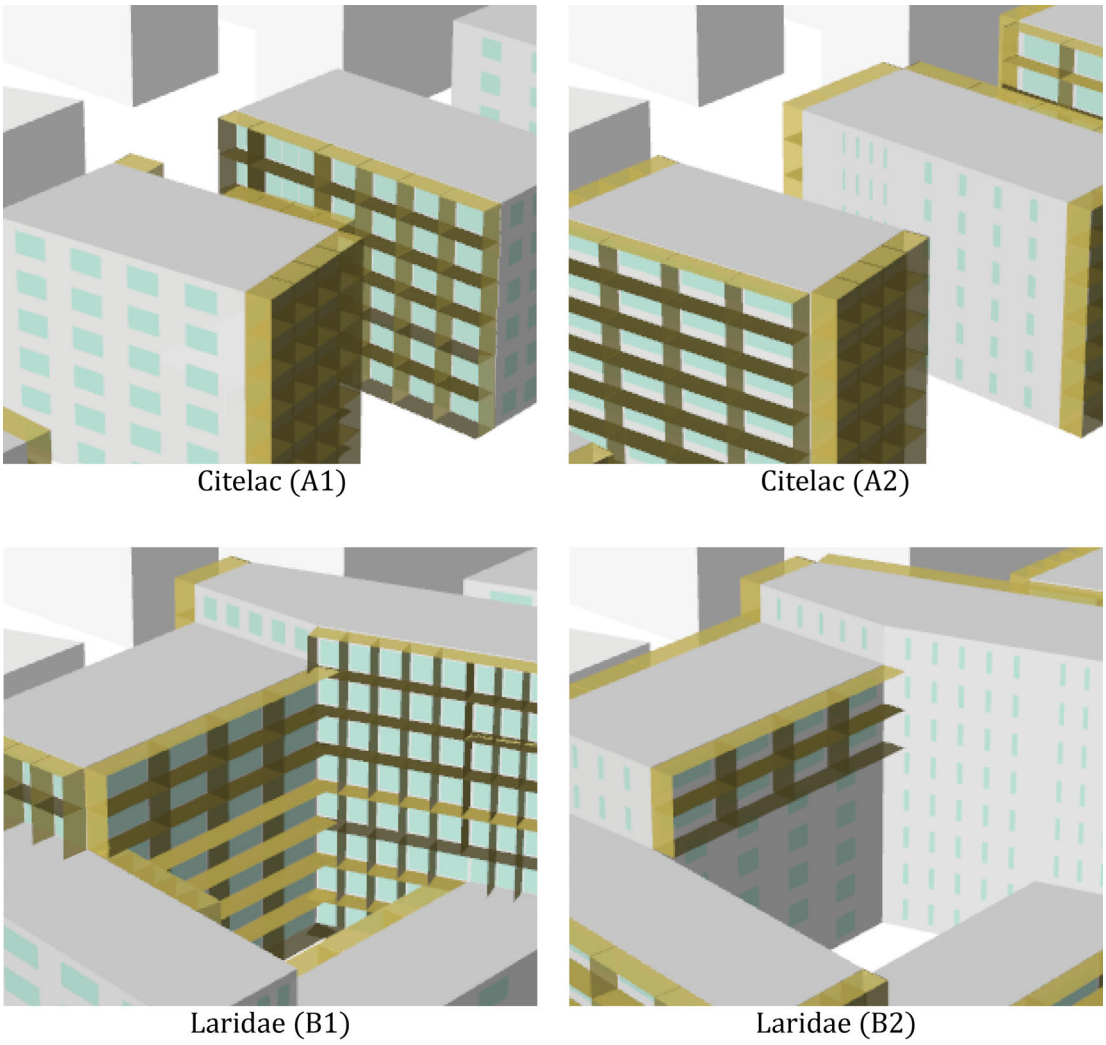


Figure 6.12 – Two out of 48 fLOD3 variants for Citelac and Laridae massing-schemes.

demand are not important criteria for ranking the design alternatives. On sDA, the design with the least performance potential appears to be Laridae. It may appear that between the other three (Citelac, Strandboden and Marais) the DM could pick any scheme if his/her minimum differentiation criteria is 10% on sDA (common industry practice, see table 3.7).

Table 6.1 shows the performance for all the design alternatives on sDA at low fLOD (section (i) of table) and at high fLOD (sections (ii) and (iii) of the table). Section (i) of table 6.1 thus shows the information would be typically available to a DM at the conceptual design stage. This data was also shown in figure 6.10. Section (ii and iii) shows information extracted after modeling the massing schemes at fLOD3. Section (ii) shows simple data summarization methods (mean, median, coefficient of variance) and expected value metrics (Maximin, maximax and the Hurvicz criterion). Section (iii) shows the risk of rank-reversal and risk of latency effect

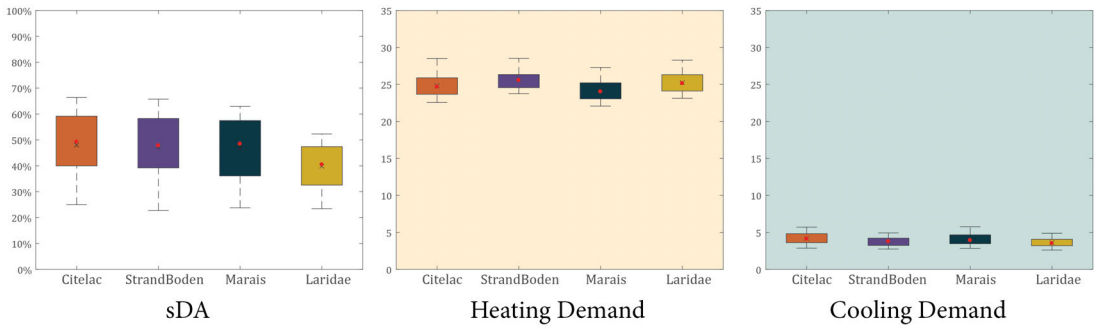


Figure 6.13 – Boxplots showing performance values for each scheme at high fLOD (fLOD3).

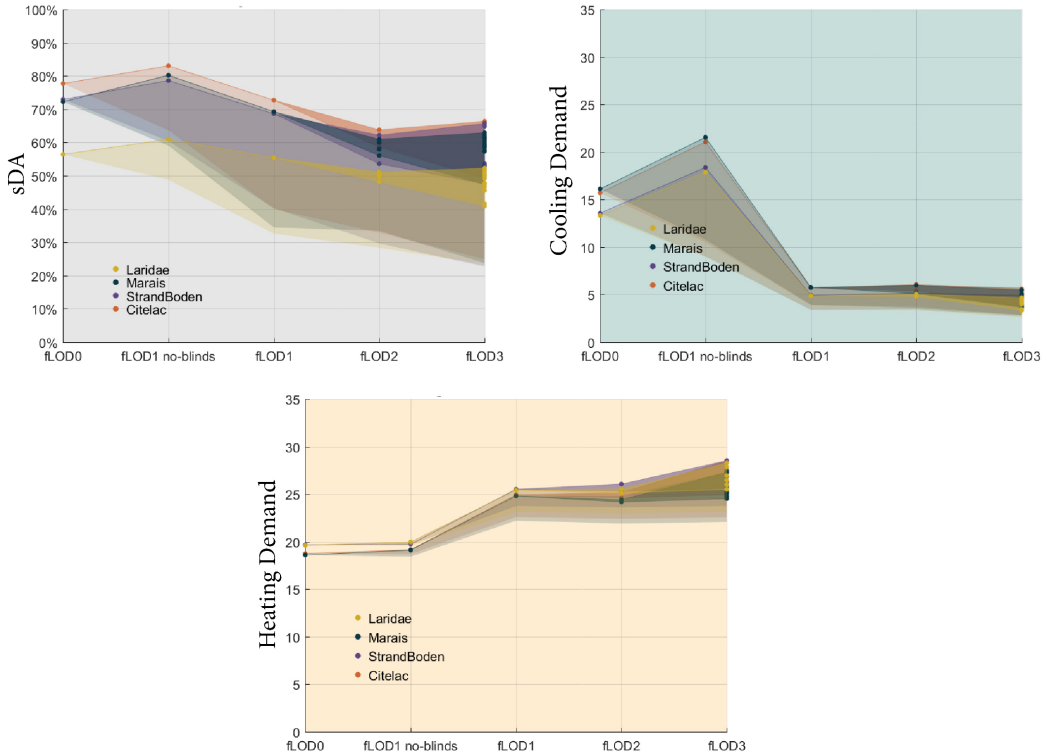


Figure 6.14 – Evolution of performance evaluation with façade level of detail.

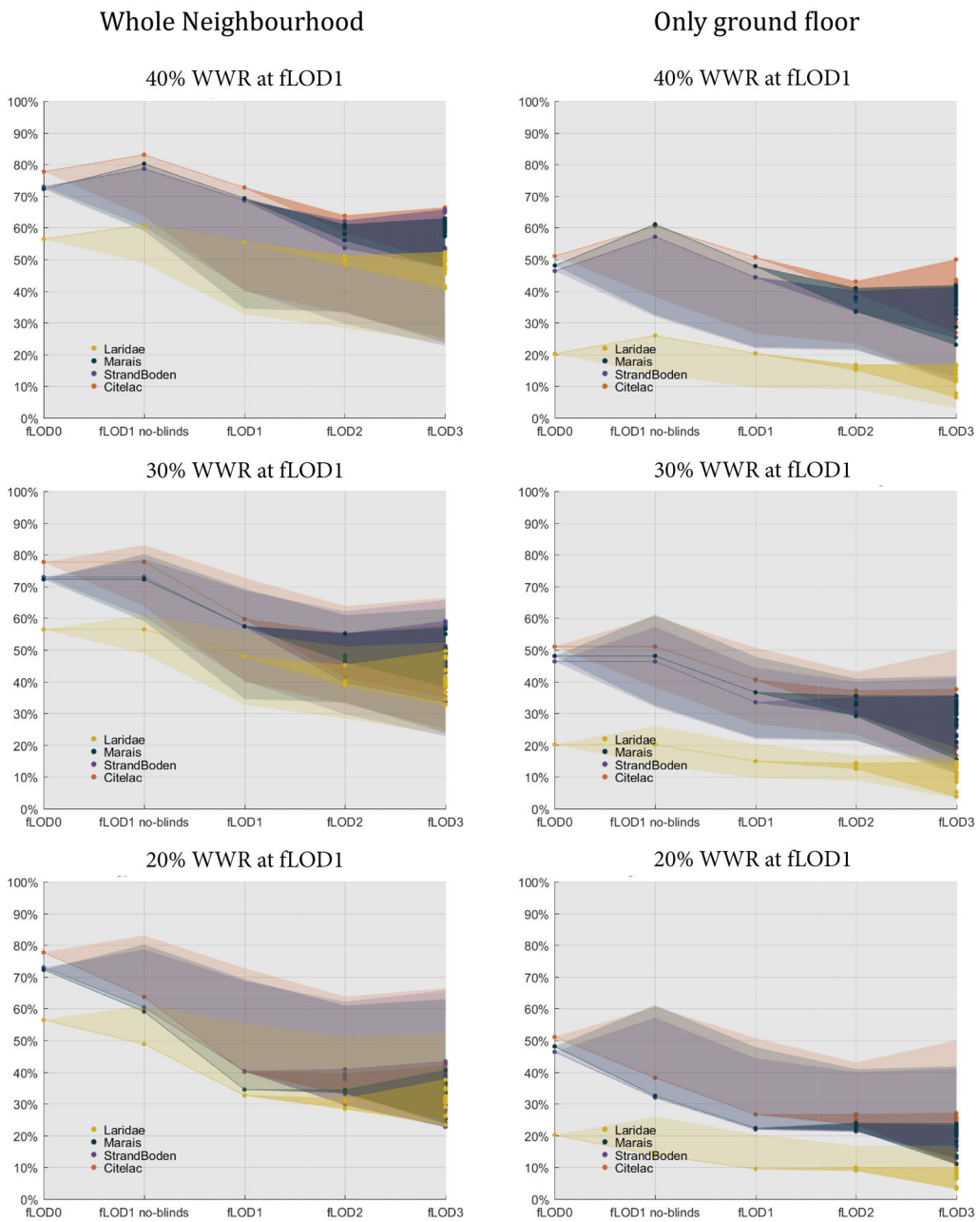


Figure 6.15 – Performance data summary for four massing scheme design proposals for AGGLOlac project.

(delayed discovery of performance difference). The risk shown is the risk in choosing Citelac over the others (since Citelac was the highest performer at fLOD0).

In section (ii) the Citelac massing scheme is found to perform the best on every statistical descriptor except on the coefficient of variance (CV) parameter. However, consistent ranking between Strandboden and Marais for the second and third position cannot be established based on criteria shown in the section (ii) of table 6.1. Also the difference between Strandboden and Marais on all criterion mentioned in section (ii) of the table is quite small (1-2% on sDA). The conclusion that could be drawn based on the data provided in section (ii) could be to assign equivalent ranks to StrandBoden and Marais and more evaluation criteria be added to make a definitive choice.

Risk based evaluations are shown in section (iii) of table 6.1. The risk measures take a comprehensive view of the payoffs (peer-to-peer comparisons between all scenarios, not just best and worst scenario) and can thus be considered a better representation of the possible outcomes of the relative comparisons than those considered above. For example, a small risk of insufficient gain is found between Citelac (highest ranking) and Laridae (the worst performing). This could not have been anticipated from any of the values shown in section (i) and (ii). Note that this risk is actually negligible.

With this information, the deliberation on ranks could be summarized as follows:

- **Should Citelac be the top ranked scheme given its high performance at fLOD0?:** The risk of latency effect between Citelac and StrandBoden is 0.94%. This means that there is risk that the performance difference between these two scheme will grow and some scenarios were found where the performance difference between them exceeds the threshold of significance (10% difference on sDA). If the question is whether Citelac should be assigned the top rank, then the presence of latency effect is not a 'threat' per se. It supports choosing Citelac over Strandboden. Between Citelac and Marais, risk of latency is smaller and so Citelac and Marias can be expected to deliver comparable performance in more scenarios at fLOD3, but Citelac has a small chance of delivering significantly (10% or higher difference on sDA) higher performance than Marais.

  On heating and cooling demand, the latency effect risk was found to be zero. Thus, the difference between all peer-to-peer comparisons were found to be less than the performance difference threshold, both at low and high fLOD, for heating and cooling demand evaluations.

  If unique ranks have to be assigned, then Citelac could be chosen although the latency effect is small (lower than risk threshold of 2.1% table 4.2) so possibility of it delivering higher performance is low.

- **Should multiple design teams be invited for detailed design development of their proposals?:** Further testing was done to see whether greater distinction could be observed between all the massing scheme proposals if the sDA was calculated only for

Table 6.1 – sDA for the four competing design alternatives at low and high fLOD. **Section (i)** shows the performance for all the design alternatives on sDA at fLOD0 that would typically be available to a DM at the conceptual design stage, **section (ii)** shows simple data summary of performance values at fLOD3, **section (iii)** show risk of performance loss if Citelac is chosen as the winner

		Citelac	**StrandBoden**	**Marais**	**Laridae**
Section (i)	sDA at fLOD0	77.7%	73.0%	72.3%	56.5%
Section (ii)	Mean expected sDA	48.0%	47.1%	46.8%	39.9%
	Median sDA	49.2%	47.8%	48.5%	40.4%
	Coefficient of variation (CV)	0.24	0.23	0.24	0.20
	Worst case*	25.0%	22.8%	23.8%	23.4%
	Best case**	66.4%	65.7%	63.0%	52.3%
	Hurwicz Criterion #	45.7%	44.2%	43.4%	37.9%
Section (iii)	Risk (Rank Reversal)	NA	0.00%	0.00%	0.00%
	Risk (Latency Effect)	NA	0.94%	0.44%	0.00%
	Risk (Insufficient Gain)	NA	0.02%	0.01%	0.01%

* Minimum performance delivered by design alternative. Maximin criterion for pessimistic DM

** Maximum performance delivered by design alternative. Maximax criterion for optimistic DM

Hurwicz criterion (α=0.5)

the ground floors. The ground floors would receive the least amount of light and thus would be potentially most sensitive to differences in the designs. Figure 6.15 shows the performance evolution for the ground floors zones of each massing scheme on the sDA metric. It was found that limiting the sDA calculation to the ground floor only increased the performance difference between the high performance group (Citelac, Strandboden, Marais) and the lowest performer, Laridae. The small risk of insufficient gain between Citelac and other schemes observed earlier (table 6.1) was now eliminated because of the increased performance difference between them. Negligible latency effect risks were again found when comparing Citelac, StrandBoden and Marais. This indicated that these are indeed close competitors and performance difference between them can only be known once more detailed designs are proposed.

6.2.3 Synthesis

In this section, the case of a design competition was presented where the selection of design solutions had to be narrowed down from a set of four design entries based on three performance criteria. The DM was assumed to have the same decision threshold values as

mentioned earlier in chapter 3 (table 3.7). The risk assessment methodology allows for robust decision making by taking into account future design development scenarios, along with rigorous comparisons between future design development scenarios, making sure that the chosen design can continue on to deliver superior performance later on in the design process.

In this case study, conceptual stage design choices needed to be narrowed down (first phase of the competition) before further detailed design development can occur (second phase). In such cases, risk assessment and its visualization can rescue the decision maker from acting on simplistic information that cannot support nuanced decision making needs. In this case-study, uncertainty quantification was found useful in confidently eliminating heating and cooling demand performance as decision criteria. On sDA, the risk assessment was useful in confirming that the top three performing forms are indeed close competitors. If sDA, heating and cooling demand were three criteria for choosing, then design teams for Citelac, Standboden and Marais schemes should all be invited for developing detailed design proposals. While a similar conclusions could have been drawn from nearly all other robustness measures that were calculated as part of this case study, a peer-to-peer comparison between all future scenarios is only possible under a regret based robustness measure.

As observed in chapter 5, the larger the difference in performance observed between the design alternatives at fLOD0, the lower the risk in decision making. Any techniques that can magnify the differences between design alternatives is likley to be helpful in identifying a clear winner. In this case-study, while limiting daylight performance assessment to ground floors did assist in increasing the the performance difference between the lowest performer and the other design alternatives, it did not help further differentiate between the set of top performers. This presents further support to the conclusion that design teams for Citelac, Standboden and Marais schemes should all be invited for further developing detailed design proposals.

7 Conclusions

The main goal of this thesis was to asses the reliability of BPS-based design decisions made at the conceptual stage of a project. Conceptual stage BPS models suffer from information deficiency and if uncertainty in future design decisions is ignored, then conceptual stage decision making can also suffer. In this thesis this problem was addressed at the neighborhood scale in the context of a typical conceptual stage decision making task - choosing one design from multiple alternatives through pair-wise comparisons.

7.1 Achievements and contribution

To address the goals of this thesis, it was first essential to identify an appropriate method for assessing and reporting reliability of the performance evaluations to the conceptual stage decision maker (DM). Several methods can be used for ranking design alternatives under uncertainty. Methods based on regret/opportunity loss based methods were found most pertinent to the problem of choosing one alternative over others. Regret estimates the loss from choosing an alternative, which could be rendered sub-optimal once important uncertainties in detailed design features are overcome. However, at the end of design evolution (where uncertain design features become known), the design possibilities are virtually unlimited. Under such conditions, a probabilistic metric of loss (i.e. risk) was found better suited. The risk metric Expected Opportunity Loss (EOL) [Su and Tung, 2012] was further built upon to propose a risk metric for Expected Relative Performance Loss (ERPL) in conceptual stage design decisions.

Risk assessment is common practice in several disciplines related to building design (e.g. structural design). However reporting confidence (or risk) in performance results is virtually absent in use of BPS tools [Clarke and Hensen, 2015]. The proposed ERPL metric reports the risk in relying on ranks assigned to design alternatives based on performance evaluations from conceptual stage BPS models. The risk is thus reported subject to the level of detail (specificity of information) present in the BPS models.

A novel experimental approach was taken for estimating the reliability of decision making at the conceptual design stage (using risk metric ERPL), given a particular design decision making process (i.e. sequential process). The following subsections present the highlights of the risk assessment method developed as part of this thesis and some key findings from the experiment conducted to assess the reliability of decision making at the conceptual design stage.

7.1.1 A method for assessing risk in performance-based decision making

ERPL, the proposed risk metric, reports three possible sources performance loss that can be encountered due to absence of important design features in conceptual stage BPS models: 1) reversal in ranks of design alternatives 2) latency effect or a delayed discovery of performance gain 3) insufficient performance gain or loss of expected performance gain.

The risk estimation method (for calculating ERPL) was further extended to evaluate the risk (as 'high' or 'low') to trigger remedial actions to be taken by the DM. For setting a threshold for high risk, maximum permissible loss at the end of the decision making process is set. The risk is also evaluated as being 'high' or 'low' based on its distribution on the tree of the many possible future design paths. Instead of relying on arbitrary thresholds of probability of loss on the overall outcomes of the decision making process (e.g. 5%, 10%), the DM has an opportunity a limit on loss that he/she could endure. The DM can set a safe limit on the percentage of future design paths that could result in loss. The DM can also define the maximum loss on such design paths. If the risk is high, the DM could increase the BPS model fLOD or take other remedial measures before choosing a massing-scheme.

The proposed risk assessment does imply a greater computational burden on the conceptual stage decision maker. On projects where BPS is being utilized mostly for compliance purposes, carrying out a risk assessment involving multiple future design scenarios maybe considered as excessive. Nevertheless, when undertaking large projects and when a number of design alternatives are being considered, risk assessment can add significant value. The risk assessment is not tied to any absolute performance threshold, and can be used irrespective of performance goals set for the project.

7.1.2 Risk in conceptual stage design decisions

An experiment was conducted to assess the risk of performance loss in design decision made based on conceptual stage design decisions. Several future design scenarios were developed for a set of conceptual design stage neighborhood-scale massing schemes, though the scope of future design scenarios was limited to facade design features. The BPS models of massing-schemes at the conceptual design stage were designated as fLOD0 models where facade design related information present in the fLOD0 BPS model is considered unreliable and subject to change.

This thesis departs from a case-study based approach that is often used to illustrate a new decision making methodologies. The experimental approach allows for greater genralizability of the findings. This approach also facilitated exploration into impact of some of the assumptions made regarding the DM during the implementation of the methodology.

Incidence of risk in conceptual design decision making

'High' risk was found in 22%, 15% and 8% fLOD0 evaluations on sDA, annual heating and cooling demand evaluations, respectively. These findings present the incidence rate (frequency of occurrence) of risk in conceptual stage decision making (within the scope of the experiment) based on the 780 massing-scheme comparisons that were conducted. The percentages of high risk evaluations mentioned here correspond to 1 in 5 cases on sDA, 1 in 7 on annual heating demand, 1 in 12 on annual cooling demand evaluations at fLOD0. These findings are relevant to DMs that could be designing multiple neighborhood scale projects every year and are conducting 5-12 comparisons between massing-scheme designs on the performance metrics considered in this thesis. If the occurrence of risky cases had been found to be 1% cases or 1 in 100 (as opposed to 1 in 5,7,12), then these findings would mainly be of interest to a policy maker who could have a wider influence on modeling and decision making processes adopted on a large number of projects. Given the significance of risk in conceptual design stage evaluation, it is found to be relevant not just at a general industry level, but at an individual organization level.

These findings correspond to the assumptions made for the 'baseline' DM for this thesis and could vary with the DM's outlook for risk and decision making related preferences. The incidence rate of risk was found to be highly sensitive to the DM's preferences (see table 5.3 and table 5.2). Better estimations of the incidence rate of high risk(at a broad industry level) could be achieved by taking a probabilistic view of the preferences of DMs rather then assume uniformity in the DM's preferences. As an early exploration in this area, this thesis presents the potential impact of the DM's preferences on the outcomes of the decision making process.

Effectiveness of fLODs in reducing risk

Risk is reported in this thesis at a given fLOD at which two design alternatives are compared. The risk assessment in this thesis further focuses on fLOD0, which is the lowest level of detail considered in this thesis. Increasing the fLOD implies that an additional design feature is added to the BPS model (e.g. active shading or fixed shading devices like balconies) or uncertainty in inputs is reduced (e.g. specific WWR is used in place of the default value at fLOD0). By increasing the fLOD, the DM is progressing towards the final design. ERPL, which is the risk in decision making due to unknown design details can also be interpreted as the potential performance loss that can be avoided by making the unknown design details, known. Results from 780 comparisons conducted based on representative neighborhood design cases indicate that the reduction of ERPL with increasing fLOD is not same on all design paths. For example, ERPL for three metrics was found to be the highest on design paths related to high WWR. If high WWR is desired for the project, then design decisions (choice between massing schemes) are better made at higher fLODs. Even so, ERPL remained high at fLOD2 for high

WWR cases, which was the penultimate fLOD for the experiment conducted in this thesis. In those design paths (where ERPL is not zero at fLDO2). The DM may, in sme cases, have to go all the way to the highest fLOD before choosing the massing-scheme. At lower WWR (20-30%), ERPL was eliminated on some design paths, implying that no further design details are needed for decision making. The need for higher fLODs thus varies not just with the massing schemes being compared, but with the further facade design details that are chosen.

The ERPL risk metric is proposed as a criterion for choosing the BPS model fLOD but not as a criterion for making design decisions. For example, lower risk observed in lower WWR cases should not lead a DM to choose low WWR. The WWR and other design decisions should be made based on the performance goals established for the design. On heating and cooling demand, fewer number of risky cases are observed at low WWR. Thus, the knowledge of WWR is valuable for heating and cooling demand assessments and progressing to fLOD1 would be a significant step in increasing reliability, *if* low WWR is preferred.

Ability of ERPL to support design process

ERPL risk metric is proposed as a means to gauge the suitability of a given fLOD for making BPS based conceptual design stage decisions. A graphical set-up was proposed to explore the potential of ERPL metric and the scenario based risk assessment methodology in aiding the design process beyond determining the appropriate fLOD. This graphical set-up could potentially allow the DM to deepen his/her understanding of the risks that may lie ahead, beyond knowing the ERPL value, and thus trigger action that is better suited to the interests of the DM. Several illustrative uses of this set-up resulted in the following insights:

- Potential utility was seen in informing the DM of the nature of risk which is not conveyed by the ERPL value. The risk due to insufficient performance gain for example, could inform the DM of the likelihood of the two design alternatives delivering equivalent performance at high fLOD. In such cases the DM acquires flexibility to make the choice between massing-schemes based on other performance metrics. The DM thus gains the confidence to disregard a particular metric from his/her purview in a multi-criteria decision making environment.
- The DM could be interested specifically in big payoffs or high relative performance gain values that can be realized at high fLOD. Alternatively, the DM could also be interested in the most likely outcome of the design process at later stages (high fLOD). To support these specific interests of the DM, the underlying distribution of possible relative performance values needs to be revealed in some manner to the DM.
- It is possible that while making a decision under uncertainty, the DM first needs assistance in identifying his/her specific interest (e.g. big payoff, design flexibility). Helping a DM in identifying his/her interests, that are also reasonable to achieve, in a given situation could also support the decision making process.

Can high risk be anticipated?

Out of all the high-risk cases found in the 780 pair-wise massing-scheme comparisons, only 5% cases had a difference of more than 18.3% in sDA at fLOD0. The same effect was observed on heating and cooling demand performance at difference of 2.8 kWh/m2-year, 8.3kWh/m2-year respectively. When performance difference at fLOD0 is smaller, the chances of encountering high-risk is higher. These results could be interpreted as the amount of performance difference that needs to be observed between two conceptual design alternatives to avoid risk of performance loss from uncertainty in facade design details. Thus a significant difference in conceptual design alternatives indicates lower risk. On heating demand though the risk was mainly due to the latency effect. That is, on heating demand evaluations at fLOD0, in 35% cases where the DM failed to obtain sufficient difference between design alternatives latency effect was found. Thus, a small performance difference in heating demand between two design alternatives at fLOD0 must be viewed with skepticism.

7.1.3 Future applications of risk assessment method

Potential method for compliance with ASHRAE 209 (Modeling cycle - 2)

The ASHRAE standard- "Energy simulation aided design for buildings except low-rise residential buildings" [ASHRAE, 2018] was recently released. This standard lays down the minimum requirements for the analysis that must be done before any conceptual design alternative is rejected. However the standard does not offer guidance on how such an analysis must be done as the developers of the standard did not want to make it prescriptive in nature. The standard relies on the user of the standard to determine a decision making process that suits his/her needs. The risk evaluation methodology presented in this thesis can be used by a DM who is looking chose between design alternatives based on relative performance differences and not based on compliance with a specific performance threshold. If the risk of performance loss in the DM's choice is found to be low, he/she can proceed to the next design stage without committing to any specific facade details (i.e. details within the scope of the risk assessment) as the DM's choice is likely to be upheld irrespective of future facade design choices.

Risk evaluation: a conceptual design stage safety measure against value engineering?

While the design process is still ongoing, all design decisions remain susceptible to unanticipated change later on. During the course of work undertaken for this thesis, several local architects were approached to better understand their design process and issues they face while making decisions (The outcomes of the interviews that took place with them are summarized in appendix A.1). One of the architects (Architect 1 in Appendix A.1) mentioned that on a recent project, in the final design phase, recessed balconies were completely eliminated from the design. Figure 7.1 shows the project she mentioned. Figure 7.1 (a) shows the design as submitted for the design competition. Figure 7.1 (b) shows the final execution of the design without balconies. In this case, the removal of the balconies was requested by the client (building owner). This is one example of many reasons that design decision taken earlier can

(a) (b)

Figure 7.1 – (a) rendering of design submitted by architect for design competition organized for the project (Image: ON Architecture)(b) The balconies were eliminated in the final design (Image:Google Street View).

be overturned in the later design stages[1]. What can a conceptual stage decision maker do to counteract such uncertainties that are clearly beyond his/her control?

While this thesis treats the sequential nature of decision making as the main source of uncertainty in design-features at the conceptual design stage, other sources of uncertainty (such as unexpected changes in client needs) also exist and can result in performance loss or 'design-gap'. The method for assessing risk of performance loss, presented in this thesis, is agnostic to the cause of uncertainty in design features and can be used to assess risk of unexpected changes in the later design stages. If the final design, due to unanticipated alterations, is found to align with scenarios that were initially identified as high risk, then the conceptual stage design choice could be rendered invalid. The risk assessment could be useful to the DM when he/she is deliberating upon altering the design at the final design stages through a process such as value engineering.

7.2 Limitations

This thesis demonstrates the improvement in reliability of decision making that can be achieved by increasing the BPS model fLOD. The risk, whether reported as ERPL, or as risk from each of loss causing effects (rank reversal, latency effect an insufficient gain) is calculated and reported with respect to specific model fLOD. BPS model fLOD is thus treated an essential part of reporting risk.

[1] After consultations with 11 construction industry experts (in the roles of clients, contractors and consultants) Yap et al. [2018] identified 39 reasons due to which unanticipated design changes could be made during the design process with the following as three most common reasons 1) "value engineering", 2) "lack of coordination among various professional consultants" and 3) "change of requirement".

7.2.1 Context and modeling assumptions

The findings of this thesis regarding the prevalence rates of high-risk cases at the conceptual design stage are a unique aspect of this study. At the same time, prevalence rates of unreliable decision making that are reported here are closely related to the context and scope of the study, especially the density (1.0), the location (Geneva, Switzerland) and the building code requirements associated with the geographical context. The simulation time for daylight BPS models needed to arrive at these findings was approximately 37.4 days on a 30 core machine. Rapid replication for new contexts and larger building scales (bigger sensor grid arrays, more thermal zones) could be challenging. However, these challenges could be partially overcome using other simulation methods [McNeil and Lee, 2013].

As stated earlier in the methodology section, limited facade design features have been considered. Excluded factors (see table 3.3) could in some cases amplify and in other cases diminish the observed risk in performance loss. As mentioned in table 3.1 several performance metrics were not found appropriate for the scope of the experiment.

Scenarios as modeled (N=48), and inclusion of more scenarios could lead to more nuanced findings. For example three WWR were considered in the design scenarios (20%, 30%, 40%). If scenarios were built using smaller intervals (e.g. at 5% increments or smaller), then the precision with which the DM needs to know the WWR to reduce risk could also be determined. Since the difference in three WWR values used in the scenarios is currently quite large (10%), they become discrete design choices.

Single, deterministic, active blind operation scenario was modeled to represent the interaction between a occupant behavior (with respect to operation of exterior blinds) and the design from fLOD1 to fLDO3. Multiple scenarios resulting from active blinds could have been included if more information were available regarding influence of design features on occupant behavior regarding blind usage. Currently available manually-operated blind-usage models rely on environmental inputs such as irradiation received at the window, depth of direct sunlight penetration [Van Den Wymelenberg, 2012] or interior area receiving daylight penetration (e.g. LM-83-12 [IESNA, 2012]). If the models could also include effect of design features such as type of blinds (e.g. fabric, venetian, sliding panels) or mode of operation (mechanized vs manual), then more design scenarios could have been considered. Multiple occupant behavior scenarios could have been used to represent the expected variance in behavior of different building occupants. However, the current common practice for consideration of occupant behavior modeling at the urban scale [Happle et al., 2018] was used assuming that given the large number of occupants in a neighborhood scale model, representative behavioral model inputs would suffice.

Impact of various massing-scheme designs on micro-climate effects such as heat-island effect and air-flow (outside and inside the buildings) was ignored. The thermal simulation models that were set up in Energy Plus account for short and long-wave radiation, convective heat transfer and heat conductance through the walls and ground surfaces. Wind flow affects

heat dissipation from the building surfaces and surroundings. To model wind flow, the thermal models needed to be coupled with a computational fluid dynamics models (CFD) (e.g. [Allegrini et al., 2015]). Natural ventilation, as a passive cooling design strategy was ignored and focus was kept on the building envelope and solar gain avoidance for achieving lower cooling demand. Modeling natural ventilation, without interior wall partitions would have been inappropriate. Interior wall partition related design scenarios were excluded from the scope of the experiment as their inclusion at the lowest level of design detail required an simplified version (of interior partitioning) that could be considered representative of detailed interior partition related scenarios.

7.2.2 Possible scope extensions

Large pieces of vegetation such as trees were excluded from the risk assessment due to limited scope in assembling an LOD framework - . The placement of trees on a site could be related to where the designer chooses to place the buildings . Trees are often used for improve the indoor environment in several ways such as seasonal solar protection, noise control and blocking unpleasant views. They can thus be treated as strategic design elements [Hongbing et al., 2010]. In order to truly represent a massing-scheme and its implications on an occupant's access to daylight and solar radiation, tree types and placement could be interesting to consider. Tree type and placement would have to be included as a design scenario as multiple placements would be possible with a give scheme. This thesis focused on facade details and incorporating vegetation elements was considered inharmonious without an aforementioned effort to do an exhaustive organization of design details to widen the scope of the risk assessment.

Aleatory (irreducible) forms of uncertainty such as weather were also excluded. Weather uncertainty is important to heating demand and overheating avoidance based performance [Pernigotto et al., 2014] but their effect on annual daylight assessments is small [Iversen et al., 2013]. Given the goals of this thesis, the sources of risk were kept consistent across metrics. This work is thus able to report reliability of conceptual design decision made based on various metrics on equal footing. Also, including aleatory forms of uncertainty in the performance evolution is typically viewed as specialized 'type of analysis' and not a model LOD. If a LOD framework were to be developed for BPS models, it could be extended beyond physical design elements to treat sources of uncertainty such weather also as 'levels-of-detail' in BPS based performance evaluations.

The proposed risk assessment methodology assumes that all future design scenarios are equally likely. It is possible that some scenarios are less likely to be implemented as the final design compared to some other scenarios. For example scenarios resulting in higher construction cost may be less likely to be chosen as the final design. An architect may be able to assign probabilities to the various scenarios based on their prior experience or intuition regarding the project at hand. If such information can be obtained and incorporated into the risk analysis, it can make the risk assessment more informative and pertinent for the DM.

Two characteristics ofa DM's persona have been explored in this thesis. However any DM's persona will invariably be composed of several other characteristics as well. Thus characteristics explored in this thesis are not expected to influence the DM's decision making practices one-at-a-time but rather in conjunction with each other and many others. Other characteristics such as degree of risk aversion and further nuances in the decision making problem (e.g. multi-criteria decision making) may influence decision making. However, the simplified interpretation of the DM's characteristics as presented in this chapter indicates that their effect can be significant. Further development of the DM's persona and acknowledgment of different personas thus appears to be important when trying to estimate the outcome of decision making using BPS tools.

As part of this work, a proposal has been given for visualizing risk in decision making. This decision making aid is targeted at a decision maker who is in a key position to monitor the progress of the design development and also trigger actions like re-evaluation of performance as and when needed. However risk based decision making is a novel concept for performance driven design and several aspects of the risk assessment need to tested with real users. Uncertainty analysis is far from common practice [Østergård et al., 2017] in the BPS domain and existing BPS users may need to be sensitized on several aspects of the approach before they can find it useful. Thus the risk based decision making approach was not tested with any real users.

7.3 Future research directions and outlook

This thesis exclusively addressed uncertainty due to future facade design decisions at the conceptual design stage. A broad and exhaustive organization of design information, that is important to BPS models, was considered outside the scope of this thesis. A larger and conscientious effort is needed to prepare a cohesive and well-rounded BPS model LOD framework. Such a framework, if adopted widely, could bring clarity and much needed standardization in reporting of BPS based performance results.

Another important area of further investigation is to examine loss-aversion [Tversky and Kahneman, 1974] characteristics of a BPS based DM. Loss-aversion is a behavioral trait of a DM and can have a large impact on the findings of this thesis. Prospect theory [Tversky and Kahneman, 1974], based on empirical studies of decision making behaviors, says that the displeasure from a unit of loss outweighs the pleasure from a unit of gain, which denotes loss-aversion. Exclusion of loss-aversion from this thesis has resulted in conservative assumptions. The maximum unacceptable loss for a DM is assumed to be equal to the gain (payoff) that the DM is looking to achieve out of the decision making process. If loss-aversion behavior was accounted for, the DM would accept a much smaller loss (as an unavoidable outcome of uncertainty) while making efforts to achieve a certain relative performance gain or payoff (e.g. generating multiple design alternatives, carrying of performance evaluations, risk evaluation). If the unacceptable loss value is lower, the risk threshold would be lower and many more cases

would be reported as being risky (e.g. currently found to be 1 in 5 for sDA for the baseline DM).

This thesis shows that the decision making practices of BPS users are a critical element of a performance-driven design process. Concepts of 'performance-gap' [Carbon Trust, 2012; Menezes et al., 2012; Cohen and Bordass, 2015] and 'design-gap' [Wright et al., 2016] are meant to bring accountability into the use of BPS tools for building design. Performance-gap draws the attention of BPS users and BPS based DMs to modeling practices that could result in large difference between performance estimated during the design process and post-occupancy. The design-gap lays emphasis on practices used for design space exploration. The quest for greater accountability is meant to culminate in better design decision making and would thus be incomplete without considering the decision making practices. More researchers in the BPS domain need to extend their work to include design decision making mechanisms as well.

In this thesis we modeled the simple decision making process involved in choosing between two design alternatives. New methodological approaches to making design decisions (e.g. [Hester et al., 2017]);[Basbagill et al., 2014]; [Chinazzo et. al., 2015]) for using BPS could potentially be evaluated by modeling the corresponding decision making context. This could provide a parallel/complementary approach to testing effectiveness of decision making methods with real-users. Methods for making design decisions can be challenging to test with real-users as it requires the development of a usable interface and finding a suitable pool of test-users. User experience, training and degree of interest in various aspects of building performance and design components can vary. Role and position (or decision making-power) that they currently hold may also vary. To achieve a harmonious group of test participants can be quite challenging. Subjective-evaluations or objectively tracking user movement (clicks and cursor movement) (e.g. [Jones and Reinhart, 2019]) [2] within the tools may not reveal the users' thought process.

Application of new methodological approaches to decision making is acquiring a role in policy-making. The ASHRAE 209 standard released in 2018 [ASHRAE, 2018] is an interesting new development for the use of BPS in building design and is quite different from prior performance standards and modeling protocols (e.g. LM-83-12 [IESNA, 2012], ASHRAE 90.1 [ASHRAE, 2012]). While modeling protocols establish best practice for various model inputs for evaluating performance, the new ASHRAE standard 209 sets requirements for the nature of analysis that needs to be done before design decisions are made. This standard has been constituted to facilitate contractual agreements between the project owner and the design team to ensure a minimum quality of BPS based analysis for decision making. Development of performance and modeling standards relies on extensive testing under various design conditions. Development of new design process related standards needs to rely on equally robust testing under various decision making mechanisms.

[2] [Jones and Reinhart, 2019] is mentioned here as an example of a study that compared the use of two glare assessment tools cursor movements and clicks. The study also compared the design decisions made by users while using the two different tools. It was not the objective of this study to understand the user's design decision making process. However, it is a good example of how detailed tracking of the user's work can only reveal limited parts of their process of arriving at a design solution.

This thesis has opened up unexplored aspects of reliability in design decision making at the conceptual/early design stage when relying on BPS based performance evaluations. The early design stage is often referred to as the most critical design stage and there is universal push to incorporate the use of BPS tools as early as possible in the design process. This thesis has revealed that without robust decision making methods, the conceptual/early stage decisions can be compromised given the high level of uncertainty in several important design details. BPS based educational curricula need to place greater emphasis on sensitizing BPS based decision makers and users on the implications of the decision making practices that they adopt. Also, this work paves the path for further development of mechanisms to qualify BPS results with the reliability in decision making that can be achieved at a given design stage.

Bibliography

Agarwal, M., Pastore, L., and Andersen, M. (2017). Suitability of neighborhood-scale massing models for daylight performance evaluation. In *International Conference on Sustainable Design of the Built Environment SDBE 2017*.

Akin, O. and Moustapha, H. (2004). Strategic use of representation in architectural massing. *Design Studies*, 25(1):31–50.

Allegrini, J., Dorer, V., and Carmeliet, J. (2015). Influence of morphologies on the microclimate in urban neighbourhoods. *Journal of Wind Engineering and Industrial Aerodynamics*, 144:108–117.

Alsaadani, S. and De Souza, C. B. (2016). Of collaboration or condemnation? exploring the promise and pitfalls of architect-consultant collaborations for building performance simulation. *Energy Research & Social Science*, 19:21–36.

Architects, T. A. I. o. (2012). Integrating Energy Modeling in the Design Process. Technical report, American Institute of Architects, USA.

Arrow, K. J. and Hurwicz, L. (1972). An optimality criterion for decision-making under ignorance. *Uncertainty and expectations in economics*, 1.

ASHRAE (2018). *Energy Simulation Aided Design for Buildings except Low-Rise Residential Buildings (ANSI Approved)ASHRAE 209-2018*. ASHRAE.

Attia, S. and De Herde, A. (2011). Early design simulation tools for net zero energy buildings a comparison of ten tools. In *Conference Proceedings of 12th International Building Performance Simulation Association, 2011*.

Attia, S., Gratia, E., De Herde, A., and Hensen, J. L. (2012a). Simulation-based decision support tool for early stages of zero-energy building design. *Energy and buildings*, 49:2–15.

Attia, S., Hensen, J. L., Beltrán, L., and De Herde, A. (2012b). Selection criteria for building performance simulation tools: contrasting architects' and engineers' needs. *Journal of Building Performance Simulation*, 5(3):155–169.

Baker, N. and Steemers, K. (1996). Lt method 3 a strategic energy-design tool for southern europe. *Energy and Buildings*, 23(3):251–256.

Baker, N. and Steemers, K. (2003). *Energy and environment in architecture a technical design guide*. Taylor & Francis.

Basbagill, J., Flager, F., Lepech, M., and Fischer, M. (2013a). Application of life-cycle assessment to early stage building design for reduced embodied environmental impacts. *Building and Environment*, 60:81–92.

Basbagill, J., Flager, F., Lepech, M., and Fischer, M. (2013b). Application of life-cycle assessment to early stage building design for reduced embodied environmental impacts. *Building and Environment*, 60:81–92.

Basbagill, J. P., Flager, F. L., and Lepech, M. (2014). A multi-objective feedback approach for evaluating sequential conceptual building design decisions. *Automation in Construction*, 45:136–150.

Ben Haim, Y. (2006). *Info gap decision theory decisions under severe uncertainty*. Elsevier.

Bleil De Souza, C. and Tucker, S. (2016). Placing user needs at the center of building performance simulation (bps) tool development: using'designer personas' to assess existing bps tools. In *Building Simulation and Optimization, Newcastle, UK, 12-14 September 2016*.

Bodart, M., Deneyer, A., Herde, A. D., and Wouters, P. (2007). A guide for building daylight scale models. *Architectural Science Review*, 50(1):31–36.

Boyko, C. T., Cooper, R., Davey, C. L., and Wootton, A. B. (2006). Addressing sustainability early in the urban design process. *Management of Environmental Quality An International Journal*.

Brembilla, E., Hopfe, C., and Mardaljevic, J. (2018). Influence of input reflectance values on climate based daylight metrics using sensitivity analysis. *Journal of Building Performance Simulation*, 11(3):333–349.

Cabannes, T., Shyu, F., Porter, E., Yao, S., Wang, Y., Vincentelli, M. A. S., Hinardi, S., Zhao, M., and Bayen, A. M. (2018). Measuring regret in routing: assessing the impact of increased app usage. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2589–2594. IEEE.

Capeluto, I. and Shaviv, E. (2001). On the use of solar volume for determining the urban fabric. *Solar Energy*, 70(3):275–280.

Chaillou, S. (2019). *Solar AI plus Architecture Towards a New Approach*. PhD Thesis, Harvard Graduate School of Design, University of Harvard, Boston, USA.

Chatzipoulka, C., Compagnon, R., and Kaempf, J. (2018). An Image-Based Method to Evaluate Solar and Daylight Potential in Urban Areas. In *Proceedings of the Symposium for Architecture and Urban Design*, pages 337–344, Delft, The Netherlands.

Chatzipoulka, C., Compagnon, R., and Nikolopoulou, M. (2016). Urban geometry and solar availability on façades and ground of real urban forms using London as a case study. *Solar Energy*, 138:53–66. 00002.

Cheng, V., Steemers, K., Montavon, M., and Compagnon, R. (2006). Urban form density and solar potential. In *PLEA 2006, Geneva, Switzerland*.

Chinazzo, G., Rastogi, P., and Andersen, M. (2015). Robustness assessment methodology for the evaluation of building performance with a view to climate uncertainties. Technical report, EPFL.

CIBSE, C. (2015). *AM11 Building Performance Modelling 2015*. CIBSE Building Services Knowledge, London,UK.

Clarke, J. (2015). A vision for building performance simulation a position paper prepared on behalf of the ibpsa board. *Journal of Building Performance Simulation*, 8(2):39–43.

Clarke, J. A. and Hensen, J. (2015). Integrated building performance simulation progress prospects and requirements. *Building and Environment*, 91:294–306.

Colyvan, M. (2008). Is probability the only coherent approach to uncertainty? *Risk Analysis: An International Journal*, 28(3):645–652.

Compagnon, R. (2004). Solar and daylight availability in the urban fabric. *Energy and Buildings*, 36(4):321–328.

Compagnon, R., Antonutto, G., Longato, P., and Rotsch, A. (2015). Assessing daylight and sunlight access in the built environment a new tool for planners and designers. In *Proceedings for PLEA 2015*.

Darren, R. (2006). Urban morphology and indicators of radiation availability. *Solar Energy*, 80(12):1643–1648.

Dave, C., Eckel, C. C., Johnson, C. A., and Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty*, 41(3):219–243.

de Souza, C. B. (2009). A critical and theoretical analysis of current proposals for integrating building thermal simulation tools into the building design process. *Journal of Building Performance Simulation*, 2(4):283–297.

De Wilde, P., Augenbroe, G., and Van Der Voorden, M. (1999). Invocation of building simulation tools in building design practice. In *Proceedings of IBPSA '99 buildings simulation conference*, pages 1211–1218.

De Wit, S. (2001). *Uncertainty in predictions of thermal comfort in buildings*. PhD, Delft University of Technology, Delft, The Netherlands.

De Wit, S. and Augenbroe, G. (2002). Analysis of uncertainty in building design evaluations and its implications. *Energy and Buildings*, 34(9):951–958.

DeKay, M. (2010). Daylighting and Urban Form An Urban Fabric of Light. *Journal of Architectural and Planning Research*, pages 35–56.

Dogan, T. and Park, Y. C. (2017). A new framework for residential daylight performance evaluation. *BS 2017, San Francisco, USA*.

Dogan, T. and Park, Y. C. (2019). A critical review of daylighting metrics for residential architecture and a new metric for cold and temperate climates. *Lighting Research & Technology*, 51(2):206–230.

Dogan, T., Reinhart, C., and Michalatos, P. (2016). Autozoner an algorithm for automatic thermal zoning of buildings with unknown interior space definitions. *Journal of Building Performance Simulation*, 9(2):176–189.

Domínguez-Muñoz, F., Cejudo-López, J. M., and Carrillo-Andrés, A. (2010). Uncertainty in peak cooling load calculations. *Energy and Buildings*, 42(7):1010–1018.

Donn, M., Selkowitz, S., and Bordass, B. (2009). Simulation in the service of design asking the right questions. Technical report, Lawrence Berkeley National Lab, Berkeley, CA, United States.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.

Fisher, R. A. et al. (1950). Statistical methods for research workers. *Statistical methods for research workers*.

for Standardization CEN, E. C. (2019). EN17037 2019 Daylight in buildings. Technical report, European Standards.

Gang, W., Wang, S., Augenbroe, G., and Xiao, F. (2016). Robust optimal design of district cooling systems and the impacts of uncertainty and reliability. *Energy and Buildings*, 122:11–22.

Gang, W., Wang, S., Yan, C., and Xiao, F. (2015). Robust optimal design of building cooling systems concerning uncertainties using mini-max regret theory. *Science and Technology for the Built Environment*, 21(6):789–799.

Ghaffarianhoseini, A., AlWaer, H., Omrany, H., Ghaffarianhoseini, A., Alalouch, C., Clements-Croome, D., and Tookey, J. (2018). Sick building syndrome are we doing enough *Architectural Science Review*, 61(3):99–121.

Gibson, M. D. (2014). Integrating geometry and light daylight solutions through performance based algorithms. *ARCC Conference Repository*, 0(0).

Gilani, S., O'Brien, W., Gunay, H. B., and Carrizo, J. S. (2016). Use of dynamic occupant behavior models in the building design and code compliance processes. *Energy and Buildings*, 117:260–271.

Hachem, C., Athienitis, A., and Fazio, P. (2011). Parametric investigation of geometric form effects on solar potential of housing units. *Solar Energy*, 85(9):1864–1877.

Happle, G., Fonseca, J. A., and Schlueter, A. (2018). A review on occupant behavior in urban building energy models. *Energy and Buildings*, 174:276–292.

Heo, Y., Choudhary, R., and Augenbroe, G. (2012). Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47:550–560.

Hester, J., Gregory, J., and Kirchain, R. (2017). Sequential early-design guidance for residential single-family buildings using a probabilistic metamodel of energy consumption. *Energy and Buildings*, 134(Supplement C):202–211.

Hitchcock, R. J., Mitchell, R., Yazdanian, M., Lee, E., and Huizenga, C. (2008). Comfen a commercial fenestration facade design tool. *Proceedings of SimBuild*, 3(1):246–252.

Hoes, P., Hensen, J., Loomans, M., de Vries, B., and Bourgeois, D. (2009). User behavior in whole building simulation. *Energy and buildings*, 41(3):295–302.

Hoes, P., Trcka, M., Hensen, J., and Bonnema, B. H. (2011). Optimizing building designs using a robustness indicator with respect to user behavior. In *Proceedings of the 12th conference of the international building performance simulation association*, pages 14–16.

Hongbing, W., Jun, Q., Yonghong, H., and Li, D. (2010). Optimal tree design for daylighting in residential buildings. *Building and Environment*, 45(12):2594–2606.

Hopfe, C. J., Augenbroe, G. L., and Hensen, J. L. (2013). Multi criteria decision making under uncertainty in building performance assessment. *Building and environment*, 69:81–90.

Hopfe, C. J. and Hensen, J. L. (2011). Uncertainty analysis in building performance simulation for design support. *Energy and Buildings*, 43(10):2798–2805.

Housing, B. (2018). BC Energy Step Code – Design Guide. Technical report, City of Vancouver, the City of New Westminster, the Province of BC.

Huang, P., Huang, G., and Wang, Y. (2015). Hvac system design under peak load prediction uncertainty using multiple-criterion decision making technique. *Energy and Buildings*, 91:26–36.

Huber, J. and Nytsch-Geusen, C. (2011). Development of modeling and simulation strategies for large scale urban districts. In *Proceedings of Building Simulation*, volume 2011, pages 1753–1760.

IESNA (2012). IES LM 83 12 IES Spatial Daylight Autonomy (sDA) and Annual Sunlight Exposure (ASE). Technical Report IES LM 83 12, Illuminating Engineering Society of North America (IESNA).

Iversen, A., Svendsen, S., and Nielsen, T. R. (2013). The effect of different weather data sets and their resolution on climate based daylight modelling. *Lighting Research & Technology*, 45(3):305–316.

Jones, N. L. and Reinhart, C. F. (2019). Effects of real-time simulation feedback on design for visual comfort. *Journal of Building Performance Simulation*, 12(3):343–361.

Jusselme, T., Antunes Fernandes, P., Rey, E., and Andersen, M. (2019). Design guidance from a data driven lca based design method and tool prototype. In *Proceedings of Building Simulation 2019, 16th Conference of IBPSA*.

Jusselme, T., Rey, E., and Andersen, M. (2020). Surveying the environmental life-cycle performance assessments: Practice and context at early building design stages. *Sustainable Cities and Society*, 52:101879.

Jusselme, T., Tuor, R., Lalanne, D., Rey, E., and Andersen, M. (2017). Visualization techniques for heterogeneous and multidimensional simulated building performance data sets. In *Proceedings of the International Conference for Sustainable Design of the Built Environment*.

Kim, Y.-J., Ahn, K.-U., and Park, C.-S. (2014). Decision making of hvac system using bayesian markov chain monte carlo method. *Energy and Buildings*, 72:112–121.

Knowles, R. L. (2003). The solar envelope: its meaning for energy and buildings. *Energy and Buildings*, 35(1):15 – 25. Special issue on urban research.

Kotireddy, R., Hoes, P.-J., and Hensen, J. L. (2018). A methodology for performance robustness assessment of low-energy buildings using scenario analysis. *Applied energy*, 212:428–442.

Kotireddy, R., Hoes, P.-J., and Hensen, J. L. M. (2019). Integrating robustness indicators into multi-objective optimization to find robust optimal low-energy building designs. *Journal of Building Performance Simulation*, 12(5):546–565.

Kvan, T. and Thilakaratne, R. (2003). Models in the design conversation architectural vs engineering. In *International Conference of the Association of Architecture Schools of Australasia, AASA*, volume 2, pages 1–11.

Lam, K. P., Wong, N. H., and Henry, F. (1999). A study of the use of performance-based simulation tools for building design and evaluation in singapore. *Architecture*, 1:11–13.

Liu, Y., Chakrabarti, A., and Bligh, T. (2003). Towards an ideal approach for concept generation. *Design studies*, 24(4):341–355.

Lomas, K. J. and Eppel, H. (1992). Sensitivity analysis techniques for building thermal simulation programs. *Energy and buildings*, 19(1):21–44.

Macdonald, I. and Strachan, P. (2001). Practical application of uncertainty analysis. *Energy and Buildings*, 33(3):219–227.

MacMillan, S., Steele, J., Kirby, P., Spence, R., and Austin, S. (2002). Mapping the design process during the conceptual phase of building projects. *Engineering construction and architectural management*, 9(3):174–180.

Madansky, A. (1960). Inequalities for Stochastic Linear Programming Problems. *Management Science*, 6(2):197–204.

Madansky, A. (1962). Methods of Solution of Linear Programs under Uncertainty. *Operations Research*, 10(4):463–471.

Mahdavi, A. (2003). Computational building models theme and four variations. In *Proceedings of building simulation 2003, Eindhoven, vol. 1*.

Mardaljevic, J. (1999). *Daylight simulation: validation, sky models and daylight coefficients*. De Montfort University Leicester.

McPhail, C., Maier, H., Kwakkel, J., Giuliani, M., Castelletti, A., and Westra, S. (2018). Robustness metrics: How are they calculated, when should they be used and why do they give different results? *Earth's Future*, 6(2):169–191.

Metz, B., Davidson, O., Bosch, P., Dave, R., and Meyer, L. (2007). *Climate change 2007: Mitigation of climate change*. Cambridge Univ. Press.

Miller, C., Quintana, M., and Glazer, J. (2019). Twenty years of building simulation trends: Text mining and topic modeling of the bldg-sim email list archive. In *Proceedings of Building Simulation 2019: 16th Conference of IBPSA*.

Mohajeri, N., Upadhyay, G., Gudmundsson, A., Assouline, D., Kämpf, J., and Scartezzini, J.-L. (2016). Effects of urban compactness on solar energy potential. *Renewable Energy*, 93:469–482.

Nault, E. (2016). *Solar Potential in Early Neighborhood Design. A Decision-Support Workflow Based on Predictive Models*. PhD Thesis, Ecole polytechnique fédérale de Lausanne, Lausanne, Switzerland.

Nault, E., Jusselme, T., Aguacil, S., and Andersen, M. (2020). Strategic environmental urban planning a contextual approach for defining performance goals and informing decision making. *Building and Environment*, 168:106448.

Nault, E., Peronato, G., Rey, E., and Andersen, M. (2015). Review and critical analysis of early-design phase evaluation metrics for the solar potential of neighborhood designs. *Building and Environment*, 92:679–691.

Nault, E., Waibel, C., Carmeliet, J., and Andersen, M. (2018). Development and test application of the UrbanSOLve decision-support prototype for early-stage neighborhood design. *Building and Environment*, 137:58–72.

Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, pages 175–240.

Nik, V. M., Mata, E., and Kalagasidis, A. S. (2015). A statistical method for assessing retrofitting measures of buildings and ranking their robustness against climate change. *Energy and Buildings*, 88:262–275.

Nikolaidou, E., Wright, J. A., and Hopfe, C. J. (2017). Robust building scheme design optimization for uncertain performance prediction. In *Building Simulation*. IBPSA.

OBrien, W., Athienitis, A., and Kesik, T. (2011). Thermal zoning and interzonal airflow in the design and simulation of solar houses a sensitivity analysis. *Journal of Building Performance Simulation*, 4(3):239–256.

of Zurich, C. (2012). Dichter a documentation of the structural change in zurich 30 examples. Technical report, Zurich city building construction departement, Urban Planning Office.

Okeil, A. (2010). A holistic approach to energy efficient building forms. *Energy and buildings*, 42(9):1437–1444.

Olgyay, A. et al. (1957). *Solar control and shading devices*. Princeton.

Østergård, T., Jensen, R. L., and Maagaard, S. E. (2017). Early building design: Informed decision-making by exploring multidimensional design space using sensitivity analysis. *Energy and Buildings*, 142:8–22.

Pacheco, R., Ordoneez, J., and Martinez, G. (2012). Energy efficient design of building: A review. *Renewable and Sustainable Energy Reviews*, 16(6):3559–3573.

Parys, W., Breesch, H., Hens, H., and Saelens, D. (2012). Feasibility assessment of passive cooling for office buildings in a temperate climate through uncertainty analysis. *Building and Environment*, 56:95–107.

Pedro, J. B., Meijer, F., and Visscher, H. (2011). Comparison of building permit procedures in european union countries. In *RICS Construction and Property Conference*, volume 356.

Perez, M. G. R. and Rey, E. (2013). A multi criteria approach to compare urban renewal scenarios for an existing neighborhood. case study in lausanne switzerland. *Building and Environment*, 65:58–70.

Pernigotto, G., Prada, A., Gasparella, A., and Hensen, J. L. (2014). Analysis and improvement of the representativeness of en iso 15927 4 reference years for building energy simulation. *Journal of Building Performance Simulation*, 7(6):391–410.

Pessenlehner, W. and Mahdavi, A. (2003). *Building morphology, transparence, and energy performance*. na.

Porritt, S. M., Cropper, P. C., Shao, L., and Goodier, C. I. (2012). Ranking of interventions to reduce dwelling overheating during heat waves. *Energy and Buildings*, 55:16–27.

Raiffa, H. (1968). *Decision analysis Introductory lectures on choices under uncertainty*. Addison Wesley.

Raiffa, H. and Schlaifer, R. (1961). *Applied statistical decision theory*. Wiley Cambridge.

Ratti, C., Baker, N., and Steemers, K. (2005). Energy consumption and urban texture. *Energy and Buildings*, 37(7):762–776.

Reinhart, C., Dogan, T., Jakubiec, J. A., Rakha, T., and Sang, A. (2013). Umi-an urban simulation environment for building energy use, daylighting and walkability. In *13th Conference of International Building Performance Simulation Association, Chambery, France*.

Reinhart, C. and LoVerso, V. (2010). A rules of thumb-based design sequence for diffuse daylight. *Lighting Research & Technology*, 42(1):7–31.

Reinhart, C. F. (2014). *Daylighting Handbook: Fundamentals, Designing with the Sun*. Christoph Reinhart.

Reinhart, C. F. and Davila, C. C. (2016). Urban building energy modeling-a review of a nascent field. *Building and Environment*, 97:196–202.

Reinhart, C. F., Mardaljevic, J., and Rogers, Z. (2006). Dynamic daylight performance metrics for sustainable building design. *Leukos*, 3(1):7–31.

Rezaee, R., Brown, J., Augenbroe, G., and Kim, J. (2015). Assessment of uncertainty and confidence in building design exploration. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 29(4):429–441.

RIBA and Sinclair, D. (2012). BIM Overlay to the RIBA Outline Plan of Work. Technical Report 978 1 85946 467 0, Royal Institute of British Architects, UK, London.

Robinson, D., Campbell, N., Gaiser, W., Kabel, K., Le Mouel, A., Morel, N., Page, J., Stankovic, S., and Stone, A. (2007). Suntool a new modelling paradigm for simulating and optimising urban sustainability. *Solar Energy*, 81(9):1196–1211.

Robinson, J. W. (1990). Architectural research incorporating myth and science. *Journal of Architectural Education*, 44(1):20–32.

Rode, P., Keim, C., Robazza, G., Viejo, P., and Schofield, J. (2014). Cities and Energy Urban Morphology and Residential Heat Energy Demand. *Environment and Planning*, 41(1):138–162.

Rysanek, A. M. and Choudhary, R. (2013). Optimum building energy retrofits under technical and economic uncertainty. *Energy and Buildings*, 57:324–337.

Sattrup, P. A. and Strømann-Andersen, J. (2013). Building Typologies in Northern European Cities: Daylight, Solar Access, and Building Energy Use. *Journal of Architectural and Planning Research*, 30(1):56.

Savage, L. J. (1951). The Theory of Statistical Decision. *Journal of the American Statistical Association*, 46(253):55–67.

Saxena, M., Perry, T., Bonneville, C., and Heschong, L. (2011). Office daylighting potential. *Public Interest Energy Research Program final project report Prepared for California Energy Commission*.

Schlueter, A. and Thesseling, F. (2009). Building information model based energy exergy performance assessment in early design stages. *Automation in construction*, 18(2):153–163.

Schwehr, P. and Fischer, R. (2010). Morphology of swiss multi-family houses of swiss multi family homes 1919 1990. Technical report, Competence Centre for Typology and Foresight Planning in Architecture (CCTP), HSLU, Switzerland, Lucerne, Switzerland.

Sharifi, A. and Murayama, A. (2013). A critical review of seven selected neighborhood sustainability assessment tools. *Environmental Impact Assessment Review*, 38:73–87.

Shi, Z., Fonseca, J. A., and Schlueter, A. (2017). A review of simulation-based urban form generation and optimization for energy-driven urban design. *Building and Environment*, 121:119–129.

SIA (2017), Z. (2017). Construction durable Batiment Norme de comprehension a la norme SIA 112. Technical report, Société suisse des ingénieurs et des architectes(SIA).

Struck, C., de Wilde, P. J., Hopfe, C. J., and Hensen, J. L. (2009). An investigation of the option space in conceptual building design for advanced building simulation. *Advanced Engineering Informatics*, 23(4):386–395.

Su, H. T. and Tung, Y. K. (2012). Minimax Expected Opportunity Loss A New Criterion for Risk Based Decision Making. *The Engineering Economist*, 57(4):247–273.

Sun, Y., Gu, L., Wu, C. J., and Augenbroe, G. (2014). Exploring hvac system sizing under uncertainty. *Energy and Buildings*, 81:243–252.

Tian, W. (2013). A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, 20:411–419.

Tregenza, P. (2017). Uncertainty in daylight calculations. *Lighting Research & Technology*, 49(7):829–844.

Tregenza, P. and Mardaljevic, J. (2018). Daylighting buildings standards and the needs of the designer. *Lighting Research & Technology*, 50(1):63–79.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.

Tzempelikos, A. and Athienitis, A. K. (2007). The impact of shading design and control on building cooling and lighting demand. *Solar energy*, 81(3):369–382.

Van Den Wymelenberg, K. (2012). Patterns of occupant interaction with window blinds: A literature review. *Energy and Buildings*, 51:165–176.

Vanhoutteghem, L., Skarning, G. C. J., Hviid, C. A., and Svendsen, S. (2015). Impact of facade window design on energy daylighting and thermal comfort in nearly zero energy houses. *Energy and Buildings*, 102:149–156.

Vartholomaios, A. (2015). The residential solar block envelope: A method for enabling the development of compact urban blocks with high passive solar potential. *Energy and Buildings*, 99:303 – 312.

Wald, A. (1950). *Statistical decision functions*. Statistical decision functions. Wiley, Oxford, England.

Weber, C. and Perrels, A. (2000). Modelling lifestyle effects on energy demand and related emissions. *Energy Policy*, 28(8):549–566.

Wilson, L., Danforth, J., Davila, C. C., and Harvey, D. (2019). How to Generate a Thousand Master Plans A Framework for Computational Urban Design. In *Proceedings of SimAUD 2019*, page 8, Atlanta, USA.

Wright, J., Nikolaidou, E., and Hopfe, C. J. (2016). Exhaustive search; does it have a role in explorative design. In *BSO2016, 3rd Building Simulation and Optimization Conference, Newcastle, UK, September 12*, volume 14.

Xia, C., Zhu, Y., and Lin, B. (2008). Building simulation as assistance in the conceptual design. In *Building simulation*. Springer.

Yap, J. B. H., Abdul-Rahman, H., Wang, C., and Skitmore, M. (2018). Exploring the underlying factors inducing design changes during building production. *Production Planning & Control*, 29(7):586–601.

Appendices

A.1 Summary of discussion with architects working at the neighborhood scale

This section presents the summary of notes captured from the discussions with five architects in Lausanne who had recently finished working on neighborhood scale residential projects to try and understand their design process. Several architects were contacted. Five agreed for discussion with the time frame of our interest. In each case the architect was requested to bring design exhibits from a recently completed residential neighborhood project that they had worked on. They were asked to speak about the early design process a few specific themes. Information regarding the architects with whom discussions took place is provided in Appendix A.1. They are referred to as architects 1,2,3,4,5. Architects 1,2 discussed small-scale medium-density projects (Density=1-2) with two buildings each. Architects 4,5 discussed high density projects (Density>3). Architect 3 discussed a low density project (Density=0.45) with a pre-exiting master plan. Following are the key topics which all architects were asked to speak about.

1. **Massing scheme:** The architects were asked to walk through the factors that they decided on their own and what was a design constraint. On the factors that they decided, such as orientation, they were asked to elaborate how they chose it. They were also asked about the order in which various design decisions were made.
2. **Motivations for exploring different massing-schemes:** All architects were asked to speak about their process for arriving at the final massing-scheme. They were especially asked if they created multiple options, how many and what were the differences between the alternatives.
3. **Facade design approach:** They were asked to speak about what factors they consider in the facade design and what was their approach was for the particular project.
4. **Managing project time and effort:** They were asked what design elements are easiest to change and which elements are hardest to change later in the design process.

Apart from the above items, several other topics came up during the discussion. Hand written

Table 1 – Inferences drawn from discussion with local architects working at the neighborhood scale.

Discussion theme	Input from discussions	Inference for facade design scenarios
Relation of facade to massing-scheme	Placement of windows and balconies could be linked to building placement within a massing scheme. The relationship could be two-way, where massing scheme could be developed to ensure privacy or windows could be placed on a massing to enhance privacy of views.	Architects were interested in exploring synergy between orientation of windows and massing scheme. Window orientation related decision could be dependent on the massing scheme.
Facade features dependent on other design factors	Placement of windows and balconies in plan is closely related to internal layout. Once window and balcony placement is decided, it is unlikely to change.	Positional relationships of facade elements (in plan) is rigid and may remain unchanged across levels of design details.
Independent facade features	Depth of balcony and size of windows appeared to be independent of other design features.	WWR and depth of balcony are specified independent of all other factors such as internal layout and there are no large dependencies on other elements. These factors (WWR and depth of balcony) can be chosen independent of massing scheme and other design choices

notes and a pre-prepared form was used to record the discussion. Inferences drawn from these discussions are presented in table 1. These were used for comparing design scenarios that were eventually developed.

The notes written during the discussion have been synthesized under two groups 1) facade design development 2) massing-scheme development:

Relation of facade features to other aspects of design: Architects 3,5 stated that window placement (in plan) and balcony placement (in plan) are closely tied to the internal layout of the building. Architect 3 said that to him facade design is a simultaneous composition exercise in plan and in elevation. He said he repeatedly goes back and forth between plan and elevation to check if the window distribution works for both the facade and the plan or not. Architect 3,5 said that once the plan is done and the project owner agrees, it takes a lot of work to change it. The balconies placement is tied to the plan and thus hard to change, but depth can be changed. Architect 1 mentioned that on her project, recessed balconies were completely eliminated in the final design. Figure 7.1 shows the project she mentioned. Figure

7.1 (a) shows the design as submitted for the design competition. Figure 7.1 (b) shows the final execution of the design without balconies. From these discussions it may be concluded when the facade has been planned in terms of window placements, we can assume that there is an internal layout associated with it. Balcony depth can be changed quite late in the design process as well. While placement of windows maybe tied to the layout, three out of five architects agreed that window sizes can remain a flexible design detail and can be changed, if needed, to accommodate energy, daylight or project budget constraints.

Development of massing-scheme and relation to facade design: Architect 4 said that the orientation of the building may also be chosen based on the internal layout. He mentioned that if there are apartment units on either side of the corridor with each apartment units getting access to one orientation only (shown as uni directional units in 3.5 (a)) then he would only place the building in east-west orientation. If the layout is of the type "bi-direction" (3.5 (a)) then he could place in any orientation. Architect 1, working on a medium density project and said that their primary criteria for building placement and orientation was the topography of the site and ensuring privacy for the building occupants. They tried to place buildings such that the balconies do not face each other. Architect 2 mentioned that they contemplated whether to have 2 or 3 buildings. Due to energy and cost considerations the scheme with 2 buildings was chosen. Total surface area of the building was discussed in relation to energy use. Placement of buildings was discussed in relation to views.

At high density, meeting the program requirements, while respecting density and requirements of the master plan for building placement seemed to overwhelm the architects. Both Architect 4,5 emphasized that it was quite difficult to arrive at a massing design that would meet all requirements. They said, a large number of massing-scheme were tried (>10) but very few solutions met all the requirements. Thus under demands of high (with respect to Switzerland) density (built density >3) the massing scheme was not discussed in relation to facade and views related concerns. For example, the project discussed by Architect 5 was adjacent to a large park but the park was only mentioned in the context of an additional set back imposed on the site, but not in terms of opportunity for views or daylight.

Additional insights relevant to early design stage performance evaluations: During these discussion we also found several instances that provide motivation for robust performance evaluations and decision making at the early design stage. The difficulty experienced by designers in balancing all design requirements at high density, need not be seen as deterrent to addition of performance evaluations. At high density there is greater need for planning for daylight and passive solar. The designer may be able to produce fewer massing-design alternatives and each design alternative takes notable amount of effort to produce. If performance evaluation is added as a decision criterion, an already stressed designer could potentially be highly motivated for robustness checks in performance evaluations. According to Architect 3, the phased permitting process is quite common (in the region) in which the massing-scheme is approved early in the design process and early design stage performance evaluation is thus a prudent step.

These discussion were also helpful in choosing the scale of the experiment (maximum 8 buildings) on site of 1.5 hectares. Architect 3 mentioned that projects larger than that, often get split into two or more projects as eventually it is not feasible for the single party to own such large projects. Separate owners, often imply separate parcels and separate design teams. Thus even in very large projects the design problem is often reduced to the scale of 5-8 buildings at the maximum (in the given regional context)

Architect 1 - Rita Cotungo
Name of Firm: ON Architecture
Position: Architect/project Head
Project Discussed: Logements 'Fiches Nord' , Lausanne
Density: Medium (1-2)
Image: ON-Architecture.ch

Architect 2 - Gael Cochand
Name of Firm: Tribu
Position: Architect-Urbanist
Project Discussed: Bochardon, Deux bâtiments de logements, Lausanne
Density: Medium (1-2)
Image: tribu-architecture.ch

Architect 3 - Diego Antonio Carrión Lobo
Name of Firm: Richter Dahl Rocha
Position: Architect/Project Head
Project Discussed: Quartier des Cèdres, Renens
Density: Medium (1-2)
www.richterdahlrocha.com

Architect 4 - Bart Daniels
Name of Firm: CCHE
Position: Urbanist/Project Head
Project Discussed: MEP Les Fèvres, Bussigny
Density: High(>3)
Image: cche.ch

Architect 5 - Gilles Humbert
Name of Firm: CCHE
Position: Architect/Project Head
Project Discussed: Petit-Mont-Riond, Lausanne
Density: High(>3)
Image: cche.ch

Figure 2 – Architects with whom early design process for neighborhood scale projects were discussed; Image indicates the project discussed with each architect.

Table 2 – Annual cooling demand for five schemes corresponding to various categories of the adaptive thermal comfort model.

Comfort category	NB1	NB2	NB3	NB4	NB5	Mean
EN15251 category I	19.0	26.5	28.4	19.1	41.0	26.8
EN15251 category II	15.8	22.9	24.3	16.1	36.9	23.2
EN15251 category III	13.0	19.5	20.6	13.3	33.1	19.9

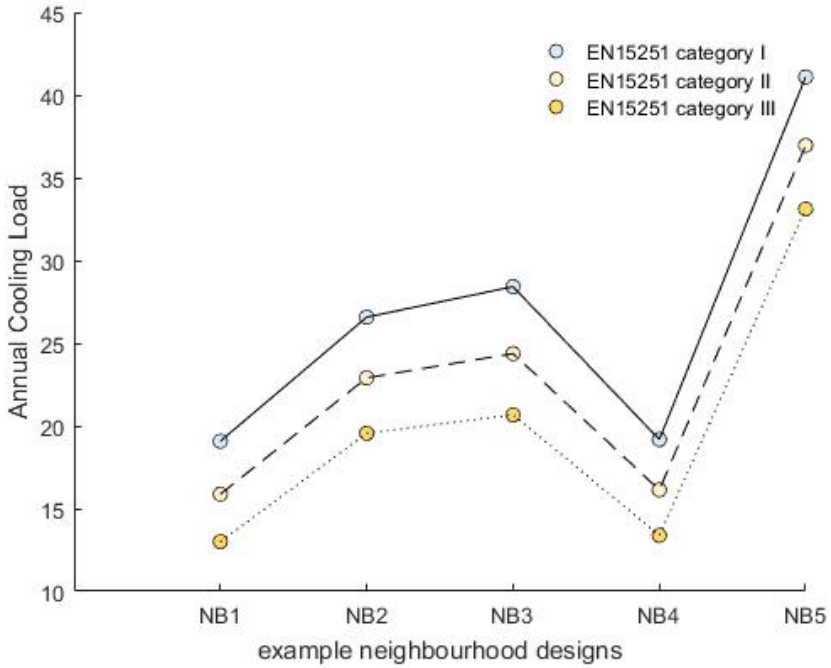


Figure 3 – Cooling load corresponding to five subject massing schemes modeled at fLOD0 to achieve varying degrees of thermal comfort in the building interior.

A.2 Data used for minimum performance differentiation threshold (dt) for annual cooling demand

A small set of simulations were done to determine a possible cooling demand based decision making threshold. The additional cooling demand that would be needed for meeting increasing stringent thermal comfort conditions (indoors) was estimated. The indoor room set-point was set up to in accordance with EN15251 categories I, II, III [Nicol et al., 2010] in five massing-schemes (fLOD0) models. The mean difference in energy demand was found to be quite consistent in all schemes (between the thermal comfort categories) even though their respective cooling demand varied significantly (-30% to +50%). See figure 3 and table 2 for outputs. The average difference in cooling demand between category I and II, and I and III were used as performance difference threshold for the high sensitivity (also baseline DM) and high sensitivity DM respectively.

A.3 Inputs for daylight and thermal simulations

Table 3 – Daylight simulation inputs.

Daylight Simulation inputs	
Location	Geneva, Switzerland
Weather File	Geneva IWEC
Opaque surface reflectance properties	
Internal surface of walls	0.5
Internal floor	0.3
Internal ceiling	0.7
Exterior wall surfaces	0.3
Exterior balcony elements	0.3
Ground surface	0.2
Glazing light transmittance	
Glazing light transmittance	0.6
Active Shading Properties	
Active blinds modelled as trans material to represent closed blinds with 10% light leak and glazing	Trans material properties: 0.06 (glazing +blinds closed) fully diffused transmission
Schedule	Hourly on/off schedule based on [Van Den Wymelenberg, 2012]
Simulation parameters	
Simulation program used	DAYSIM 4.0 [Reinhart, 2012]
Sensor array spacing	1m x 1m grid
Sensor array height from floor	0.76 m
Ambient bounces	6

Table 4 – Thermal simulation inputs.

Thermal Simulation inputs	
Location	Geneva, Switzerland
Weather File	Geneva IWEC
Envelope properties	
Opaque wall U-value (W/m2K)	0.14
Roof U-value (W/m2K)	0.36
Ground slab U-value (W/m2K)	0.36
Glazing U-value	1
Glazing SHGC	0.58
Window-to-wall ratio (WWR)	as per fLOD variant
Exterior reflectance value	0.3
Internal Loads	
Internal lighting (W/m2)	6.6
Lighting schedule	Off during unoccupied hours
Miscellaneous equipment (W/m2)	3.0
People density (p/m2)	0.025
Hourly schedules	As per SIA 2024-2015 [SIA, 2015]
Occupant Controls	
Blind controls	Hourly on/off schedule based on [Van Den Wymelenberg, 2012]
Fixed shading	as per fLOD variant
HVAC system inputs	
Heating setpoint with setback, setpoint (°C)	18, 20
Cooling setpoint with setback, setpoint (°C)	28,25
System Type	Purchased heating and cooling
Ventilation rate	30 m3/h.person
Heat recovery	not included
Demand controlled ventilation	included
Heating/Cooling system efficiency	1
Simulation parameters	
Program used	Energy plus ver 8.4
Simulation time step	1 hour
Shadow calculation frequency	7 days
Solar distribution calculation	Full exterior with reflections
Warm up days	7 days (default)

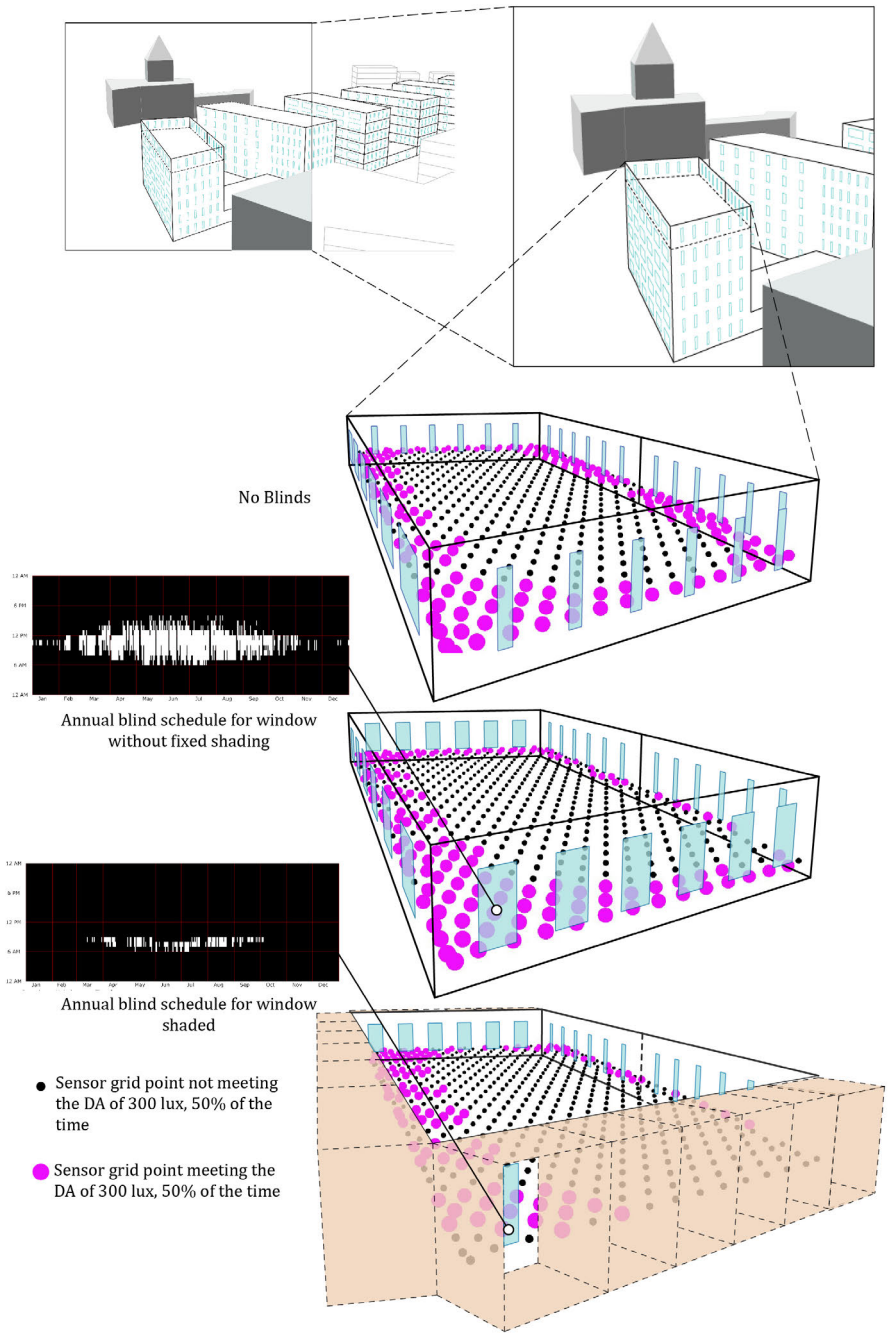


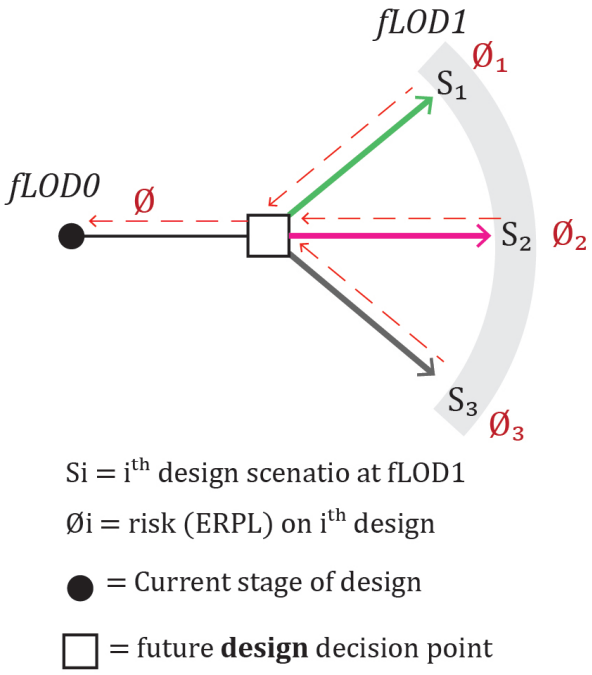
Figure 4 – Detail of one zone of the Citelac massing scheme; blind schedules corresponding to one window are also shown.

A.4 Methodology for calculation of risk threshold for multiple comparisons

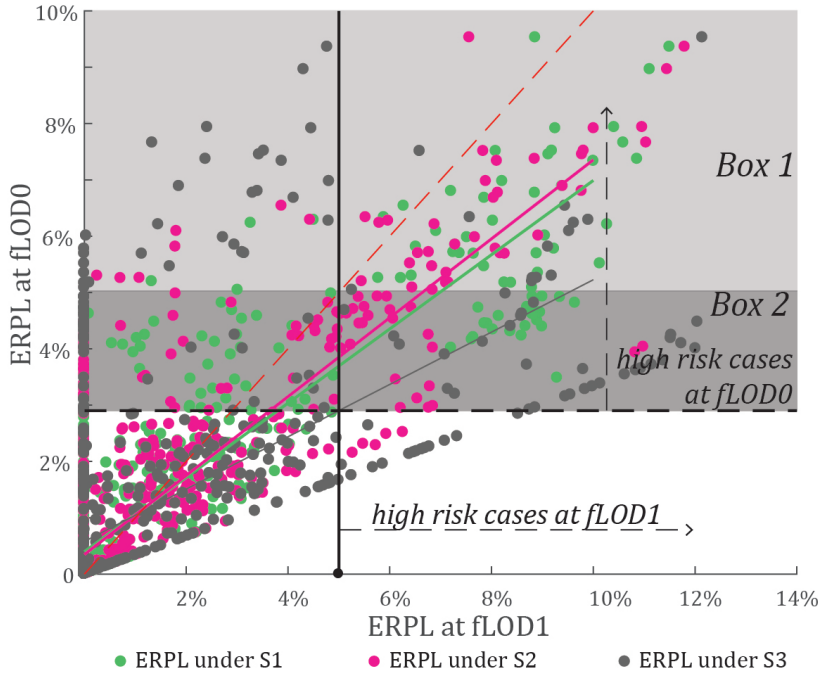
At fLOD0, all future design scenarios are considered equally probable. Some scenarios result in opportunity loss (regret). However, at the onset of the design process, the DM does not know which design path they are going to follow that leads them to particular design scenario. They can only know the collective risk from all the scenarios. At the 'final' stage of design, the maximum acceptable risk has been assumed to be 50% chance of unacceptable loss. It is important to relate this maximum permissible risk at the end of the process to the risk seen by the DM at fLOD0, which is the average risk from all design paths.

Figure 5 (a), as an example, shows the risk values $\theta_1, \theta_2, \theta_3$ from future design paths leading to scenarios S1,S2 and S3 (fLOD1). Since the DM does not know which design scenarios are going to be realized in the future, they are all treated as equally probable. The risk θ at fLDO0 is thus the average of the risk at S1,S2 and S3 (fLOD1). Figure 5 (b)shows the ERPL values found at fLOD1 in the experimental data gathered from this study. $\theta_1, \theta_2, \theta_3$ from the comparison are plotted on the x-axis and the corresponding average value for each case (each instance of the experiment) is shown on the y-axis (figure 5 (b)). If the maximum permissible risk is set to 5% (sDA) (50% change of 10% loss on sDA) at fLOD0, 'Box 1' (figure 5 (b)) indicates the comparisons that would qualify as high-risk. A lot of comparisons with high risk ($\theta_1, \theta_2, \theta_3$ >5%) get left out of 'box 1'. The risk threshold at fLOD0 needs to be lower in order to warn the DM that there is high-risk under some future design scenarios. A new threshold can be established for fLOD0 based on the correlation lines (solid colored lines in figure 5 (b)). At this new lowered threshold, more comparisons with high-risk under future design scenarios can be captured. Now, comparisons falling in 'box 1' and 'box 2' both, would rightly qualify as high-risk. Some high-risk comparisons are still left out and is a limitation of this method. However, lowering the risk threshold further could imply that some low risk comparisons may get classified as high risk.

This example shows a simplified version of this calculation only including adjustment of risk threshold based on observations at one fLOD (fLOD1) to another (fLOD0). For the risk threshold used in Chapter 4, this process was repeated through all the fLODs meaning from fLOD2 to fLOD1, from fLOD1 to fLOD1-no blinds and from fLOD1-no blinds to fLOD0 to calculate the risk threshold at fLOD0.



(a)



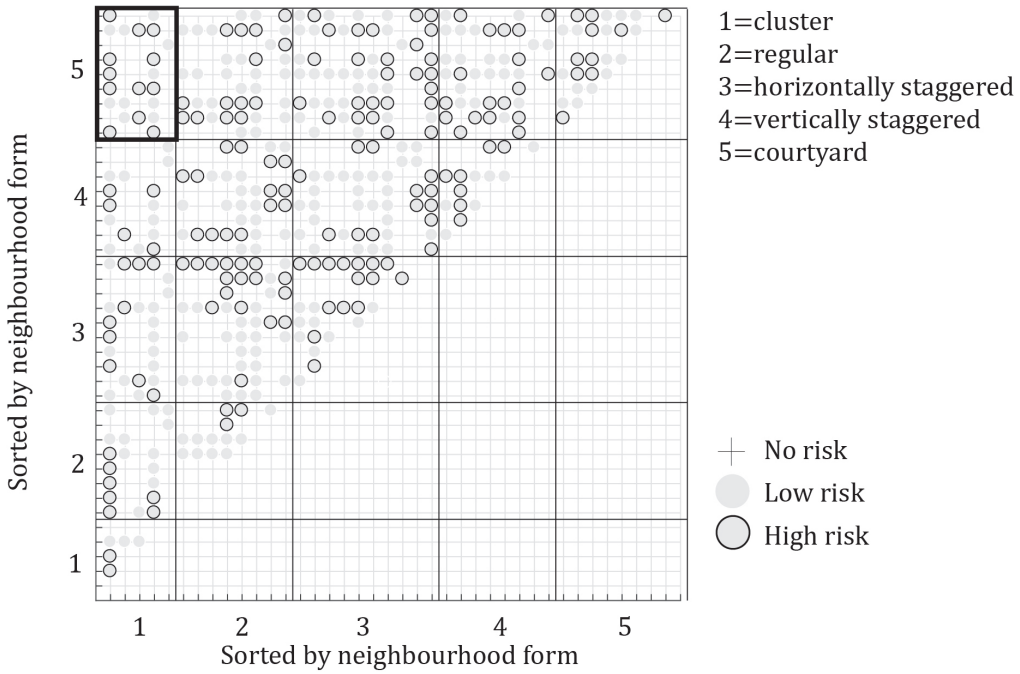
(b)

Figure 5 – Illustration of methodology for calculating risk-threshold used in Chapter 4 (a) diagram showing risk at lower fLOD (fLOD0) as the average of the risk on future facade design scenarios. Any of the paths S1, S2, S3 could be taken by the DM when increasing the model fLOD to fLOD1 (b) ERPL values at individual design scenarios S1,S2,S3 are shown on x-axis, and the corresponding risk at fLOD0 is shown on y-axis.

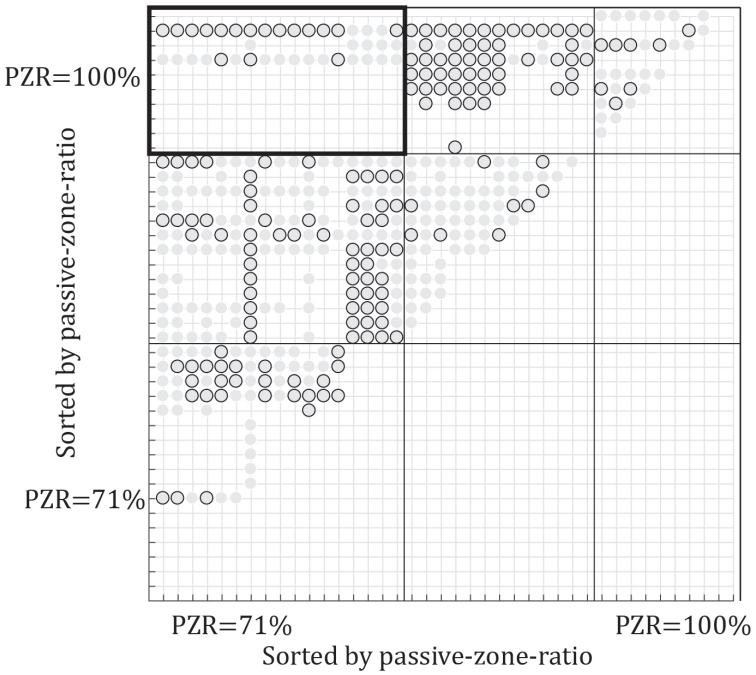
A.5 Graphical exploration of risk inducing neighborhood massing scheme characteristics

The figures in this appendix show initial exploration of risk found in the comparison of 780 pairs of massing schemes. Figure 6 shows all high risk (high ERPL) and low risk on sDA. Every grid intersection in this figure marks a comparison between two massing schemes. A grid intersection with a circular marker shows a case where risk (ERPL) of performance loss was observed. Circular markers with the black edge show high risk cases. When the comparisons are ordered by the form of the massing-scheme (Figure 6-a) (1. cluster, 2. regular shapes, regularly placed, 3. horizontally staggered blocks, 4. blocks with varying heights 5. courtyard) no apparent trend or clustering of high risky cases can be seen. Risky cases, including low risk cases are found in all cells, whether comparisons were done within the same typologies or a comparison is done between schemes of different typologies. However, when the schemes were sorted by passive-zone-ratio(PZR) (Figure 6-b), the distribution of risky cases does get differentiated and to some extent limited to medium PZR cases being compared to other PZR cases. The risk in pair-wise comparison thus exhibits some patterns when the pairs are sorted by specific design characteristics. Similar effects are seem on heating and cooling demand evaluations as well, shown in figure 7.

The highlighted cell on the top-left corner indicates comparison between two 'extreme' or different types of massing schemes. For example, the highlighted in figure 6 (a) shows comparisons between 'cluster' type and 'courtyard' type neighborhood schemes. The highlighted Figure 6 (b) shows comparison between schemes with the lowest passive zone ratio, to those having the highest passive zone ratio (within the scope of the experiment). If more risky cases are found in this cell, then differences between proprieties of massing massing schemes induces risk. If more risky cases are found in the top-right cells or the bottom-left cells, then risk could be induced when two massing schemes having similar properties are compared.



(a)



(b)

Figure 6 – Risk (ERPL, sDA) in pair-wise comparison of massing scheme sorted by (a) massing typology (b) passive-zone-ratio.

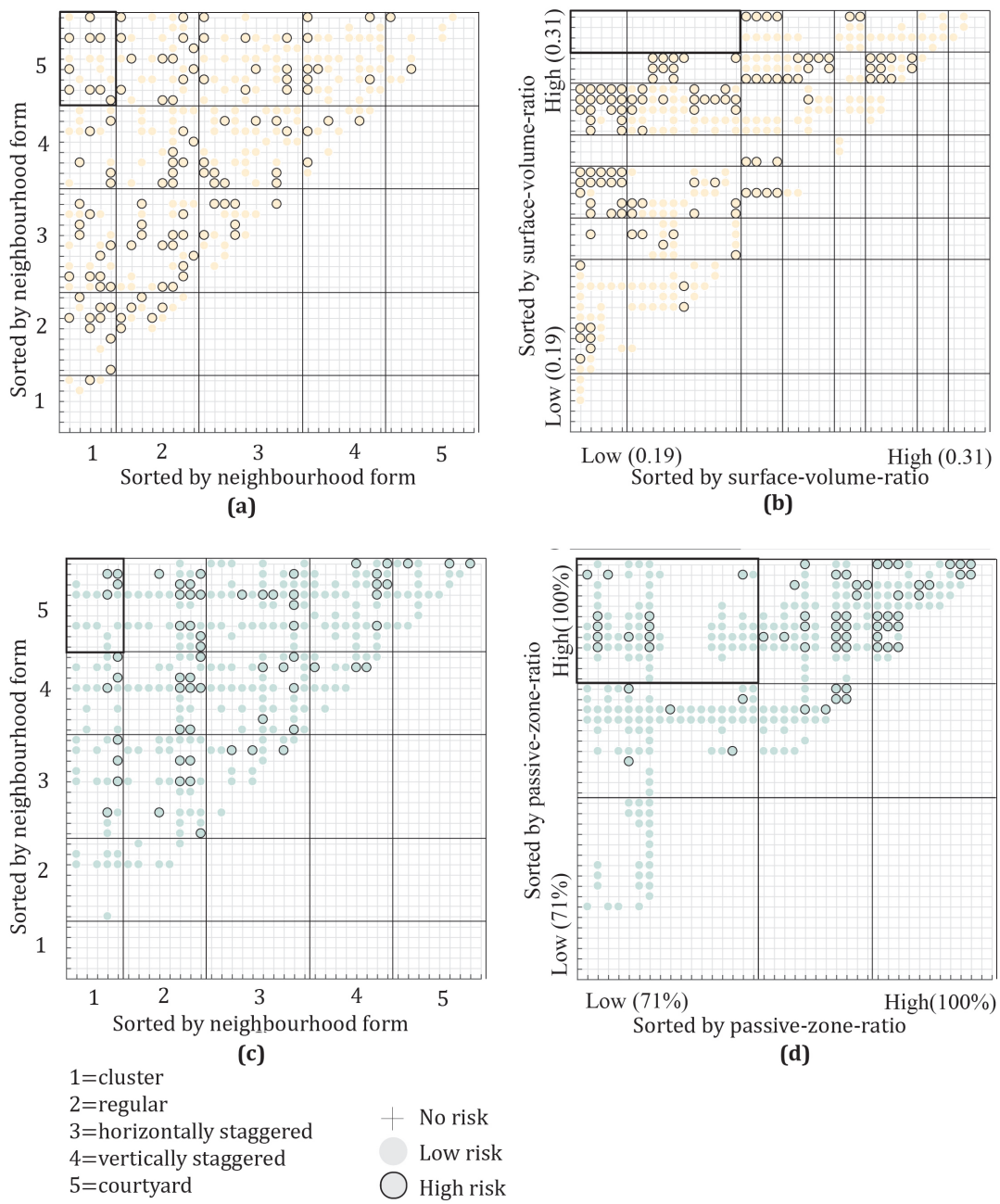


Figure 7 – Risk in pair-wise comparison of massing scheme sorted by (a) massing typology , evaluation metric = cooling demand (b) passive-zone-ratio, evaluation metric = cooling demand (c) sorted by massing typology , evaluation metric = heating (d) sorted by surface-volume-ratio, evaluation metric = heating demand.

A.6 Summary of properties of subject massing schemes used in experimental evaluation of risk

Table 5 – Properties of subject massing schemes extracted for statistical analysis. The variable type is an alphabetical code used in this thesis for easier referencing of variable types.

Variable type	Factor name-code	Factor/early design performance indicator	Maximum Observed value	Minimum Observed Value	Mean
NA	TYP	Typology (categorical)	C NA	NA	NA
K	SVR	Surface Volume Ratio (m2/m3)	0.31	0.19	0.24
K	MEH	Mean Building Height (number of floors)	8.4	3.5	5.68
K	SCO	Site Coverage Ratio (%)	28%	14%	20%
K	NBO	Number of Buildings	8	3	4.68
K	PZR	Passive Zone Ratio (%)	100%	71%	84%
K	CEX	Complexity (m2/m2)	0.86	0.52	0.65
K	MOD	Mean Outdoor Distance (m)	18	6.5	11.29
K	DIR	Directionality [-]	781	118	339
L	TIVf	Mean annual irradiation on vertical surfaces per unit area of vertical façade area (kWh/m2)	618	561	593
M	TIFa	Mean annual irradiation on vertical surfaces per unit area of floor area (kWh/m2)	568	307	383
N	pHEAT_1	Passive solar heating potential (heating period irradiation threshold =208 kWh/m2) (%)	75%	49%	62%
N	pHEAT_2	Passive solar heating potential (heating period irradiation threshold =178 kWh/m2) (%)	91%	61%	78%
N	pHEAT_3	Passive solar heating potential (heating period irradiation threshold =156 kWh/m2) (%)	91%	62%	80%
O	pOVERHEAT_1	Overheating avoidance potential (cooling period irradiation threshold =201 kWh/m2) (%)	41%	13%	30%
O	pOVERHEAT_2	Overheating avoidance potential (cooling period irradiation threshold =251 kWh/m2) (%)	91%	61%	77%
O	pOVERHEAT_3	Overheating avoidance potential (cooling period irradiation threshold = 301) (%)	92%	62%	81%
P	pDAYLIGHT_1	Daylight potential (total annual irradiation threshold = 5k lux) (%)	100%	95%	99%
P	pDAYLIGHT_2	Daylight potential (total annual irradiation threshold = 10k lux) (%)	91%	61%	78%
P	pDAYLIGHT_3	Daylight potential (total annual irradiation threshold = 15k lux) (%)	81%	9%	22%

Minu Agarwal LEED AP(BD+C)

Doctoral (Ph.D) Assistant at Laboratory of Integrated Performance in Design
École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland
Email : minu.rke@gmail.com, Phone +41 79 136 66 87

Professional Experience

Ph.D. Assistant at Laboratory of Integrated Performance in Design 2015-June 2020
EPFL, Lausanne, Switzerland

Project Consultant at Integrated Environmental Solutions Ltd., Atlanta USA 2009-2015
Consulting for green building performance rating and energy use optimization
Interfaced with developers, design and engineering firms globally
Applications for green design certification, rebuttal on reviews (GBCI, USGBC)
Extensive experience in modelling HVAC systems and controls

Building Physicist at Buro Happold Consulting Engineers, New York, USA 2007-2009
Building engineering for low energy design, façade optimization for solar gain & daylight penetration
Developed an MS excel based tool for life cycle cost analysis and cost payback analysis

Education

Doctor of Philosophy (Civil and Environmental Engineering) 2015-June 2020
EPFL, Lausanne, Switzerland
PhD successfully defended on May 4th, 2020.

Master of Science in Sustainable Design 2007
Carnegie Mellon University, Pittsburgh, USA

Bachelor of Architecture 2005
Indian institute of technology (IIT), Roorkee, India

Peer Reviewed Conference Papers

1. M. Agarwal, G. Danseux, L. Pastore and M. Andersen. *Reliability of daylight and energy demand evaluations for decision making at the conceptual design stage.* Proceedings of BS2019, Rome, Italy, 2019
2. M. Agarwal, L. Pastore and M. Andersen. *Influence of façade details on early design decisions regarding daylight performance of neighborhoods.* Proceedings of the Building Simulation and Optimization Conference 2018, Cambridge UK, 2018
3. M. Agarwal, L. Pastore and M. Andersen. *Suitability of neighborhood-scale massing models for daylight performance evaluation.* Proceedings of the International Conference on Sustainable Design of the Built Environment (SDBE) 2017, London, UK, 2017.
4. M. Agarwal, P. Rastogi, M. M. V. Peltier, L. Pastore and M. Andersen. *Examining Building Design Decisions Under Long Term Weather Variability and Microclimatic Effects: A case-based exploratory study.* Proceedings of PLEA, Los Angeles, USA, 2016.

Journal Papers

1. *How useful are massing models? Performance evaluation at the conceptual stage for neighborhood scale design* with L. Pastore and M. Andersen **(revise and resubmit)** : Journal of Building Performance Simulation
2. *Green Hotels: An Overview*. P. Das, M, Agarwal: Boston Hospitality Review, Winter 2019

Book Chapter

1. Book Chapter- 'Green Real-Estate Trends in India', Book: 'Real Estate Finance in India,' with D. Sharma, *Sage Publications,* India, 2014

Teaching Experience

EPFL | École polytechnique fédérale de Lausanne, CH

Teaching assistant for Masters level course: Comfort and Architecture, Sustainable Strategies (2017-2019)
Master semester project supervision

- Visual-Thermal comfort evaluation and comparison: application to an open-space office (2018)
- Energy modeling of the Swiss Solar Decathlon Pavilion: testing the thermal resilience under unplanned use (2017)
- Energy modeling of the Swiss Solar Decathlon Pavilion: development and fine-tuning of solar passive strategies (2017)
- Improvement of solar passive behavior of raw earth in Zanskar, India (2015)

Realism Real Estate Consultancy Pvt Ltd, India

Online educator: Single handedly developed and managed course for LEED accreditation exam preparation (2011-2013)

Cooper Union Continuing Education, NY, USA

Course Instructor: Whole building energy performance analysis for introduction to building performance modeling (Spring 2009)

Presentations

- *Human centric building design: Meaning and implementation (online)* Anant National University, Ahmedabad, India, 2020
- *How to kick-start a LEED project?* Corporate Social Responsibility (CRS) strategies class, Ecole Hoteliere Lausanne, Switzerland, 2020
- Panel discussion on *Simulation to Support Regulations – ASHRAE 209-2018* at IBPSA Rome, 2019

- *Research applications of parametric modeling tools* at School of Architecture, Civil and Environmental Engineering (ENAC) Research Day, EPFL, Lausanne, 2018
- *LEED for beginners*, School of Architecture and Planning, IIT Roorkee, 2015
- *Multidimensional analysis of a low energy naturally ventilated buildings*, Energy Systems Laboratory, Texas A&M University, College Station 2009

Academic Service

- Contracted peer-reviewer for Green Building Certifications Institute, USA (2012-2013)
- Conference paper peer-reviews for PLEA, IBPSA International
- *Exam jury for Corporate Social Responsibility(CRS) strategies class,* Ecole Hoteliere Lausanne, Lausanne, Spring/Fall, 2019

Editorial Experience

- Managed white paper series 'Gyanism' for Realism Real Estate Consultancy Pvt Ltd, India (2011-2013)
- Co-editor for 'A Contemporary review of Mithila Art' Cognitio Publications, India (2021 expected)

Skills

- Building Simulation Software: IES VE with extensive use of HVAC systems modeling modules, Energyplus, DesignBuilder, eQUEST, ArchSIM (for teaching only), Energy-10, Radiance (Rhino plugins such as Honeybee and DIVA)
- CAD tools: Rhino (with Grasshopper plugin (intermediate), Revit and AutoCAD
- Programming Languages: Matlab (intermediate), Python (beginner) and R (beginner)
- Statistical modeling
- Graphic Design Tools: Adobe Illustrator, In-Design and Photoshop